The University of Hull University

Educational Technology: towards an understanding of effective technologies, with particular reference to literacy.

being a Thesis submitted for the Degree of Doctor of Philosophy

in the University of Hull University

by

Kenneth A. Spencer, MSc

June, 2000.

TEXT BOUND CLOSE TO

THE SPINE IN THE

ORIGINAL THESIS

TEXT CUT OFF IN THE ORIGINAL

TABLE OF CONTENTS

Chapter 1: Educational Technology: towards an understanding of effective technologies

Chapter 2: The Papers

Chapter 3: Discussion of the papers

Chapter 4: Conclusion

Chapter 5: References

Appendix 1: Published work 1977-2002(in press)

Appendix 2: Supplementary Materials 1999-2002(in press)

1 Educational Technology: towards an understanding of effective technologies

Introduction

The term educational technology is so ambiguous that its use often results in confusion. Recently it has been seen as a synonym for Information & Communication Technologies, and this has led to misunderstandings as to the outcomes that can be anticipated when ICT is applied to education.

"Technology in education" and "technology of education" (see Spencer, 1988, Chapter 1) are also often seen as being synonymous, whereas they are very different conceptions (although complementary).

In order to understand why the embedding of Educational Technology has not been ubiquitous, it is necessary to analyse educational research during the past century and to relate this to expectations of what could be achieved through the use of technology in education. By reviewing 150 years of research in this field it is clear that technologies can be employed as supplements or aids for teachers, or as replacements for teachers, and there will be no detrimental effects on pupil performance. When the technology of education is considered it is clear that such methodologies not only equal traditional approaches, but more usually produce educationally significant improvements (i.e. effect sizes of +0.5).

This thesis will consider research evidence in the following fields: audio-visual approaches to learning (including film, television, audio-tutorial methods and radio) and the realism theories of the 1950s; models of human information processing and the change in perspective; behavioural objectives, mastery learning and interactive learning systems. It will conclude with a consideration of the applicability of mastery techniques to the teaching of reading and writing, and present an analysis of the written form of the English language with a view to establishing a new methodology for literacy teaching and learning.

Early educational technologies

The first technology must be language, the very powerful tool that allowed accumulated knowledge to be passed from one generation to the next, speeding up the evolutionary process. Then, after thousands of years, came written language, about 5,000 years ago, allowing thoughts and ideas to be transmitted across almost limitless vistas of time. Again, the process of human evolution was speeded up:

> The invention of writing was highly significant for the development not only of language, but of society, and favoured the progress of commerce. It confirmed the power of the priests through the trained scribes, and even more the

might and prestige of the ruler. (Singer, C., Holmyard, E.J. & Hall, A.R. (1954) A History of Technology, Volume 1, Part 1, Chapter 4: Speech and Language.)

The next major step forward was the invention of moveable type. Printing allowed a dramatic increase in the rate and volume of information distributed, and drove the European Renaissance. As many books were printed in the fifty years after Gutenburg's invention as had been produced by the scribes of Europe during the previous thousand years. And after such a revolutionary educational invention, things did not change a great deal for 500 years. Teachers taught a variety of subjects using basically the same tools: books and writing materials. Indeed, The Visible World in Pictures (Orbis Sensualium Pictus), produced by Comenius in 1658, was still widely used in schools the 19th century (Saettler, 1990).

So, to a certain extent, teachers have always used technology, but the available technology was rather limited. With new technologies being introduced, in the 19th century, to boost production and aid in the distribution of the products of industrialisation, there was also a demand for improvements in education to meet the need for skilled workers and clerical staff. This was met by the monitorial training system introduced by Lancaster, based on Bell's ideas: it was a very efficient, but labour intensive, system, capitalising on human resources rather than the introduction of machines. Indeed, while Lancaster was introducing his system of schooling Charles Babbage and Ada Lovelace, were laying the foundations for the future revolution in educational technology, with their design for the precursor of the modern programmable computer, the analytic engine, which ran programmes from punched cards.

The twentieth century has been obsessed with the idea that the new communications technologies, such as film, radio and television, should make a significant impact on education, and that the introduction of such audio and visual aids should raise the level of achievement of pupils. The main early theory governing this approach, the realism theory, was summarised by Carpenter (1953):

> Sign similarity hypothesis: that films whose signals, signs, and symbols have high degrees of similarity ("iconicity") to the objects and situations which they represent will be more effective for most instructional purposes than films whose signals, signs and symbols have low degrees of "iconicity"... Signs (or symbols) vary in "iconicity" to the degree to which they are similar to the things or situations signified. Thus, for example, sound motion pictures have potential capacities for high degrees of "iconicity" in representing objects in motion as well as reproducing authentic sound. (p. 41)

Unfortunately, most of the research, which compared the new media, such as film and television, with traditional ways of teaching, found that there were no significant differences in student performance.

Early Film Research

With the advent of sound films for educational purposes, in the early 1930s, it was natural for researchers to consider the relative effectiveness of lectures and filmed recordings of such lectures. Hoban and Van Ormer's report, subtitled *Rapid Mass Learning*, summarised the research available during the period 1918 to 1950. The earliest research investigating comparisons of sound films and lecture demonstrations was conducted by Clark in 1932. Three sound films, *Radioactive Substances*, *Liquid Air*, and *Characteristics of Sound*, were found to be as effective as the lecture demonstrations given by regular class instructors, in tests designed to measure thinking and reasoning ability.

A later study by Hall and Cushing in 1947 investigated the difference between a sound film presentation and a lecture with enlarged illustrations, dealing with 3 science topics. No differences were found.

In an experiment conducted by Vernon (1946) information test scores of seamen who witnessed two showings of a film demonstrating two methods of taking depth soundings, a total of 50 minutes of instructions, were only 6 per cent below those of groups which had usual instruction lasting 3 hours. Vernon concluded that an hour's film appears to be as effective as three hours weak oral instruction.

Hoban and Van Ormer summarised their review of the research, in which films are compared with demonstrations or lectures, and conclude that films reduce instruction time and are often equivalent to good instructors. They hasten to add that this should not be interpreted as meaning that films can eliminate the need for instructors but rather that the effectiveness of instructors of average or below-average ability can be improved (and instructional time can be saved). Also, they suggest that films can be projected on large screens, increasing the size of the viewing group but with no loss in instructional effectiveness. Finally, they conclude that films, used alone, can offset a shortage of instructors.

Films Reduce Instruction Time:

The conclusion which continually recurs in these studies is that films reduce instruction time with little or no sacrifice of instructional results. In some of the experiments in film presentation, from one-half to two-thirds of the instructional time was saved by the use of films in place of lecture or demonstration. There thus appears to be considerable support for the Navy slogan "More learning in less time", although this may not always be true for all films.

Films are Often Equivalent to Good Instructors:

A second conclusion that is recurringly supported by the research data is that, in communicating facts and demonstrating concepts, films (or filmstrips) are about

equivalent, and sometimes better than superior instructors using the best non-filmic materials at their disposal.

Hoban, C.F. and Van Ormer, E.(1950) Instructional Film Research, 1918-1950. Port Washington, N.Y.: US Navy Special Devices Centre.

From the results of the first thirty years of research comparing the comparatively new educational media there is strong confirmation of their effectiveness when compared with traditional methods. Indeed, Hoban and Van Ormer suggest that by using film recordings of above average teachers there can be some compensation for poorer quality personnel.

Peggie Campeau's stringent review of the literature concerning audio-visual media, in 1966, cites nine studies conducted at university, senior high school, junior high school and elementary school levels in which no significant differences in achievement were found when students were taught by either motion pictures or conventional instruction. This is confirmed by Greenhill in his introduction to a volume of abstracts on film and tv research by MacLennan & Reid (1964).

Early Television Research

Just as research was indicating quite conclusively that there is no disadvantage in studying from filmed courses, a new technological innovation was entering the mass media market place. Television, although invented in the late 1920s, was only beginning to make headway in the early 1950s, and it was natural for researchers to turn their attention to a medium which offered all the potentialities of film, at a possible lower cost, for, as Lumsdaine and May concluded in 1965, film and tv can be considered substantially identical media for many purposes.

The first research effort used television as a means of expanding the total audience for a given lecture via closed circuit television, in which the transmission was carried from the lecture room, by cable, to several locations, enabling a single teacher to communicate with many hundreds of students. In fact, at one stage, audiences of 7,000 were taught via closed circuit tv in New York University. In Greenhill's report of the Pennsylvania experiment he acknowledges that the major use of closed circuit tv was for the presentation of regular classroom instruction to students located in multiple classrooms, as a means of coping with mounting enrolments. Greenhill also argues that the standard of instruction would also be raised by:

(1) extending the influence of its best professors to large numbers of students and

(2) by making it possible for these professors to present demonstrations and other teaching materials that it would be impossible or impractical to use under normal classroom conditions.

This new medium, which is substantially the same as motion film, in that it presents moving pictures with sound accompaniments, should produce substantially the same results when compared with traditional teaching. The shift towards tv, which was viewed as a panacea for all the educational ills during the mid 1950s and early 1960s, resulted in a proliferation of research studies. Stickell, in 1963, reviewed 250 comparisons of educational television and conventional face-to-face instruction from 31 research reports. Overall, 75% of the studies showed no difference, with equal percentages favouring tv or face-to-face instruction. Chu and Schramm found that, by 1968, of 421 separate comparisons taken from 207 published reports, 308 showed no difference, 63 showed tv to be superior and 50 found conventional instruction superior.

The research reviewed was so wide-ranging that they concluded that tv can be used efficiently to teach any subject matter where one-way communication will contribute to learning.

These results lead to the inevitable conclusion that course enrolment can be greatly expanded by the use of educational television and that student performance will not suffer. However, the second point referred to by Carpenter, concerning the possible improvements in instruction, does not seem to be tenable. If this is the case, what are the advantages of using film or television? The answer is: when there is no difference in performance, the most obvious measure then to come under scrutiny is the cost.

An early cost analysis is available in Greenhill's final report on tv teaching at Pennsylvania State University, during the period 1956-57. Comparisons were made between actual costs of televised instruction and the costs that would have been incurred in courses had they been taught in the usual way. The analysis showed a total saving, in favour of the tv service, of $40,000, which represented more than the total cost of running the service. However, it was necessary to have at least 200 students per course before any savings were made. The system required large recurrent finance to keep it in operation and, if the system was not fully operational for most of the academic year, it rapidly became less cost-effective.

Radio and Audio Recordings

It is generally recognised that for teaching via radio or audio recordings essential graphic information, usually in the form of printed materials, must be provided in order to fully exploit the potential of the medium. This means that comparisons of radio and traditional teaching are often comparisons of audio plus print with traditional teaching.

Beginning in the 1920s instructional radio was widely used in the United States and Britain, but with the advent of television its use dwindled in the US, although it has continued to be widely employed by the BBC schools service. Developing countries, however, are making increasing use of radio, its principal attraction lying in its low cost when compared with television. It is also an effective instructional medium, as much of the research confirms.

Carpenter, in 1937, prepared 15-minute radio lessons on science for pupils ranging from fourth grade to senior high school level. The results of the end-of-term examinations indicated that pupils taught by radio did as well as, or better than, those taught by conventional methods. Attitude reports from pupils showed a high degree of interest in radio lessons. Heron & Ziebarth (1946) found that the radio was as effective as the face-to-face instruction in college psychology courses.

Another example of the effectiveness of radio as a teaching instrument, this time in a developing country, is reported by Mathur & Neurath (1959). A total of 145 villages in Bombay state, averaging about 850 people per village were chosen as the experimental group and were provided with radio sets. A similar number of villages without radio sets served as the control group. Twenty special farm programmes were broadcast twice a week for 30 minutes. Comparison of test results, both before and after the broadcast programmes, found a significant increase in knowledge in the radio villages, but only negligible increases in the non-radio groups. This and many other reports indicate the efficacy of instructional radio. Forsythe's (1970) review concluded:

> Research clearly indicates that radio is effective in instruction. Experimental studies comparing radio teaching with other means or media have found radio as effective as the so-called "conventional methods". Even though radio has been criticised for being only an audio medium, studies have shown that visual elements in learning are not uniformly important. In many educational situations visuals may be more harmful than helpful. Also, the efficiency of combined audio and visual media has been challenged by studies which show that multi-channel communications may not be inherently more effective than single channel presentations.
> Forsythe, R.D. (1970) Instructional Radio: a position paper. Stanford University. ERIC ED 044 933

These conclusions can be extended to include the equivalent recorded form of instruction, such as discs or tapes. Popham, in 1961, divided an introductory graduate course into two sections. In one he taught in a lecture-discussion format; in the other, he played a tape-recorded version of the lecture and then led a brief discussion period. The two groups were carefully matched on scholastic aptitude and two achievement pre-tests. Following instruction several post-tests were administered and it was found that there were no differences between the groups.

By 1969, the rapid expansion of cassette tapes was making the medium ideally suited to individualised learning and it was this aspect which was of interest to Menne, who used it in an introductory psychology course at Iowa State University. The lectures were recorded on tape and notes were taken from the blackboard material used by the instructor during the presentation of his lectures. The blackboard notes were then assembled to form a booklet. Each member of the experimental tape group was issued with a tape recorder, a complete set of lecture tapes, a booklet containing the

transcribed blackboard material and a schedule of the lecture topics to be given to the lecture group. The audio-taped group was self-paced, though they were required to take 3 objective tests during the course. Information was available concerning student performance on several measures, which enabled a covariance analysis to be applied to the results for the regular exams, class points and a final grade. There were no significant differences.

Ackers and Oosthoek (1972) report a similar experimental course. The subjects in the taped group again had individual access to recorders and tapes on the subject of micro-economics, and were able to follow the course at their own speed, within certain broad limits. Ackers does not elaborate on the broad limits and we can consider the taped course to be substantially student-paced. Both groups had ample opportunity to participate in test problems, which formed an integral feature of the instruction, and were encouraged to take part in fortnightly group discussions. The performance was assessed in a June examination which, according to the authors, called for the sub-categories 'Application' and 'Analysis' of the category 'Comprehension' from Bloom's Taxonomy of Objectives. The results indicated a slight advantage in favour of the tape group.

General Conclusion from early media research

Much of the early research was concerned with the new mass media, and it was clear from this research that these new approaches produced results similar to more traditional methods of teaching. Traditional approaches had a limited number of technologies embedded in them, such as books, and writing and drawing materials; the new approaches incorporated the teachers themselves, whose performance had previously been evanescent. The new media removed the need to have teachers actually present in the classroom, because they can produce a facsimile of the teacher. This represents a significant change, because prior to the introduction of the new media, if a teacher was not present, instruction could only be given through the medium of print, which need a decoding skill, the ability to read, to be present in the student. The new media literally spoke directly to the students, and did so as effectively as if the teacher was actually present. This lies as the heart of the embedding process: the new media could actually replace teachers, although researchers usually denied that this was a possible outcome, and their results were euphemistically disguised as demonstrating that the new media could provide a means to compensate for a lack of teachers. Nonetheless, teachers and producers of media programmes (e.g. the BBC) were well-able to perceive the threat from the new media.

Media Attribute Research

Early research into the effectiveness of media considered the effectiveness of the new medium compared to traditional approaches. In addition to this, groups of media which differed along one dimension only, were compared in order to determine the effectiveness of a particular attribute, such as visual motion or pictorial colour.

Motion

One of the earliest accounts of the experimental investigation of media effectiveness is Freeman's (1924) Visual Education, published 20 years after the first public demonstration of moving pictures by the Lumiere brothers. The results are not as scientifically valid as later research, but they do confirm later findings.

McClusky's (1924) experiment represents one of the first comparisons between film and a lecture illustrated with slides. He used a film on the life history of the Monarch butterfly and compared this with two lecture conditions: a slide lecture using eight slides, each illustrating a step in the life cycle; an oral presentation illustrated with two pictures and two blackboard sketches. Each was presented to 20 pupils in grades 6-8 in two schools and lasted 12 minutes. The results failed to show any difference between the methods.

Brown (1928) found similar results when comparing films and filmstrips for teaching factual information about the physiology of seeing to high school students. In the filmstrip group discussion was free and questions were asked both by the teacher and the students. A multiple choice test indicated a superior performance for the filmstrip group and Brown concluded that this was because of the greater exchange of comment within the teacher-paced filmstrip group.

A more satisfactory approach to determining the effectiveness of the visual motion attribute was undertaken by Twyford (1954). The topic investigated was methods of riot control, under the title Military Police Support in Emergencies, and introduced the problems of training soldiers to cope with such complex situations as restless, disturbed city populations, agitated groups, mobs and rioting crowds. The film had a Hollywood budget and expensive crowd scenes organised in an American city. The question of simpler and less expensive production methods was raised and Twyford was charged with determining the effectiveness of other methods. Twyford's group suggested an alternative approach using stock film or newsreel coverage of riots, and even the use of still pictures if motion film was not available. The project eventually compared the Hollywood style film with two filmograph versions, which were similar to sound filmstrips. In all three versions the soundtrack was identical. One filmograph was based on the motion film and consisted of individual still frames taken from the original. The second filmograph was made up of stock still pictures of riots taken from news libraries, or simple diagrammatic representations of troop movements. The groups of recruits were tested using a 42 question test with 10% of the questions using pictures. The full motion version scored 4% more than the filmographs, which were equally effective. The difference in performance was educationally insignificant, but the difference in cost was very substantial.

This is one of the earliest well-controlled experiments that shows that motion aids such as film and television will not automatically improve student performance when compared with simpler aids such as filmstrips. The essence of this argument is that film and tv can teach many different groups and subjects about as effectively as traditional methods, but so can simpler aids such as sound filmstrips or sound tapes with booklets, and these simpler aids cost less to produce.

Colour and Pictorial Quality

It is also worth, at this point, considering the experimental evidence concerning comparisons between colour illustrations and their counterparts in monochrome, and the effects of changing pictorial quality.

Vander Meer's (1954) work has been described as demonstrating a rigorous methodology which sets the standard for similar studies. The first experiment involved 500 students, 14 and 15 years old. One half of the students saw colour versions of 5 films, whilst the other half simultaneously saw black and white prints made from the original colour materials. The films were commercially produced titles including: Maps are fun; How man made days; Rivers of the Pacific Slope; Snakes, and Sulphur & its compounds. Two types of tests of perceptual and conceptual learning were developed for each film: non-verbal and verbal. The results for the verbal tests indicate that in only one case was there a statistically significant result in favour of the colour film version. The non-verbal test results reverse the statistically significant results for the verbal tests, with two of the three films favouring the b & w version. However, the differences do not persist and the delayed recall test indicates no difference between the two versions.

The main conclusion reached by Vander Meer was:

> 1. The use of colour in instructional films which may superficially seem to 'call for colour' does not appear to be justified in terms of greater learning on the part of those who view the films. If colour is to be used effectively in films there must be careful preproduction consideration of the probable psychological impact of specific uses of colour upon the learner.

A similar project was undertaken at Yale University (May and Lumsdaine, 1958). In the "Learning from Films" report the effects of pictorial quality and colour is considered, especially the importance of factors generally regarded as entering into the degree of polish or quality of the pictorial component of teaching films, both factors being related to the cost of producing and printing films.

The Yale team produced a colour film 'Seasons', which dealt with the causes of seasonal change, and was to be used to investigate the efficacy of colour instructional films. During the production phase a story board was produced as a guide for the eventual production of animated and live colour footage. The story board consisted of very crude b & w pencil sketches for each scene. In order to aid in visualising the content of the final film a so-called pencil test running reel was made by photographing these sketches on motion picture film in the planned sequence.

Before the final film became available a silent print of the pencil test version was shown to a sixth grade class, with a staff member reading the commentary. The post test scores were later compared with the performance of a similar class who viewed the full colour version and the result was that the learning from the crude, jerky b & w version was substantially as great as from the full colour version. These surprising

results were at first only accepted as being very tentative, because the groups compared were not selected to assure equivalence or provide any valid measure of error. A second experiment utilised 4 classes of fifth grade pupils, and Lumsdaine concluded that the difference between the two mean scores was so small that it is interpreted as the result of mere chance fluctuations — in other words they are not statistically significant.

Kanner and Rosenstein in 1960 evaluated the need for colour rather than b & w instructional television in the US Army. The report indicates that reliable colour television equipment was available but that the cost was higher than monochrome equipment although costs were expected to fall with technological developments as indeed they have done.

The experimental study beamed the eleven lessons from a mobile colour television facility into two classrooms, one viewed the lessons on colour receivers while the other group viewed them on monochrome receivers. Immediately following a lesson the subjects were tested using multiple-choice questions. Every effort was made to incorporate colour items into the tests. A total of 368 trainees took part in the experiment and pairs of subjects were matched on electronics aptitude or general technical scores and then were randomly assigned to one of the two experimental conditions.

Ten out of 11 comparisons show no significant differences and the single statistically significant result is considered to be unimportant in view of the overall picture and small differences in test performance. The overall mean scores are remarkably similar, bearing out the results of May and Lumsdaine.

Stephen Cox produced a survey of the research into the effects of colour in learning from film and television in 1976, and from the results of twenty or so studies concluded that overall there is no marked difference in learning from colour or black and white film or television.

Dwyer took this research forward in an attempt to improve visualised instruction. He planned and carried out his programme of systematic evaluation of the effects of a variety of pictorial types over an extended period of time, involving 100 separate studies with a total population of 23,000 students.

The first piece of research in this programme is recorded in Dwyer's 1967 research report: Adapting Visual Illustrations for Effective Learning, published in the Harvard Educational Review. He describes an experiment which compares four different audio-visual presentations of the same material. The commentary is the same for each presentation but the pictorial image is different.

Initially, only black and white illustrations were considered. The four conditions used in the experiment were:

1   Oral/verbal presentation. The students in this group saw no accompanying illustrations, but the name of the parts of the heart mentioned were projected on the screen.

2   Abstract linear representations, later called simple line drawing: the line drawings used in this presentation were similar to instructional drawings used in many current science books.

3   Detailed shaded drawing presentation: the illustrations were more complete than the simple line drawings, and they represent the heart and its related parts more realistically rather than merely identifying and locating them as in the simple line drawings.

4   Realistic photographic presentation: photographs of the heart were used.

Dwyer also designed four individual criterial measures, which were administered in the following order: drawing, identification, terminology, comprehension tests. After the presentation of the instructional materials each student was permitted to take as much time as he required to complete one criterial measure before proceeding to the next. The important question, a re-statement of the 'Sign Similarity Hypothesis', was: do students learn more if illustrations are more realistic?

The results of this study indicated that when students viewed their respective instructional presentations for equal amounts of time, the simple line drawing presentation was significantly more effective in facilitating achievement than was the oral presentation without visuals on the drawing, identification, and total criterial tests. The oral presentation without visuals of the heart was found to be as effective as each of the visually complemented treatments on both the terminology and comprehension tests. Dwyer also concluded that, contrary to previously stated theories of visual communication (e.g. Carpenter, 1953), the more realistic illustrations were found to be the least effective in complementing oral instruction.

The purpose of this and other experiments was to test the hypothesis that an increase in realistic detail in visual illustrations increases the probability that learning will occur. Eventually nine slide sequences, possessing differing degrees of realistic detail and colour, were produced so that variations in visual stimuli could be assessed in terms of their ability to facility student achievement on five criterion measures. The results indicated that increasing the amount of realistic detail in visual illustrations does not necessarily lead to greater learning.

WHY ARE THERE NO DIFFERENCES?

There are no differences because much of what is happening in mediated instruction hardly differs from what is happening in the classroom. And the classroom is a most inefficient device for education. The reason colour, motion, or pictorial quality adds little to understanding a range of topics, is that the human information processing system has processing limits and best deals with information that has been simplified. This actually matches better with the types of tests which are administered to measure learning, such as those used by Dwyer: simple line drawings are best for instruction when assessment uses simple line drawings. Travers (1964) successfully linked this to the emerging discipline which was seeking to apply information theory to psychology. He demonstrated that much of the information that is attended to by the sense organs

is actually filtered out before it reaches the higher levels of cognitive processing. In many cases, as exemplified by Dwyer's line drawings, simplification makes the world more comprehensible because it places less demands on the processing system: it is, by its very nature, partly processed, the extraneous information having been stripped out.

There are no differences because the information passed to the student by the teacher, the television or radio programme, book or picture, is not usually sufficiently well-adapted to the student's needs. The information is often too much, in quantity or speed of delivery, and the student perceives only a fraction of it, and understands even less.

This represents the context for the first series of papers submitted for the thesis (Spencer, 1977, 1981a, 1981b, 1981c, 1981d, 1988 Chapter 5). These papers, which will be discussed in greater detail in the conclusion for the thesis, draw together the research available at the time of writing and synthesise a theoretical basis for understanding the failure of the new media to produce significant educational gains in performance.

CAN THERE BE ANY DIFFERENCES?

There must be an unequivocal "yes" to this question. Students who are average in a class can be turned into above-average "A"-grade students, but not by merely changing the medium of instruction. Clark (1983) concluded from the research on learning from media that:

> "Consistent evidence is found for the generalisation that there are no learning benefits to be gained from employing any specific medium to deliver instruction."

"Media in education" is often misconstrued as being what educational technology is all about. In fact, it is only one aspect of educational technology and should be more properly termed: technology IN education. It does not have a good record for showing significant educational gains when introduced into the classroom. However, there is a technology OF education that has been much more successful in raising levels of achievement. Its roots are in the behavioural sciences, and can be traced back to the work of Skinner and Bloom; it is associated with mastery learning methodologies. A review of the relative effect sizes of different media and methods (Spencer, 1991) shows that the Learning for Mastery (LFM) approaches advocated by Bloom (1968) and Keller's (1968) Personalised System of Instruction (PSI) produced educationally significant results when compared to the media approaches.

The Personalised System of Instruction is a totally student-paced system, in which students can take course units at any time, and then receive immediate feedback on their performance on tests taken directly after each unit of study. The feedback is provided by proctors, who are paid students who have already completed the course of study. This approach is very similar to that advocated by Lancaster in the monitorial system. Students can only move on from one unit to the next when they

have achieved a high level of proficiency in the work being studied: a criterion level of at least 80% correct is usual, and Kulik (1986) has shown that raising that criterion level to >90% produces a much larger effect size of 0.8, compared to 0.4 for 70-80%.

Learning for Mastery, Bloom's approach, is teacher-paced and more suitable for use in the classroom. It requires constant formative evaluation of each pupil's performance, with support for failing students being provided by the use of audio-visual aids and other suitable remedial materials, and extra sessions with pupils and adults providing remedial help. Bloom is adamant that the tests must be diagnostic:

> For students who fail to master a given unit, the tests should pinpoint their particular learning difficulties — that is, the specific questions answered incorrectly and thus the particular ideas, skills, and processes which need additional work. We have found that students respond best to diagnostic results when the diagnosis is accompanied by very specific prescription of particular alternative instructional materials and processes they can use to overcome their learning difficulties. (Bloom, 1968)

It is interesting to compare these mastery approaches with Postlethwait's (1972) Audio Tutorial (AT) method. Postlethwait's system, based on the provision of audio-taped lectures, has many of the characteristics of LFM and PSI: it is student-paced within a set time for particular units of work (often 1 week), with students visiting the learning centre whenever they need to study; there is continual formative assessment, usually weekly, with students completing a written and oral test; the course is broken down into small units; and students are provided with a set of learning objectives for each unit. However, there is no mastery requirement and students can move from one unit to the next obtaining very low grades. When compared with traditional approaches, this system shows only a small effect size, and the reason seems to be that no matter how flexible a system is, no matter how carefully instructional materials are prepared for students, if the students do not invest sufficient mental effort in the learning process they will fail to master all the material presented. If later units require full understanding of earlier materials, then those students who have previously achieved only partial understanding will inexorably develop a cumulative learning deficit, and the gap between top and bottom students will widen.

A Synthesis for the Future?

We must begin to accept that what takes place in the classroom can be replaced by a whole host of alternative media, without a deterioration in pupil performance. Research has demonstrated this using different media, teaching subjects and ages of students. If we wish to improve levels of performance, we must look to the new methodologies. Those systems that incorporate mastery learning strategies seem to offer the most hope for such improvements. Bloom and Keller have demonstrated mastery systems based on traditional media, essentially human resources, backed up with written materials, but this mastery methodology can also be automated and applied to computer-based systems.

The second series of papers (Spencer, 1990a, 1990b, 1992, 1996) presented for this thesis evaluate the effectiveness of such computer–based systems which incorporate strategies for boosting performance levels based on the paper which presented the synthesis of previous research (Spencer, 1991). Spencer (1992,1996) demonstrated two such effective systems for teaching biology and reading. Integrated learning systems, which manage student progress by constantly assessing performance and indicating suitable learning materials, are also following this path. At this stage results are variable, but encouraging:

> In the first phase of the evaluation pupils using SuccessMaker made learning gains in numeracy above that of equivalent control groups (an effect size of 0.4 which was equated to progress of 20 months over a six-month period. (P.14, Integrated Learning Systems: A report of phase II of the pilot evaluation of ILS in the UK, NCET, 1996)

Table 1.1   Summary of Effect Sizes (based on Spencer, 1991)

	Effect Size	
Programmed learning (Secondary)	0.08	*
Individualized instruction	0.10	*
Ability groups	0.10	*
Visual-based instruction	0.15	**
Programmed learning (Higher)	0.24	**
Audio-Tutorial	0.25	***
Assigned homework	0.28	***
Computer Assisted Instruction	0.29	***
Tutoring	0.40	****
Integrated Learning Systems	0.40	****
Computer Simulations	0.49	*****
Personalised System of Instruction	0.58	******
Learning for Mastery	0.80	********

* represents the size of effect in tenths of a standard deviation

[An effect size of 1.0 improves an average pupil's performance such that they achieve results previously associated with the most able 10% of the class; an ES of 0.25 and less is considered educationally insignificant]

We must remember that the development of sophisticated computer-based systems is in its infancy, but even so, as Table 1.1 illustrates, it is now beginning to equal the effectiveness of individual tutoring methods. It is not surprising that the computer begins to excel when it is used to provide simulations in which feedback processes are an integral part.

Skinner was the most effective early proselytiser of mechanical methods using feedback mechanisms, demonstrating his teaching machines in 1954. He argued that a

country that could mass-produce washing machines and cars could surely develop a machine for providing sufficient feedback to students to enable them to reach high levels of performance, especially in basic numeracy and literacy. Programmed learning was only moderately successful. Later computerised methods, using essentially the same limited psychology, have shown themselves to be at least as effective as teachers, and in some cases even more effective.

However, the newer integrated learning systems, combining mastery strategies with the ability to provide rapid feedback and make decisions about suitable remedial materials, represent the true state of Educational Technology. It is these technologies, combining technology of education with technology in education, that will have the greatest impact on the basic skills (literacy and numeracy) that Skinner was motivated to try and improve.

Solving the literacy problem: writing technologies

The UK has large numbers of adults who fall into the lowest 2 categories on a range of international literacy measures, and the final series of papers in this thesis investigate contributory factors which influence literacy levels in UK schools, with a view to producing a new theoretical perspective for the development of mastery-based literacy materials. This perspective is described in the concluding section of the thesis.

The UK literacy rate is similar to other English-speaking countries, and the OECD (1997) concluded:

> Just over half of the UK's adults show relatively low levels of literacy (Levels 1 or 2), a similar proportion to the United States and New Zealand. Within the large group of people with literacy weaknesses, which is common to most countries, the UK has a particularly large number of very low literate adults: over 20 per cent only reach Level 1. But the UK also has high numbers of adults scoring at the top Levels, 4 and 5. For example 17 per cent score at this level in prose literacy compared to 13 per cent in Germany. In quantitative literacy, on the other hand, the proportion of Germans (24%) with high literacy is greater than in the UK (19%). In the United Kingdom, the influence of family background on literacy, over and above the influence of education, appears to be stronger than in most other countries. Among adults who have completed secondary school, 75 per cent of those whose fathers were educated to this level are at literacy Level 3 or above, compared to 59 per cent of those with less educated fathers. This gap is wider than in most other countries in the survey. The UK labour market pays a high premium for literacy skills, more than in other countries except Ireland and the United States. (OECD, 1997)

At prose level 1 most of the tasks require the reader simply to locate and match a single piece of information in the text that is identical to or synonymous with the information given in the question. More than 20% of the adult population in the UK can only function at this level when dealing with prose. Level 2 tasks require the reader to locate one or more pieces of information in the text; however, several distracters may be present and the reader may need to make low-level inferences. Tasks at this level also begin to ask readers to integrate two or more pieces of information, or to compare and contrast information. Fifty percent of UK adults can only operate at these low literacy levels.

Prose level 3 tasks tend to direct readers to search the text to match information, requiring the reader to make low-level inferences or to locate text that meets specified conditions. Sometimes the reader is required to identify several pieces of information that are located in different sentences or paragraphs rather than search for information located in a single sentence. Readers may also be asked to integrate or to compare and contrast information across paragraphs or sections of text. A little over 30% of UK adults operate at this level.

The higher levels, 4 and 5, require readers to perform multiple-feature matching or to provide several responses where the requested information must be identified through text-based inferences. Tasks at these levels may also require the reader to integrate or contrast pieces of information that are sometimes presented in relatively lengthy texts. Some tasks require readers to make high-level inferences or use specialised knowledge. The UK results show 17% of adults operating at this level and this compares favourably with most other countries. This in many ways is a worrying factor when considering the distribution of adult literacy skills in the English-speaking world: typically, English-speaking countries have a high proportion, about 50%, of the population operating at the two lowest levels. Table 1.1 demonstrates this by comparing the UK and US figures with those for Sweden, which has a much greater proportion of the population operating at the higher levels, and comparatively small numbers possessing the limited skills at levels 1 and 2.

Sweden actually has only 7% of adults in the lowest category, demonstrating that high levels of functional illiteracy need not be tolerated. The UK is, however, in the same league as other countries, such as the United States and New Zealand, that have English as their common language.

There is increasing evidence which supports the view that part of the failure of English-speaking countries to sustain acceptable levels of literacy can be attributed to the written form of the language. Paulesu et al (2000) indicate that there are 1120 ways of representing the 40 sounds (phonemes) in English by different letters or letter combinations. Whereas, Italian has only 35 graphemes to represent its 25 phonemes.
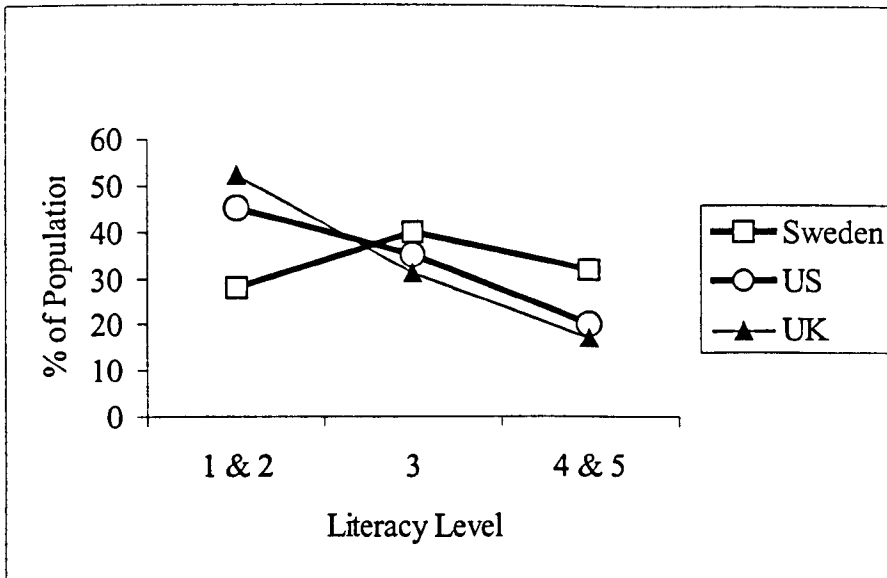
Table1.2   Adult Literacy Levels for UK, US and Sweden (Source: Organization for Economic Co-operation and Development and Statistics Canada, Literacy, Economy and Society, Results of the International Adult Literacy Survey, 1995.)

McGuinness (1998) has suggested that there 'is no diagnosis and no evidence for any special type of reading disorder like dyslexia', and that teaching techniques which do not take account of the basic codification of English are the main source of literacy problems in the English-speaking world. When techniques based on an understanding of the code are employed, with a wide range of failing subjects, McGuinness (1996) demonstrates striking improvements within comparatively short interventions.

Oney and Goldman (1984) demonstrated the superiority of a coherent systematic alphabetic system over the English code when they compared the performance of first and third grade Turkish and American pupils. Turkish is an almost perfectly transparent orthography, having been reformed in 1928. This has been confirmed by Landerl et al (1997) and Frith et al (1998) in comparisons with the more regular German orthography.

> Clearly, the success of phonological recoding depends on whether the word to be identified conforms to the code. In the case of irregular words, which are frequent in English, phonological recoding, if strictly applied, leads by definition to an incorrect word sound... Learning to read and write in a consistent orthography should not lead to such difficulties. Here, grapheme-phoneme recoding is reliable, and the assembly of phonemes results in pronunciations close to that of the target word. (Frith et al, 1998, p.32)

Landerl's results actually demonstrate that for low-frequency words German dyslexic children out-perform normal English children.

The second series of papers in this thesis, based on the theoretical synthesis, demonstrate the successful application of mastery learning procedures. In particular, Spencer (1996) illustrated how such techniques could rapidly ameliorate literacy deficits built up over many years. This paper considered a restricted set of words (the most common 120 words in the English language), but what of the wider literacy concerns beyond this set? If techniques are to be used to raise overall literacy performance then an understanding of the factors which contribute to failure (such as orthography) must be explored in order to provide a sound underlying structure for teaching programmes. The third series of papers in this thesis (Spencer, 1998, 1999a, 1999b, 2000) looks at the complexity of written English and attempts to identify factors which contribute to the low levels of literacy, especially spelling, in the UK (and other English-speaking countries). Having identified contributory factors in these papers, the concluding section of this thesis will present new data which indicates a possible structure for mastery-based reading/spelling programmes.

2 The Papers

1. Spencer, K.A. (1977) An evaluation of the cost-effectiveness of audiotaped and videotaped self-study media presentations, British Journal of Educational Technology, 8(1), 71-77.

2. Spencer, K.A. (1981a) An evaluation of the cost-effectiveness of film and tape-slide instructional presentations for the mentally handicapped, Programmed Learning and Educational Technology, 18(1), 23-9.

3. Spencer, K.A. and Clarke, A. (1981b) A BBC television series for the mentally handicapped, Educational Broadcasting International, 14(3), 110-113.

4. Spencer, K.A. (1981c) Alternative media: a review of the instructional effectiveness, Educational Broadcasting International, 14(2), 90-93.

5. Spencer, K.A. (1981d) AV communication: small is beautiful, Audiovisual Librarian, 7(1), 22-28.

6. Spencer, K.A. (1988) The Psychology of Educational Technology and Instructional Media, pp198, London: Routledge.

7. Spencer, K.A. (1990a) HyperCard: teaching technology for successful learning, Journal of Audiovisual Media in Medicine, 13, 25-30.

8. Spencer, K.A. (1990b) HyperHeart: does animated illustration contribute to mastery learning?, British Journal of Educational Technology, 21(3), 227-8.

9. Spencer, K.A. (1991) Modes, media and methods: the search for educational effectiveness, British Journal of Educational Technology, 22(1), 12-22.

10. Spencer, K.A. (1992) Information technology and human information processing: a critical perspective, Proceedings of the European Conference about Information Technology in Education, Barcelona, November, 512-522.

11. Spencer, K.A. (1996) Recovering reading using computer mastery programmes, British Journal of Educational Technology, 27(3), 191-203.

12. Spencer, K.A. (1998) Predictive models of spelling behaviour in 7- and 11-year-olds, Journal of the Simplified Spelling Society, 24(2), 23-27.

13. Spencer, K.A. (1999a) Predicting spelling behaviour in 7 to 11 year-olds, Reading, July, 72-77.

14. Spencer, K.A. (1999b) Predicting word-spelling difficulty in 7- to 11-year-olds, Journal of Research in Reading, 22(3) 283-292.

15. Spencer, K.A. (2000) Is English a dyslexic language? Orthographic transparency and multilingual teaching, Dyslexia, 6, 152-162.

1. Spencer, K.A. (1977) An evaluation of the cost-effectiveness of audiotaped and videotaped self-study media presentations, British Journal of Educational Technology, 8(1), 71-77.

An Evaluation of the Cost-Effectiveness of Audiotaped and Videotaped Self-Study Media Presentations

Kenneth Spencer

Kenneth Spencer was, at the time of writing, Lecturer in Media Development at the Audio-Visual Centre, University of Hull. He is at present Lecturer in the Department of Educational Studies, University of Hull.

Abstract

Two self-study media presentations, booklet-tape and videotape recording (two formats) were evaluated in order to determine the relative cost-effectiveness of the presentations. Three topics from an elementary electronics theory course were prepared in three different versions, and relevant costs were calculated. The results of the study indicated that the differences in mean scores for all comparisons did not justify the greatly increased costs associated with videotape recording.

Introduction

Models for making instructional decisions, with particular reference to media selections, have been proposed by Gerlach (1966), Briggs, Campeau, Gagné and May (1967) and Tosti and Ball (1969). Such models are useful, though it is recognized that most objectives may be attained through instruction presented by any variety of different media, as a great many studies have shown no significant differences between one medium and another in facilitating the attainment of a wide range of objectives. Tosti and Ball indicate that, using their framework, there is often no one best medium, and that several operational systems may convey equivalent presentational designs within the constraints specified. In such cases they suggest that the final selection between the systems should be based on external considerations such as cost, availability of media, and user preference.

The purpose of this study was to investigate the relative effectiveness of two self-study media presentations, booklet-tape and videotape (two presentation formats) in relation to the cost of production and transmission. There are few studies which have compared audiotape and videotape presentations. Most studies have compared either audiotape or videotape with conventional class lectures. Popham (1961, 1962) used audiotaped lectures in both graduate and undergraduate classes and found no significant differences in achievement between the taped and live lecture presentations. Similarly, Menne, Hannum, Klingensmith and Nord (1969) found no significant differences between audiotaped and live lectures in an introductory college course.

Comparisons between videotaped and face-to-face, live instruction have also generally shown no significant differences (Allen, 1971; Chu and Schramm, 1967; Dubin and Taveggia, 1968). These results suggest that as there were no differences between live lectures and either videotaped or audiotaped lectures, there will be no differences for comparisons between videotaped and audiotaped lectures.

MacDonald (1972) offers evidence which confirms this. Videotape recordings were compared with audiotapes plus relevant slides, to determine whether the additional cost of videotape was justified. The two groups did not differ significantly from each other on an information test, and it was concluded that the audiotape version was

more cost-effective, because the transmission equipment cost for the audiotape plus slides was about one half that for videotape.

Materials
Topics
Three different topics were chosen from a first-year course in elementary electronics theory, for electronic engineering and applied physics students. The topics, dealing with basic theory, were:
(i) four terminal networks
(ii) the amplification properties of a junction transistor
(iii) the load line.

Production
Three versions of each topic were prepared, using the facilities available in the Audio-Visual Centre at Hull University.

(i) Booklet-tape. This consisted of printed sheets, containing all relevant equations and diagrams, and an accompanying audiotape with a recorded commentary relating to the contents of the booklet.
(ii) Videotape recordings. Two formats were used for each topic, high-cost VTR and low-cost VTR. Both formats were recorded on 1in videotape and these master recordings were transferred to ½in copies for self-study.
(iia) Low-cost VTR. The lecturer delivered a spontaneous lecture, guided by outline notes. Two cameras were used, one directed to the lecturer and the other towards a large pad on which the lecturer wrote relevant equations and diagrams.
(iib) High-cost VTR. The lecturer read from a fully prepared script. Three cameras were used, with one directed to the lecturer, and two available for captions. The captions varied in their complexity, ranging from straightforward Letraset on card to complicated captions involving moving parts and critical superimposition from up to three separate sources.

Costs
It has been suggested by Dunn (1971) that the unit for comparison, the unit cost, should be the cost per student-hour, and that evaluations of instructional strategies should all be on this basis, so that comparisons of unit costs can be made. A blueprint for cost-effectiveness studies in educational technology has been proposed (Cumming and Dunn, 1970) which corresponds closely with the analysis of media costs presented by Zeckhauser (1970). This blueprint has been used in this study as the basis for the cost evaluation (Table 1).

As the system in this study was to provide supplementary material, rather than act as a lecture course substitute, the figures suggested by Zeckhauser (1970) for hardware and software have not been adhered to. Instead, operating with a total of 20 programme hours it was assumed that three transmission machines with four self-study copies of each programme would allow a total of 30 students reasonable access time. Capital equipment costs have been written off over a period of five years, and programme costs over a period of four years.

The estimates for the production of each master copy included figures for the time involved for academic and technical staff, capital equipment, consumable supplies,

etc. However, estimates of the capital cost of the building, heating, etc, were not included.

Table 1. Estimated costs per student-hour

	booklet-tape	low-cost VTR	high-cost VTR
transmission			
3 machines plus maintenance	£55	£625	£625
cost/annum	£11	£125	£125
production			
salaries	£18.50	£49.50	£227.50
consumable materials	£4	£13	£20
capital equipment	£3.50	£15	£60
4 self-study copies	£6.50	£40	£40
cost/½-hour programme	£32.50	£117.50	£347.50
*cost/annum/20 programme hours	£325	£1175	£3475
total cost/annum	£336	£1300	£3600
cost/student-hour	£0.56	£2.17	£5.84

* 20 programme hours = 40 ½-hour programmes
 costs of programmes written off over 4 years

therefore cost/annum/20 programme hours $= \dfrac{\text{cost/½-hour programme} \times 40}{4}$

Method
Subjects
The subjects were thirty-three first-year electronics and applied physics students studying elementary electronics theory.

Evaluation instruments
In order to assess the learning following exposure to each topic a multiple-choice test was prepared. Each test consisted of ten questions, each with four alternatives. The questions were designed to assess the theory and application of information presented, and care was taken to ensure that the relevant sections in each topic were covered equally well in each version, following the procedures outlined by MacIntosh and Morrison (1969). Subjects were instructed to answer all questions, with one response per question. A guessing correction was not applied to the test scores.

A questionnaire, which required students to rate their preferences for the different presentations, was also prepared.

Procedure
Subjects were randomly assigned to one of three treatment groups and the following

plan, exposing each subject to a different media version for each topic, was adopted:

TOPIC	GROUP A	GROUP B	GROUP C
four terminal networks	booklet-tape	low-cost VTR	high-cost VTR
amplification properties of a junction transistor	high-cost VTR	booklet-tape	low-cost VTR
load line	low-cost VTR	high-cost VTR	booklet-tape

The learning material was presented for revision seven days after a class lecture. The topics were presented one per week, for three consecutive weeks.

Although a control condition for subjects was desirable the small number of students available precluded such a condition, the procedure concentrating on the comparison of different media conditions.

A total of nine subjects could study each topic during a single session, using six videotape carrels and three booklet-tape carrels. Each presentation lasted approximately 25 minutes, and subjects were required to study the material for a total of 50 minutes. Subjects were encouraged to take full advantage of the medium used, especially the use of still frames for note-taking with the videotape recordings, and the making of supplementary notes on the printed sheets provided with the booklet-tape.

Following the instruction period, subjects completed the ten-question multiple-choice test, a period of 20 minutes being allowed. The questionnaire was administered following the final test.

A pre-test would have given some indication of the amount of learning following the class lecture. However, a post-test only was administered because of difficulties in preparing an equivalent form pre-test and also the possible effects of administering the same test as both pre- and post-test.

Results

The mean scores and standard deviations for each topic are given in Table 2. To examine whether differences of statistical significance were obtained a standard 3×3 Latin Square analysis of variance was carried out. The raw data was binomial and could be expressed as a proportion of the total possible score. It was therefore transformed in order to stabilize the variances, using the transformation

$$Y = 2 \sin^{-1} \sqrt{X}$$

where X is the observed proportion. In the case where $X = 1$ (ie the raw score is 10) the following transformation was used

$$Y = 2 \sin^{-1} \sqrt{X - \tfrac{1}{4} \cdot n}$$

where $n = 10$.

The results of the calculations in Table 3 show that there were no statistically significant differences.

The mean scores of the tests for the three versions of the topics are presented in Figure 1 in relation to their relative cost per student-hour (expressed as a percentage of the high-cost VTR).

Table 2. Mean scores and standard deviations

	booklet-tape	low-cost VTR	high-cost VTR
four terminal networks mean score SD	8.5 1.2	8.1 1.0	8.4 1.3
amplification properties of a junction transistor mean score SD	7.9 1.9	8.4 1.0	8.4 1.3
load line mean score SD	7.7 1.3	8.1 0.9	8.8 0.7

Table 3. Standard 3 × 3 Latin Square analysis of variance

source of variation	df	SS	MS	F	P
between methods	2	0.0380	0.0190	0.0820	N.S.
between topics	2	0.1607	0.0804	0.3464	N.S.
between groups	2	0.0042	0.0021	0.009	N.S.
residual	2	0.1083	0.0542	0.234	N.S.
within cells	90	20.8901	0.2321		
total	98	21.2013			
$F_{2\,90}\,(p = 0.05) = 3.10$					

Figure 1: Percentized mean scores and estimated costs per student-hour
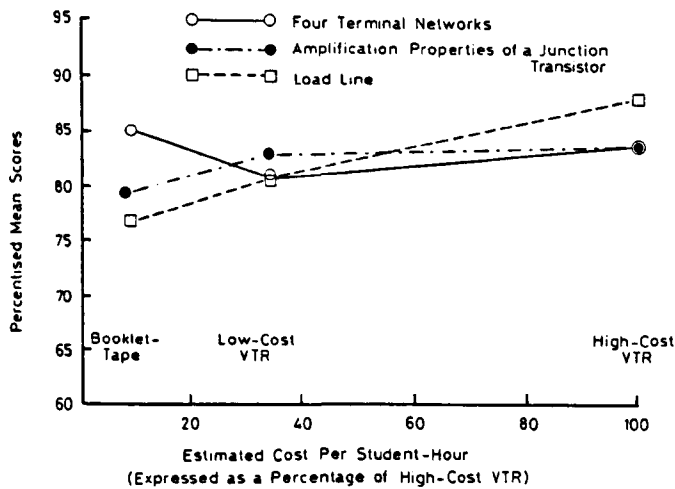
The questionnaire indicated that a majority of the subjects (59%) preferred the high-cost VTR, compared with low-cost VTR (26%) and booklet-tape (15%). However, although the booklet-tape was *preferred* by a comparatively small number of students, a high percentage (65%) found the booklet-tape suitable for the presentation of the concepts involved in the topics chosen and the same proportion also felt that booklet-tape had definite advantages over the more traditional printed material used to supplement lectures.

Discussion

The results of this study tend to confirm the general 'no significant difference' results in other media comparison studies. Clearly the additional costs of either high- or low-cost VTR presentation is not justified when compared with the booklet-tape presentation.

Twyford, Davis and Seitz (1954) presented much the same argument following the evaluation of an expensive film treatment and a much less costly filmograph treatment. It was suggested that it was questionable whether the more elaborate and expensive production justified the small difference (two points) in informational learning. They suggested that, as it was not possible to demonstrate any great differences in favour of the more elaborate and expensive treatments, a point of diminishing returns in effectiveness is reached, beyond which substantial increases in cost result in only slight increases in training effectiveness. This suggestion appears to be substantiated by the results of the present study.

A further consideration is the expressed preference for the VTR presentation in general and the high-cost VTR in particular. These results are similar to those reported in other studies, in which subjects expressed a preference for such attributes as motion (Miller, 1969) and colour (Vandermeer, 1952). These studies indicate that the strong preferences expressed do not appear to influence learning. While subject preferences are important in deciding instructional strategies, they would be more relevant if they were associated with definite increases in learning.

This study has looked at only one aspect of cost-effective media comparisons, with the materials being restricted as back-ups to class lectures. There are still more comparisons to be made, having eliminated VTR presentations: eg comparisons between booklet-tape, printed hand-outs, revision from students' notes, a repeat lecture and a no further instruction condition. However, even if the last condition proved to be equally effective the cost saving would still be small when compared to that following the elimination of VTR presentations for such topics.

Wilkinson (1973) had suggested that with the long history of no significant difference between instructional strategies, the cost-effectiveness question is often reduced to finding the cheapest method of presentation, a generalization supported by this study. However, this will not always be the case and further research is necessary to investigate the complex interaction between subject, medium and topic.

Acknowledgement

(Department of Mathematical Statistics) for his suggestions concerning the analysis of the data.

References

ALLEN, W H, 'Instructional media research. Past, present and future', *Audio Visual Communication Review*, 19, pp 5–18, 1971

BRIGGS, L J, CAMPEAU, P L, GAGNÉ, R M and MAY, M A, *Instructional media: a procedure for the design of multimedia instruction, a critical review of research and suggestions for future research*, American Institutes for Research, Pittsburgh, ED 024 278, 1967

CHU, G C and SCHRAMM, W, *Learning from television: what the research says*, Stanford University, Institute for Communication Research California, ERIC ED 014 900, 1967

CUMMING, C and DUNN, W R, 'The application of cost-effectiveness techniques to educational technology', in A C Bajpai and J F Leedham (eds), *Aspects of Educational Technology IV*, Pitman, London, 1970

DUBIN, R and TAVEGGIA, T C, *The teaching–learning paradox*. University of Oregon Press, Eugene, 1968

DUNN, W R, 'Costs: comparison with conventional instruction', in D Packham, A Clearly and T Mayes (eds), *Aspects of Educational Technology V*, Pitman, London, 1971

GERLACH, V S, 'Selecting an instructional medium', in W C Meierhenry (ed), *Media competencies for teachers*, Teachers College, University of Nebraska, Lincoln, 1966

KALKOFEN, H, 'Instructional film research: some notes on methodology', *Educational Media International*, 2, pp 2–6, 1972

MACDONALD, R, *A comparison of two methods of presenting occupational information*, Appalachia Education Lab, Charleston, West Virginia, ERIC ED 062 574, 1972

MACINTOSH, M A and MORRISON, R B, *Objective testing*, University of London Press, London, 1969

MENNE, J W, HANNUM, T E, KLINGENSMITH, J E and NORD, D, 'Use of taped lectures to replace class attendance', *Audio Visual Communication Review*, 17, pp 42–46, 1969

MILLER, W C, 'Film movement and affective response and the effect on learning and attitude formation', *Audio Visual Communication Review*, 17, pp 172–181, 1969

POPHAM, J W, 'Tape recorded lectures in the college classroom', *Audio Visual Communication Review*, 9, pp 109–118, 1961

POPHAM, J W, 'Tape recorded lectures in the college classroom II', *Audio Visual Communication Review*, 10, pp 94–101, 1962

TOSTI, D T and BALL, J R, 'A behavioural approach to instructional design and media selection', *Audio Visual Communication Review*, 17, pp 5–25, 1969

TWYFORD, L C, DAVIS, J S and SEITZ, C P, *Evaluation of the film: 'Military Police support in emergencies'*, Technical Report No SDC 269-7-52, Special Devices Centre, Port Washington, New York, 1954

VANDERMEER, A W, *Colour vs black and white in instructional films*, Technical Report No SDC 269-7-28, Special Devices Centre, Port Washington, New York, 1952

WILKINSON, G L, 'Cost evaluation of instructional strategies', *Audio Visual Communication Review*, 21, pp 11–30, 1973

WINDER, B J, *Statistical principles in experimental design*, (2nd ed), McGraw-Hill, London, 1971

ZECKHAUSER, S H, *A look at media in higher education*, Harvard University, Cambridge, Massachusetts, 1970

2. Spencer, K.A. (1981a) An evaluation of the cost-effectiveness of film and tape-slide instructional presentations for the mentally handicapped, Programmed Learning and Educational Technology, 18(1), 23-9.

An Evaluation of the Cost-effectiveness of Film and Tape-slide Instructional Presentations for the Mentally Handicapped

Kenneth A. Spencer *Department of Educational Studies, University of Hull*

Abstract: Two media presentations, tape-slide and film, were evaluated in order to determine the relative cost-effectiveness when used as automated teaching aids for the mentally handicapped. The results indicated that significant learning and retention occurred following repeated presentation of the materials, but that no differences were observed between the two methods of instruction. It was concluded that tape-slide was more cost-effective than film.

Introduction

An audiovisual approach to the teaching of the mentally handicapped offers considerable advantages to both teacher and pupil, particularly for such topics as social skills training. A major advantage is the opportunity to bring the outside world into the classroom, where it can be studied without the distractions and difficulties experienced in the real-life situation.

There are, however, several problems associated with the selection of suitable audiovisual aids which are of especial concern for the instructor of the mentally handicapped. One major problem stems from the opposing views of the realist and reductionist theorists. The realism proponents, as exemplified by Dale (1946) and Finn (1953), claim that the more realistic a presentation, the more effective will be the transmission of the desired message. The reductionist view, as expressed by Miller *et al.* (1957) and Travers *et al.* (1967), is that it would be a mistake to assume that one cue added to another would increase learning by a linear increment, and they contend that additional cues or excessively realistic cues may be distracting or posssibly even evoke competitive responses in opposition to the desired learning. The reductionist view has received support from Dwyer (1972, 1978) during an extensive research project involving thousands of normal subjects. Porotskaya (1967) has also tended to support this view for mentally handicapped subjects, concluding that the contents of silent films are easier to assimilate by comparison with slides, which are in turn better than sound pictures.

Another factor which has a bearing on this question is the relative cost of the instructional materials. In general, the more realistic aids are more expensive. However, research suggests that the more realistic aid is usually no more effective than the less expensive, less realistic form. A good example of this can be found in the literature which deals with comparisons between motion films and still pictures. The results of many studies (e.g. James, 1924; Vernon, 1946; Carson, 1948; Twyford *et al.*, 1954) indicate that the more realistic, expensive film versions are no more effective than the slide or filmstrip versions. Clearly, in the present climate of financial stringency, cost-effectiveness criteria demand that the more expensive presentation must be more effective.

The implications of such results, if confirmed with mentally handicapped subjects, will necessarily impinge on media selection strategies of instructors. Driscoll (1968) has demonstrated that factual learning takes place at a rate which merits the use of film as a teaching tool with the mentally handicapped. The present research investigates the relative effectiveness of film and the less expensive, less realistic (lack of motion) tape-slide medium.

Materials

'Tundra Animals' tape-slide

A short tape-slide was prepared to enable the selection of matched pairs of subjects for the two experimental conditions. The programme lasted five minutes and consisted of ten 35mm slides of tundra animals, each accompanied by 30 seconds of commentary. Synchronization between picture and sound was obtained using a Phillips N2209 AV recorder and LFD 3442 synchronizer, in conjunction with a Kodak Carousel 'S' projector.

'Zoo Animals' film

A 16mm colour sound film lasting nine minutes was prepared for presentation to the film group. The film was divided into nine sections, each lasting one minute and dealing with one animal. The associated commentary was recorded on a magnetic stripe. A Bell and Howell 658Q projector was used to project the film.

'Zoo Animals' tape-slide

Sixty-six selected frames from the 16mm film were enlarged using a commercial liquid-gate printer. This process produced excellent half-frame transparencies, which formed the basis for the tape-slide presentation. The 16mm film soundtrack was transferred to audiocassette and synchronization between picture and sound was accomplished using the tape-slide equipment described above. This version of the 'Zoo Animals' material accurately reflected the film version, but with the absence of pictorial motion.

Costs

An outline for costing media has been presented by Zeckhauser (1970). Using the suggested cost per student per hour, Spencer (1976) calculated the comparative costs for various media which have been used by institutions for the production of audiovisual materials for the mentally handicapped. The costs, given in Table 1, are based on the assumption that the institution produces a total of ten hours of instructional material, and that pupils view each programme a minimum of five times. Programmes can be watched either by groups (in-class) or by individuals (individual study) and costs are given for both conditions. Full account has been taken of all aspects of the production of the materials, including salaries, consumable materials and recording and transmission equipment.

Evaluation instruments

1. 'Tundra Animals' identification test: subjects were shown the transparencies of the ten animals in random order. Following the presentation of each transparency the subject was asked to name the animal shown on the screen. The total identification score for each subject represented the number of animals correctly named.

2. 'Zoo Animals' identification test: the test was similar to the 'Tundra Animals' identification test. The total score represented the number of animals correctly named. The slides were full-frame 35mm slides of the nine animals taken at the time of the 16mm filming.

3. 'Zoo Animals' information test: the experimenter explained to the subject that he was to try and remember all he could about the animals he had seen on the film. An audio recording was made of each test session.

The subject was asked: 'What can you tell me about the ... (animal name)?' The names were presented in the same order during each test session, and the order was randomized between days. A period of five seconds was allowed and if no reply was given the question was repeated. Following each response the subject was asked if he could remember anything else about the particular animal. If no further response was given during the next ten seconds the question was

	*Cost/student/hour (£)	
Medium	In-class	Individual study
Tape-slide	0.74	1.94
Super-8mm film	1.37	3.66
½" videotape	1.89	4.16
16mm film	1.82	—

*costs based on 40 students/annum

Table 1. Media costs per student per hour

Method

Subjects

Forty subjects were instructed using the 'Tundra Animals' tape-slide. The group was heterogeneous in terms of intellectual abilities, with IQ scores ranging from 47 to 85 and chronological ages ranging from 18 years to 57 years. The group consisted of 22 women and 18 men.

repeated with another animal name.

The recorded responses were analysed according to 80 specific categories derived from the commentary. Additional categories based on purely visually presented information were also devised. However, these additional categories were not required because all the responses given by subjects could be coded into the categories derived from the commentary (see Appendix).

Procedure

1. Pre-experimental
To facilitate the experimental procedure the subjects were randomly divided into two groups. Each group received an identical presentation of the 'Tundra Animals' tape-slide. The presentation took place in a blacked-out room with the transparencies projected to fill the width of the 5' x 5' screen. The sound was presented through an amplified loudspeaker and seating was arranged so that all subjects had an unrestricted view of the screen and received the sound at an acceptable level, without distortion.

The experimenter explained to the subjects that they would see pictures of animals on the screen and would hear what the animals were called and how they lived. It was also explained that later in the day they would be asked questions about what they had seen. Following the first presentation the programme was reset and shown for a second time. This procedure was based on previous evidence which suggested that subjects benefit from an immediate second showing of such programmes (Hoban and Van Ormer, 1950).

In the afternoon following the presentation of the tape-slide material, subjects were tested using the 'Tundra Animals' identification test.

The procedure was repeated the following day.

From the results of the identification test on the two consecutive days it was possible to select 13 pairs of subjects showing matched score profiles. Subjects who did not learn during the two pre-experimental sessions were discarded, as were subjects whose profiles could not be matched. Table 2 provides details

assigned at random either to the film or to the tape-slide group.

Pre-experimental scores were obtained for the 'Zoo Animals' identification test and, as expected, these were very low and permitted a good range of experimental scores. The pre-experimental identification scores are also given in Table 2.

Each period of instruction on the topic 'Zoo Animals' was followed by an information and identification test session. The information test was administered prior to the identification test in order to preclude information being relearned during the identification test.

The 13 subjects in the group receiving instruction were admitted to a blacked-out projection room and were told that they would see a *film* about some more animals and that they would be asked some more questions about the animals later in the day. Both tape-slide and film groups were told that they would see a film to ensure that both groups thought that they were receiving the same treatment. Tape-slide subjects accepted the presentation as being a film, despite the lack of motion.

Following this introduction the experimenter checked that all subjects had a clear view of the screen before showing the programme. The projection for both presentations was arranged so that the visual images were the same size and filled the width of the 5' x 5' screen. Following the first showing of the programme it was reset during a three-minute interval and then shown again. Thus the subjects saw each programme twice during each of the four instructional sessions.

	Tape-slide group	Film group
'Tundra Animals': identification test A, mean score S.D.	2.09 1.81	2.36 2.33
'Tundra Animals': identification test B, mean score S.D.	3.36 2.20	3.64 2.58
'Zoo Animals': pre-experimental identification test 1, mean score S.D.	0.64 1.12	0.91 1.44
IQ mean S.D.	59.54 6.93	57.18 9.75

Table 2. Mean scores for 'Tundra Animals' identification tests 1 and 2, 'Zoo Animals' pre-experimental identification test, and IQ

of the performance of the two groups at this stage.

2. Experimental
To create the two experimental groups subjects were

The information and identification tests were administered during the afternoon following the morning period of instruction.

Previous pilot experiments had indicated that mentally handicapped subjects respond well to repetitions of such instructional materials and that incremental learning can be demonstrated for each repetition. In this study the instructional materials were shown twice per session for four sessions, a total of eight viewings of the programme.

Six weeks following the final session, subjects were again tested, thus providing a single measure of retention.

The experiment spanned a total of 59 days and it was realized that some subject drop-out was inevitable. As the groups had been formed on the basis of matched pairs, subject drop-out from one group was followed by the removal of the remaining member of the pair in the other group.

instruction and this is confirmed in Tables 6 and 7, which show highly significant differences within the test sessions, for both identification and information tests. A significant interaction is also observed between the method of instruction and test session for the information test, reflecting the intially low rate of learning for the tape-slide group.

Discussion

Perhaps the most striking result of this experiment is the effectiveness of either film or tape-slide when used as an instructional medium for the mentally handicapped. Clearly there is little difference between the two media presentations for the identification or information tests, a result which confirms those previously obtained with normal subjects, which run

Method of Instruction	Test	Test session	Day
Tape-slide (Tundra Animals)	Identification (Tundra Animals)	A	1
Tape-slide (Tundra Animals)	Identification (Tundra Animals)	B	2
No instruction	Identification (Zoo animals)	1	7
Tape-slide; film (Zoo Animals)	Identification and information (Zoo Animals)	2	8
Tape-slide; film (Zoo Animals)	Identification and information (Zoo Animals)	3	11
Tape-slide; film (Zoo Animals)	Identification and information (Zoo Animals)	4	14
Tape-slide; film (Zoo Animals)	Identification and information (Zoo Animals)	5	17
No instruction	Identification and information (Zoo Animals)	6	59

Table 3. Experimental Procedure

Table 3 summarizes the experimental procedure.

Results

The mean scores and standard deviations are given in Tables 4 and 5. To examine whether differences of statistical significance were obtained during the period of instruction a two-factor analysis of variance, with repeated measures, was carried out on data from both identification and information tests, Tables 6 and 7. The data obtained from the delayed retention tests were analysed separately and the one-way analysis of variance summary is given in Table 8 for the information test and Table 9 for the identification test.

There is a general lack of statistically significant differences when comparing the two methods of presentation on either test during the period of instruction or following the six-week retention period. There is evidence from Tables 4 and 5 that incremental learning takes place throughout the period of

counter to the 'realism' theories. The results from the present experiment confirm Driscoll's conclusion that factual teaching proved an easy task for the film medium and extends this to demonstrate that such teaching proves to be an equally easy task for tape-slide, but at lower cost.

The results for both tape-slide and film are impressive. There is an increase in mean score from session to session for the information test and a similar result for the identification test. Following the six-week retention period the information scores fall but still remain at a level similar to that following the second period of instruction, confirming the results obtained by Vergason (1966) who also found that retention remained at a relatively high level following automated instructional procedures.

It is apparent that, in the education of mentally handicapped persons, repetition will be particularly important for the presentation of information and concepts, though number of repetitions will vary with

Test session	Tape-slide group			Film group		
	Mean	SD	N	Mean	SD	N
1	0.64	1.12	11	0.91	1.44	11
2	1.73	1.79	11	2.72	2.93	11
3	3.27	2.80	11	3.73	3.10	11
4	4.27	2.83	11	4.45	2.98	11
5	4.27	3.38	11	5.27	2.97	11
6	3.01	3.00	9	4.81	3.22	9

Table 4. Mean scores and standard deviations for identification tests, tape-slide and film groups

Test session	Tape-slide group			Film group		
	Mean	SD	N	Mean	SD	N
2	7.73	7.68	11	10.77	11.71	11
3	15.81	13.92	11	19.00	10.94	11
4	26.18	17.23	11	24.36	11.51	11
5	30.54	21.29	11	31.45	15.40	11
6	16.55	15.89	9	18.33	13.25	9

Table 5. Mean scores and standard deviations for information tests, tape-slide and film groups

Source of variation	SS	df	MS	*F	
Between subjects	584	21			
(A) Method of instruction	9	1	9	0.31	N.S.
Subjects within groups	575	20	28.75		
Within subjects	176	66			
(B) Test session	83	3	27.67	18.07	P<0.01
A B	4	3	1.33	0.09	N.S.
B x subjects within groups	89	60	1.48		

$$*F._{95}(1,20) = 4.35$$
$$F._{99}(3,60) = 4.13$$

Table 6. Summary of analysis of variance — identification test, film v. tape-slide group

Source of variation	SS	df	MS	*F	
Between subjects	16170	21			
(A) Method of instruction	38	1	38	0.047	N.S.
Subjects within groups	16132	20	806.8		
Within subjects	6186	66			
(B) Test session	5928	3	1976	705.7	P<0.01
A B	90	3	30	10.7	P<0.01
B x subjects within groups	168	60	2.8		

*F.$_{95}$ (1, 20) = 4.35
F.$_{99}$ (3, 60) = 4.13

Table 7. Summary of analysis of variance — information test, film v. tape-slide group

Source of variation	SS	df	MS	* F	
Method of instruction	14	1	14	0.665	N.S.
Experimental error	3426	16	214		
Total	3440	17			

*F.$_{95}$ (1, 16) = 4.49

Table 8. Summary of analysis of variance — information test, six weeks retention, film v. tape-slide group

Source of variation	SS	df	MS	* F	
Method of instruction	14	1	14	1.43	N.S.
Experimental error	156	16	9.75		
Total	170	17			

*F.$_{95}$ (1, 16) = 4.49

Table 9. Summary of analysis of variance — identification test, six weeks retention, film v. tape-slide group

the type and complexity of the materials to be learned. Such repetition could prove to be boring and tedious for the teacher and the application of automated techniques can clearly be advantageous in such circumstances. The experimental results demonstrate the equal effectiveness of the two media in promoting learning and the retention of information. The major factor discriminating between the two methods is cost, and on this basis tape-slide is the more cost-effective method of preparation.

References

Carson, D. (1948) *The American way of life as portrayed in filmstrips — an experiment in visual education.* Research Publication No. 2, Scottish Educational Film Association, Glasgow.

Dale, E. (1946) *Audio Visual Methods in Teaching.* Holt, Rinehart, and Winston Inc., New York.

Driscoll, J. (1968) Educational films and the slow learner. *Mental Retardation,* **8**, 32-4.

Dwyer, F.M. (1972) *A Guide for Improving Visualized Instruction.* Learning Services, State College, Pennsylvania.

Dwyer, F.M. (1978) *Strategies for Improving Visual Learning.* Learning Services, State College, Pennsylvania.

Finn, J.D. (1953) Professionalising the audio-visual field. *Audio-Visual Communication Review,* **1**, 6-17.

Hoban, C.F. and Van Ormer, E.B. (1950) *Instructional Film Research 1918-1950.* Technical Report No. SDCM 269-7-19, Special Devices Centre, Port Washington, New York.

James, H.W. (1924) The relative effectiveness of six forms of lesson presentation. In Freeman, F.N. (ed.) *Visual Education.* University of Chicago Press, Chicago.

Miller, N.E. *et al.* (1957) Graphic communication and the crisis in education. *Audio Visual Communication Review,* **5**, 1-20.

Porotskaya, T. (1967) Selection of films and lessons prepared utilizing the cinema in the auxiliary school. *Spetsial 'naga Shokola* **13**, 86-9.

Spencer, K.A. (1976) A feasibility and cost-effectiveness study of the use of audio-visual aids in the education of the mentally handicapped. MSc dissertation, University of Newcastle-upon-Tyne.

Travers, R.M. *et al.* (1967) *Research and theory related to audio-visual information transmission.* Revised edition. Project No 3-20-003, US Department of Health, Education and Welfare, Office of Education.

Twyford, L.C., Davis J.S. and Seitz, C.P. (1954) *Evaluation of the film: 'Military Police support in emergencies'.* Technical Report No. SDC 269-7-52, Special Devices Centre, Port Washington, New York.

Vergason, G.A. (1966) Retention in educable retarded subjects for two methods of instruction. *American Journal of Mental Deficiency,* **70**, 683-8.

Vernon, P.E. (1946) An experiment on the value of film and filmstrips in the instruction of adults. *British Journal of Educational Psychology,* **16**, 149-62.

Zeckhauser, S.H. (1970) *A look at the media in higher education.* Harvard University, Cambridge, Massachusetts.

Appendix

Example: The following categories were used for the Prairie Dog responses. The total maximum score is 8, one point being awarded for each correct response.

1. The Prairie Dog lives in America
2.   is a small animal
3.   is brown
4.   has a short tail
5.   has a black tail
6.   holds its food in its paws
7.   makes noises when frightened
8.   lives in holes (burrows) in the ground.

Biographical note

Kenneth Spencer is Lecturer in Educational Technology in the Department of Educational Studies, University of Hull.

3. Spencer, K.A. and Clarke, A. (1981b) A BBC television series for the mentally handicapped, Educational Broadcasting International, 14(3), 110-113.

A BBC TELEVISION SERIES FOR THE MENTALLY HANDICAPPED

Can a single television series adequately cater for a wide range of mentally handicapped adults and children? KEN SPENCER and ANN CLARKE of the Department of Educational Studies, the University of Hull, evaluate a recent BBC experience.

Background

Since at least the early 1970s, BBC Schools Television had been receiving requests, both from organizations and from individual teachers, for a regular series for mentally handicapped children. The numbers of children in schools for the educationally subnormal were, however, felt to be too small to justify a special series and it was only when the educational needs of mentally handicapped adults began to be brought to the attention of BBC Continuing Education that the idea arose of a series intended for both adolescents and adults and jointly produced by the two departments. After a brief feasibility exercise, such a series was proposed to, and accepted by, the BBC's two education councils in 1976. Since *Let's Go* ... was such a new venture, it was felt that an independent evaluation of its value and significance was a priority, and funding was allocated from prize money awarded to the BBC from the Institut de la Vie based in Paris. The University of Hull was approached in June 1978 to undertake this evaluation.

The first questionnaire (programmes 1-10)

Questionnaires relating to the first ten programmes were sent in December 1978 to three types of establishment:

- Schools for the Educationally Subnormal (Severe) — ESN(S)
- Schools for the Educationally Subnormal (Mild) — ESN(M)
- Adult Training Centres — ATCs.

In total, questionnaires were sent to 494 ATCs and 1,234 schools. In order to ensure that the conclusion drawn from the questionnaire returns would be valid, special efforts were made to elicit information from non-respondents, including telephone calls to a randomly selected sample. It was established that 50 per cent of the schools did not find *Let's Go* ... helpful and therefore did not use it. Teachers reported that ESN(M) pupils found it hard to identify with the mentally handicapped actors, indeed some of them were insulted by the idea that they should watch the programmes. It has to be remembered that ESN(M) schools cater for a wide ability range, and teachers often found 'ordinary TV' programmes to be more appropriate.

The response rate for the ATCs was approximately 40 per cent and it seemed that, of the non-respondent sample, 47 per cent did not use the programmes. About 33 per cent of all ATCs did not use the series for a variety of reasons, including lack of facilities and a bias towards work-orientated instruction. The data obtained for evaluation were consequently heavily weighted towards the 'user' category.

From the review of all aspects of the available information it seems likely that the data obtained were predominantly from a group who both liked and

used the programme. It is important to bear in mind that there were also those who found the programme unsuitable for their purposes and needs. In the case of the schools this was a majority, for the ATCs a minority.

The responses to several open-ended questions give an overall perspective to the perceived strengths and weaknesses of the series and the following brief summary presents the conclusions from these responses for the schools and ATCs.

The schools' response

The majority of respondents were very positive about having a programme specially designed for the mentally handicapped and felt that the subject matter and its mode and manner of presentation made it eminently suitable for their purposes. It was seen as a motivational aid, assisting routine learning, and making use of 'the magic of the media'. There was mention of the reinforcement value of the series and, particularly where video facilities were available, its long-term teaching value.

Points mentioned less often, but nevertheless worthy of emphasis, were that *Let's Go* ... could help parents of mentally handicapped people realize the full potential of their children and also that the public could use the programme to achieve an understanding of the problems faced by the mentally handicapped in their everyday lives.

The major criticisms were concerned with the length, pacing and form of the programmes. The general impression was that there was really too much for ESN(S) pupils to assimilate and that points should have been made more slowly, clearly and in more detail, and that topics should have extended over more than one programme. Alternatively the programme could have been

longer, to allow for more reinforcement and repetition.

The least satisfied group of respondents were teachers from ESN(M) schools; they formed the largest group of total non-users and often appeared to regard the programmes as irrelevant or positively inappropriate for their pupils. Although many of the teachers found the pace and vocabulary level appropriate, several indicated that their pupils needed a wider range of topics presented at a more sophisticated level. In both 'M' and 'S' schools, teachers would have appreciated more time for follow-up work between programmes. There was a moderate number of complaints about subject matter, mainly along the lines that many viewers would find it hard to relate to what they saw which would consequently diminish its teaching value. The subject-matter was perceived as tending to be idealisitc, middle class and almost exclusively urban.

Finally, there was recurrent criticism that the series, not surprisingly, had failed to cater for the entire group of mentally handicapped viewers. It fell between stools and elicited such comments as 'subject matter too advanced for ESN(S)' or 'We would use a more advanced programme for our school leavers'.

The Adult Training Centres' response

Once again the opinion was expressed that the media could help capture interest hard to kindle in the teaching of over-worked subjects. For this group the level of presentation and subject matter seemed appropriate to a large number of respondents. Many instructors realized the potential value of the series for parents and the general public; they seemed to be very aware of the need to integrate the handicapped within society.

'Let's go a little more slowly' was the phrase used by one of the instructors, and this summed up the general criticism that more reinforcement, attention to detail and continuity were necessary. The level of teaching was said to be a little too high for many trainees who would benefit from a more basic series dealing with time, money, numeracy, etc. Once again, there were a few criticisms of the subject matter. It was not 'adult' enough, too 'idealistic' (nothing ever goes wrong) and too urban. More help was needed with the teaching of basic skills and a large number of respondents mentioned the need for programmes to help the trainee lead an independent life within the community (ie, budgeting, housework, etc).

Conclusions from the first questionnaire

We believe the following points represent the most important implications of the first questionnaire.

1. There is a great variation in opinions about the programmes among teachers in establishments for slow learning children and adults, ranging from outright disapproval to warm enthusiasm. We know something about those whose attitude is negative, but perhaps not sufficient in view of their number.

2. There was also considerable variation in the ways in which the *Let's Go...* programmes were used, from occasional viewing of a particular transmission to regular use with video back-up as part of an educational programme.

3. Of the three groups of respondents those from the ATCs were clearly the most satisfied. This will presumably surprise nobody since the programmes portray mentally handicapped adults in the context of Adult Training Centres. We are in no doubt about the instructors' gratitude towards the BBC for producing for the first time programmes specially designed to cater for the needs of their clients. However, our evidence suggests that a substantial minority of ATCs did not find the programme valuable.

The second questionnaire (programmes 11-20)

Our procedure for the evaluation of the second half of the series of programmes differed from that adopted with the first. The poor response to our request for information together with a desire to obtain more specific comments on individual programmes led us to solicit only those respondents in ATCs and schools who had replied fairly rapidly to our first request and had indicated their willingness to collaborate further.

In January 1979 a letter was sent to 90 ATCs, 33 ESN(M) schools and 33 ESN(S) schools (all of which had earlier proved willing and efficient) giving a list of programmes and transmission dates together with ten brief questionnaires, one to be completed after each programme had been used.

To our dismay, despite our efforts, we failed to obtain the immediate and full response for which we had confidently hoped. Many of the packets were returned weeks after the programmes had ceased, suggesting that our instructions had not been followed.

However, from all the evidence before us, we conclude that approximately 80 per cent of the respondents, who were known to be very favourably disposed to the first half of the series, used at least some of the last ten pro-

grammes, although it is probable that only about 60 per cent were frequent users. It seems, therefore, that the second part of the series was less well received than the first.

In general, the responses to the second questionnaire reinforce the findings of the first. There was widespread appreciation and enjoyment of the series, with some predictable reservations about the level, content and pacing of the programmes. This was, of course, inevitable as the series could not hope to satisfy the needs of all its viewers.

An overview of the responses to the questionnaire makes it clear that those concerned with teaching the more severely handicapped have been delighted that the BBC had the imagination and sympathy to produce *Let's Go...*, while teachers in ESN(M) schools found the programmes less relevant to their needs. However, it is also clear that the great heterogeneity of age, severity of handicap and availability of educational resources has led some teachers to a view that the programmes were superfluous.

The experimental evaluation of Let's Go...

A further investigation was designed to determine the relative effectiveness of *Let's Go...* programmes when used as audio-visual aids in a teaching situation. The major experimental design consisted of comparing two groups of trainees. One group was taught by a class instructor making use of aids normally available in the classroom; this is referred to as the Experimental I (EI) group. The other group received instruction from the same teacher who used a video-tape recorded version of *Let's Go...* as an additional aid; this group is the Experimental II (EII) group. From this design it may be argued that any differences which emerge from testing following the instructional periods are attributable to the use of the *Let's Go...* audio-visual materials. Such differences, if they are found, need not necessarily be solely due to the content of the programmes but may be due to the novelty effect which was not controlled for in the experiment, on the understanding that, should consistent differences emerge in favour of Experimental II group, a further series of experiments would be undertaken in order to determine the relative contributions of the many factors associated with the introduction of the video-tape recordings. In order to increase the validity of the results, four instructors took part in the experiments, two from each of two ATCs in different cities, and each instructor taught two small groups of trainees.

The results of previous research suggest that mentally handicapped

students derive great benefit from repeated showings of a variety of audio-visual materials, without the intervention of an instructor. A further small group, in each ATC, received instruction only from the video-tape. This group, referred to as the Control (C) group, saw the video-taped programme twice during each session and, as each programme lasted only 12 minutes, received only half the instructional time per week when compared with the other groups.

Selection procedures

Three *Let's Go ...* programmes were selected for the experiment and were used to teach skills over a two-week period with pre-testing to control for prior knowledge.

The nature and subject of many of the programmes in the series precluded their use in the experimental project for various reasons, eg *Let's Go and Cook a Meal*, though ideal in some ways for testing improvement in skills, included insufficient task analysis. Several of the other programmes, such as *Let's Go and Find a Hobby*, were obviously intended as a focus for discussion or as motivational aids. Three programmes selected as suitable — *Let's Go by Bus, Let's Go to the Post Office*, and *Let's Go and Telephone*. Only *Let's Go and Telephone* offered opportunities for behavioural measurement. The instructors were told some of the specific points to be assessed to counteract any possible bias towards those using the programmes and to ensure that they covered the right material in their teaching.

The number of potential subjects was inevitably reduced for several reasons, eg the subjects needed enough basic skills to make participation possible but those who had already mastered the relevant skills could not be included. It was impossible, however, to exclude the possibility that, in some cases, what was being measured was not the attainment of a new skill, but the recall of an old one.

Teachers were deliberately encouraged to use their preferred methods, and given no direction on how to teach the skills or use the recorded material. This was to further our aim of evaluating the use of *Let's Go ...* as it naturally occurred in the different educational settings.

Besides the initial screening already described, the groups were matched for potential learning ability. The 65 potential subjects were in the upper part of the moderately handicapped group, though some had in fact come from ESN(M) schools. The ages ranged from 16 to 52, the majority being under 35.

The instructors were chosen by the Centre as those with the most involvement with the 'educational programme' ment with the 'educational programme'

Table 1

	Group C	Instructor A		Instructor B	
		EI	EII	EI	EII
	PRE-TEST : ALL GROUPS				
Four days of instruction consisting of:	Video only (24 minutes per day)	Instruction only (40 minutes per day)	Instruction plus video	Instruction only (40 minutes per day)	Instruction plus video
	POST-TEST 1				
A further four days of instruction consisting of:	Video only (24 minutes per day)	Instruction only (40 minutes per day)	Instruction plus video	Instruction only (40 minutes per day)	Instruction plus video
	POST-TEST 2				

rather than industrial skills. All the teachers were qualified teachers of the mentally handicapped, and the group included one graduate and three people formerly employed in industry.

Experimental design

The experimental design described in Table 1 was used with each of the three programmes. Thus, there were five matched groups, formed on the basis of subjects' scores on the pre-test and assessment procedures. Two instructors participated, with two groups assigned randomly to EI and EII and the fifth group to C. Each instructor taught two groups, one from both EI and EII to control for teacher differences across the treatments. This also meant that the groups were small enough for effective teaching. Eight lessons were given per skill over a period of two weeks. The subjects were tested after four days tuition and again after a further four days. In this way the amount of learning which had occurred was systematically checked.

As the instructors used their own teaching methods, there was some variation in the way the skills were taught and the programme used. In one ATC the instructors used many visual aids and a lot of role play and practical work, taking the trainees out to use buses and a public telephone. Both tended to use parts of the programme, stopping and starting to reinforce specific points. In the other city, both teachers used similar methods and materials and only for telephoning was there any practical work. The lessons consisted mainly of discussions, with the programme shown at the beginning of the session.

Results

Members of the control groups, who watched each recording a total of eight times without any added instruction, significantly improved their performance over the testing time. There is a slight but unlikely possibility, more likely in the case of the final test, that the improvements could be in part attributed to familiarity with testing procedures. The girl who profited most had had experience with a private phone and acquired the necessary skills to cope with a pay phone as well as the new ability to answer relevant questions. However, for some of the control groups, learning was non-adaptive and inappropriate, eg, they concentrated on irrelevancies such as specific bus numbers.

The surprising result is that the control group, which received only half the instructional time per week, were not inferior, in a statistical sense, in all comparisons with the experimental groups. The results from one ATC indicate that overall there was no difference between the five groups for the *Bus* programme, or the *Telephone* programme when the ability to actually telephone was used as the criterion. In the other city the results indicate no difference between the five groups for the *Post Office* programme and the *Telephone* programme when the behavioural measure is used. When the performance of the control group after two weeks' instruction is compared with the two experimental groups after only one week (in which case all groups have received the same total amount of instruction) the results indicate that there are no differences between the five groups in either ATC setting. These

results should be treated with caution, however, in view of the small number of subjects involved in the control groups and further research is needed to clarify the potential use of video-tape recorded programmes as teacher substitutes.

We are more confident of the results concerning the advantages of providing video recordings of the *Let's Go...* programme as additional aids in a conventional teaching situation. The results do not indicate a single instance of an overall superior performance for the EII group. This result is wholly consistent across ATC settings, instructors and the three selected programmes. Therefore, we conclude that the use of the programme by the teachers did not improve performance. These findings are not in fact surprising, given the results of previous experiments involving the efficiency of television aids.

Conclusions

It seems that the results can be taken to demonstrate quite clearly that teacher involvement and the use of a variety of methods are at least as important as the use of television aids. Perhaps the most interesting aspect of this piece of investigative research is that there is a general improvement in the group using television alone without any instruction.

These findings may very well imply that for this type of audience at least, programmes like *Let's Go...* are most useful when they carry the burden of instruction rather than when used as adjuncts to teaching. This is by no means to suggest that teachers might become redundant, merely that the admixture of natural teaching with televised material is not necessarily a profitable educational method. In an age where individualized instruction is increasingly used in various educational settings, carefully devised programmes may profitably release teachers from certain repetitious activities in order that they have more time for others.

The experimental evaluation showed a number of instances of inappropriate learning in the unsupervised group. It may well be very difficult for an organization such as the BBC to design programmes which will produce wholly appropriate learning in all situations, and it may thus be necessary to consider the value of locally produced audio-visual materials (including video-recordings, films and tape-slide programmes) designed for automated instruction.

This article is based on the research report '"Let's Go..." An Evaluation of the BBC Television Series for the Mentally Handicapped', copies of which can be obtained from the authors.

EVALUATION ON A TIGHT BUDGET

DR TONY BATES, Head of the Audio-Visual Media Research Group at the Open University, describes low-cost evaluation methods that many educational broadcasting organizations could use to make their programmes more effective.

A challenge for educational broadcasters

I would like to throw the challenge down to managers of educational broadcasting systems: *you can run an effective evaluation programme in your educational broadcasting unit, without having to increase your budget at all.*

If you do not have an evaluation programme, it is almost certainly not because of lack of money. Economists will tell you that everything has a cost, and an evaluation programme is no exception. However, in comparison with the overall operating cost of any educational broadcasting system, in even the poorest country, the cost of an evaluation programme will be peanuts. Without too much disturbance to the educational broadcasting unit, these costs can be met by quite small changes to existing arrangements. In many cases, of course, it would probably be better to increase the budget slightly to allow for an evaluation programme to be developed with less disruption, but I want in this article to show how evaluation can be effective on a very tight budget.

Attitudes to evaluation

There can be lots of reasons, some genuine, many no more than rationalizations, why an educational broadcasting unit will not have an effective evaluation programme. The main reason, though, is that the educational broadcasting unit does not want it enough. There are two common attitudes I have come across: the first is that it is not the job of educational broadcasters to do 'research'; the second is that a 'scientific' evaluation would probably be very harmful to the unit, or to individuals, showing up all their weaknesses, possibly even leading to the unit being closed down. Lastly, it must be said that some units have very bad evaluation programmes, where the activity is no more than a ritual or public relations exercise, and the broadcasters know this.

These attitudes are often very strongly held. It is impossible to set up an effective evaluation unit on a tight budget unless the value to the educational broadcasting unit of an evaluation programme is fully understood and accepted, by managers and producers, so it is important to tackle these issues right at the start.

The educational broadcaster's responsibility

The majority of educational broadcasting organizations – in both developing and developed countries – have no evaluation programme. Why should they? Is not their job to make educational broadcasts? No, it is not. The programmes are made for a purpose, and that purpose is not to keep producers and equipment fully employed, but to help train teachers, or educate children, farmers, adult illiterates, etc.

The aim of evaluation is to make sure that the broadcasts do this as effectively as possible. The broadcasters' responsibility does not end at the point of transmission. An educational broadcasting unit without an evaluation programme is like a million dollar bridge with the middle section missing.

The difficulty for broadcasters is that the technology separates the producers from the users of the product. One aim of an evaluation programme is to bridge the gap, to provide feedback, in such a way that judgement is based on good information. Furthermore, evaluation is not something which is the responsibility of someone other than the educational broadcaster or producer. They are evaluating all the time: judgements are made about the quality of the script, the suitability of transmission times, the excitement that the programme might generate. These are professional judgements, but they need to be based on accurate knowledge of the potential audience and that audience's capacity to respond in the ways anticipated by

4. Spencer, K.A. (1981c) Alternative media: a review of the instructional effectiveness, Educational Broadcasting International, 14(2), 90-93.

Alternative media:
a review of the instructional effectiveness

KEN SPENCER
Department of Educational Studies, University of Hull

This article was originally prepared as a tape-slide programme for self-study

In this review I will be considering the relative effectiveness of what Schramm has called the 'big media' (films, TV) and the 'little media' (radio, discs, audio-tapes).

Different approaches

The traditional lecture or lesson is shown in Figure 1 as being suitable for a range of situations, from a single individual, through small groups, as found in most classrooms, to large groups in excess of several hundred students. The audience size can be dramatically increased, as indicated in Figure 2, by the use of broadcast media such as radio and television. In this case, although the total audience may be several thousand or indeed hundreds of thousands, the lecture may be received by individuals, small groups or even large groups gathering at reception sites.

Figure 3 shows another way in which the audience can be greatly extended — by recording the lecture, for example on film or audio-tape, and then distributing it to other locations where the material is replayed, again either to individuals or small or large groups.

The recorded lecture can offer a degree of flexibility which is not inherent in the traditional or broadcast form. The tape can be stopped and replayed if necessary, or it may be used on more than one occasion. Figure 4 gives a more comprehensive view of the most advantageous use of media with modern technological innovations. The lecture is recorded and possibly edited; the recording is then broadcast and received at distant locations where it is again recorded for later use by individuals or groups. In order to understand some of the advantages of the systems shown in Figures 2, 3 and 4 I will be considering first the research dealing with comparisons of traditional lessons with recorded versions of the same material, which will lead to conclusions concerning the relative effectiveness of new and old methods.

Film

With the advent of sound films for educational purposes in the early 1930s it was natural for researchers to consider the relative effectiveness of lectures and filmed recordings of such lectures. Hoban and Van Ormer's report, subtitled *Rapid Mass Learning*, summarized the research available during the period 1918 to 1950. One of the earliest researches investigating comparisons of sound films and lecture demonstrations was conducted by Clark (1932), who found three sound films to be as effective as the lecture demonstrations given by regular class instructors, in tests designed to measure thinking and reasoning ability. A later study by Hall

and Cushing in 1974 investigated the difference between sound film presentations and illustrated lectures dealing with three science topics. Again, no differences were found. In this study the oral lecture was based on the film scripts and the illustrations were also taken from the film artwork.

In an experiment conducted by Vernon (1946), information test scores of seamen who witnessed two showings of a film demonstrating two methods
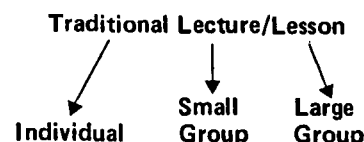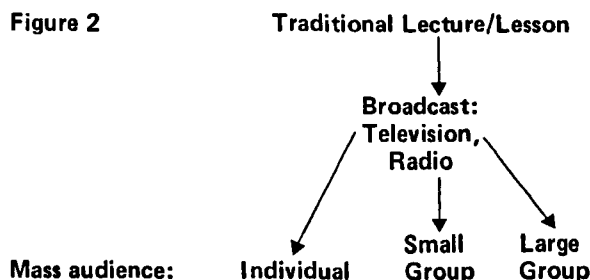
Figure 1

Traditional Lecture/Lesson

Individual   Small Group   Large Group

Figure 2

Traditional Lecture/Lesson

Broadcast:
Television, Radio

Mass audience: Individual   Small Group   Large Group

Figure 3

Traditional Lecture/Lesson

Recording:
Film, videotape, audio-tape, disc

Mass Audience: Individual   Small Group   Large Group

Figure 4

Traditional Lecture/Lesson

Recording

Broadcast

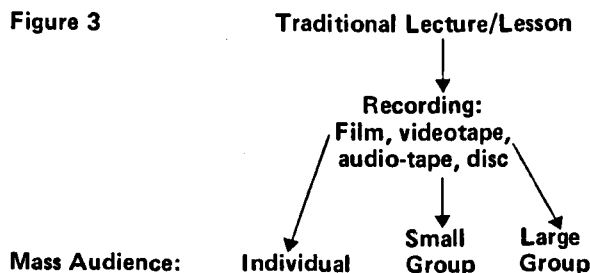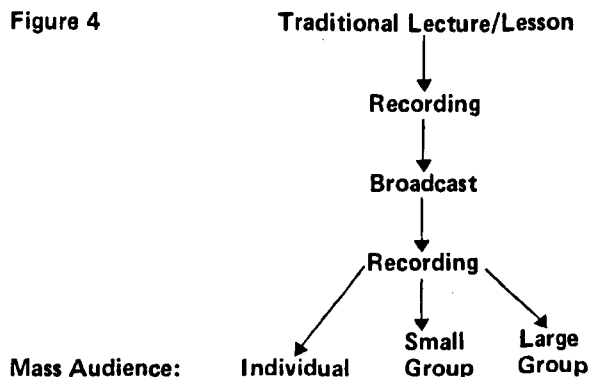Recording

Mass Audience: Individual   Small Group   Large Group

of taking depth soundings — a total of 50 minutes of instruction — were only 6 per cent below those of groups which had usual instruction lasting 3 hours. Vernon concluded that an hour's film appears to be as effective as 3 hours' weak oral instruction.

Hoban and Van Ormer summarize their review of the research, in which films are compared with demonstrations or lectures on science topics, and conclude that films reduce instruction time and are often equivalent to good instructors. They hasten to add that this should not be interpreted as meaning that films can eliminate the need for instructors, but rather that the effectiveness of instructors of average or below average ability can be improved.

Greenhill, in his introduction to a volume of abstracts on film and TV research by MacLennan and Reid (1964), comments that in general no significant differences were found in the bulk of studies comparing filmed courses with direct instruction in such subjects as college-level psychology, high school chemistry, physics, history and industrial arts. This is confirmed by Peggie Campeau's stringent review (1967) of the literature concerning audio-visual media.

Television

Just as research was indicating quite conclusively that there is no disadvantage in studying from filmed courses, a new technological innovation was entering the mass media market place.

Television, although invented in the late 1920s, was only beginning to make headway in the early 1950s, and it was natural for researchers to turn their attention to a medium which offered all the potentialities of film at a possible lower cost, for, as Lumsdaine and May concluded, film and TV can be considered as substantially identical media for many purposes.

The first research efforts used television as a means of expanding the total audience for a given lecture via closed circuit television, in which the transmission was carried from the lecture room by cable to several locations, enabling a single teacher to communicate with many hundreds of students. In fact, at one stage, audiences of 7,000 were taught via closed circuit TV in New York University.

In Greenhill's report of the Pennsylvania experiment he acknowledges that the major use of closed circuit TV is for the presentation of regular classroom instruction to students in several different classrooms. Here, television is used as a means of coping with mounting enrolments by making it possible for an instructor to teach larger numbers of students than would be possible by conventional means, Greenhill argues that the standard of instruction would also be raised by making it possible for these professors to present demonstrations and other teaching materials that it would be impossible or impractical to use under normal classroom conditions.

The shift towards TV, which was viewed as a panacea for all the educational ills during the mid 1950s and early 1960s, resulted in a proliferation of research studies. Stickell reviewed 250 comparisons of educational television and conventional face-to-face instruction from 31 research reports. The results are given in Table 1, which shows the number of studies classified as interpretable, partially interpretable or uninterpretable. Overall, 75 per cent of the studies showed no difference, with equal

Table 1*

THE COMPARISONS

The results were:	Favoured Face-to-Face Instruction	Showed No Difference	Favoured ETV	Totals
Interpretable	0 (0%)	10 (100%)	0 (0%)	10
Partially interpretable	0 (0%)	20 (87%)	3 (13%)	23
Uninterpretable	28 (13%)	158 (73%)	31 (14%)	217
Total Sample	28 (13%)	188 (75%)	34 (14%)	250

*From: Stickell, D.W.: 'A Critical Review of the Methodology and Results of Research Comparing Televised and Face-to-Face Instruction', doctoral dissertation, Pennsylvania University, 1963.

percentages favouring TV or face-to-face instruction. The 10 results which are fully interpretable, according to Stickell's criteria, clearly indicate that, over a wide range of subjects, there is no difference between class lectures and TV instruction. In a review of teaching by television Chu and Schramm found that, by 1967, of 421 separate comparisons taken from 207 published reports, 308 showed no difference, 63 showed TV to be superior and 50 found conventional instruction superior. These results confirm Stickell's earlier review of the literature. The research reviewed was so wide ranging that they were also able to conclude that TV can be used efficiently to teach any subject matter where one-way communication will contribute to learning.

These results lead to the inevitable conclusion that course enrolment can be greatly expanded by the use of educational television and that student performance will not suffer. However, the second point referred to by Carpenter, concerning the possible improvements in instruction, does not seem to be tenable. If this is the case, what are the advantages of using film or television? The answer is rather complex. When there is no difference in performance between the two modes the most obvious measure then to come under scrutiny is the relative cost of implementing either system. And costs depend on how the system is to be used, in particular the total number of students receiving instruction. In 1970 Sally Zeckhauser produced her policy document *A Look at Media in Higher Education* which suggested that for classes of 500 students television was between 5 and 25 times more expensive than the traditional lecture. However, if the audience is greatly extended, TV teaching can be delivered at a cost below that of conventional means. Mayo, McAnany and Klees indicate that for the Mexican 'Telesecundaria' the break-even point is 10,000 students. With 28,000 students enrolled in classes of 50 the cost is 1.5 times as much by conventional methods as by television. With a reduction in the staff-student ratio to 30 conventional costs would be 3 times as much.

Radio

Bates has suggested that the full potential of radio has been missed in the Open University because producers have allowed the more interesting pressures of television to dominate their activities. This is also

true in the field of the media research: as Hancock laments, of more than 500 documented studies comparing educational media most are concerned with instructional television and relatively few with radio. However, the results of research on instructional radio show it to be an effective medium, its principal attraction lying in its low cost when compared with television. For example, with a hypothetical system transmitting 500 hours of programmes annually to 200,000 students, television costs 4 times as much as radio (Jamison, Klees and Wells).

An early study of the effectiveness of radio was reported by Carpenter in 1937. He prepared 15-minute radio lessons on science for pupils ranging from fourth grade to senior high school level. The results of the end of term examinations indicated that pupils taught by radio did as well as, or better than, those taught by conventional methods. Attitude reports from pupils showed a high degree of interest in radio lessons.

An interesting experimental design for a comparison of radio and conventional teaching is reported by Heron and Ziebarth in 1946. 98 college students in general psychology were divided so that one half attended classroom lectures while the other half listened to the same lectures over the radio. Halfway through the course, the two groups changed places. Mid-session and final examination results indicated that the radio was as effective as the face-to-face instruction.

Another example of the effectiveness of radio as a teaching instrument, this time in a developing country, is reported by Chu and Schramm, concerning the work of Mathur and Neurath in 1959. A total of 145 villages in Bombay state, averaging about 850 people each, were chosen as the experimental group and were provided with radio sets. A similar number of villages without radio sets served as the control group. Twenty special farm programmes were broadcast twice a week for 30 minutes. Comparison of test results, both before and after the broadcast programmes, found a significant increase in knowledge in the radio villages, but only negligible increases in the non-radio groups. This and many other reports indicate the efficacy of instructional radio, summarized by Forsythe thus:

'Research clearly indicates that radio is effective in instruction. Experimental studies comparing radio teaching with other means or media have found radio as effective as the so-called conventional methods. Even though radio has been criticized for being only an audio medium, studies have shown that visual elements in learning are not uniformly important. In many educational situations visuals may be more harmful than helpful. Also, the efficiency of combined audio and visual media has been challenged by studies which show that multi-channel communications may not be inherently more effective than single channel presentations.'

Audio recordings

The research results from instructional radio are further confirmed when audio recordings of lectures are compared with traditional methods.

Popham, in 1961, divided an introductory graduate course into two sections. In one he taught in a lecture-discussion format; in the other, he played a tape-recorded version of the lecture and then led a brief discussion period. Following instruction several tests were administered and it was found that there were no differences between the groups. During a second experiment (Popham, 1962) the discussion was led by an untrained student, and again there were no differences between tape-taught and conventionally-taught students. Students, in both experiments, had generally favourable attitudes to the taped lectures. They thought that the lectures were better organized and they felt free from distractions during the taped presentation. However, they did miss the opportunity to question or disagree with the instructor.

Popham concluded that there may well be a legitimate place for taped teaching at college level. In view of this optimistic report it is surprising to learn that John Menne and his colleagues, in 1969, could find no mention of further taped lecture courses in the literature during the intervening 7 years. By 1969, the rapid expansion of cassette tapes was making the medium ideally suited to individualized learning, and it was this aspect which was of interest to Menne. The course used was an introductory psychology course at Iowa State University. The lectures were recorded on tape and the blackboard notes were assembled to form a booklet. The hypothesis was that there would be no difference between the groups listening to the tapes and those who attended the lectures.

A total of 290 students chose the experimental audio-tape presentations, with 408 going for the traditional lecture. However, the drop-out rate was much greater for the traditional group, representing 14 per cent of the total, compared with only 2 per cent of the tape-recorded lecture group. Each member of the experimental tape group was issued with a tape recorder, a complete set of lectures tapes, a booklet and a schedule of the lecture topics to be given to the lecture group. The audio-tape group was self-paced though they were required to take 3 objective tests during the course. Information was available concerning student performance on several measures. There were no significant differences between the two groups.

The noticeable difference in drop-out rate is clearly a benefit of this type of approach and the authors suggest that students tend to leave courses if they have fallen behind in their work and see little hope of recouping. This does not occur as readily if the material is always available on tape.

Ackers and Oosthoek report a similar experimental course to the one at Iowa, but with subjects randomly assigned to either a taped or live lecture. The subjects in the tape group again had individual access to recorders and tapes on the subject of microeconomics, and were able to follow the course at their own speed, within certain broad limits. The performance of both groups was assessed in a June examination which, according to the authors, called for the sub-categories 'Application' and 'Analysis' of the category 'Comprehension' from Bloom's taxonomy of objectives. The results indicated a slight advantage in favour of the tape group. The difference is statistically significant and amounts to a superior mean score of 6 per cent.

The authors indicate that it had been anticipated that the well-programmed tape system would not only produce a higher mean score in the microeconomics examination, but, compared to the lecture group, also a significantly greater percentage of students passing the examination. This was not the

case: approximately 50 per cent of students passed the examination in both treatments. However, an analysis of the number of students failing the examination showed that the percentage failing because they did not sit the examination was only 22 per cent for the taped-lecture group, compared with 38 per cent of the total lecture group. Thus, when account is taken of failure by default the result is clearly in favour of the tape group, with a 38 per cent pass rate for the tape lectures and 27 per cent pass rate for the conventional lectures.

Media comparisons

The research on TV and radio seems to suggest that both media can be as effective as traditional instructional methods. If this is the case, do comparisons between TV and radio show that they are equally effective? If so, there will be little justification for using the more expensive medium to attain substantially the same objectives.

McLuhan reports an experiment conducted in Toronto in which information was presented via television, radio, lecture and print. When each medium was allowed 'full opportunity to do its stuff' (p.311), television and radio showed results high above the lecture and print presentations. Unexpectedly to the testers, however, radio was found to be superior to television. McLuhan attributes this to the fact that radio is a 'hot' medium, which can engage the students, allowing more participation.

A further example of this comes from a study in which introductory lectures were presented to electronic engineering students either by video or audio-tape (Spencer, 1977). The audio version was essentially the same as the video version but with formulae and illustrations presented in printed handouts. There were, in fact, two video versions: a low-cost version in which the lecturer spoke directly to camera and wrote formulae on a board; and a high-cost version in which there was extensive use of complex graphics. Responses to a questionnaire indicated that a majority of students (60 per cent) preferred the high-cost video version of the programmes. However, there was no evidence from the tests that they had learned more from this version, and it was concluded that the audio presentation, produced at a fraction of the cost of the TV programmes, was more cost-effective.

A similar result was obtained by Foxall, who compared students viewing either a TV maths programme or a comparable tape-slide version. She concluded that tapes with slides were at least as effective as television and were favoured by teachers as they were more directly under their control.

MacDonald also offers evidence which confirms these results. Video-tapes were compared with audio-tapes plus relevant slides to determine whether the additional cost of video-tape was justified. The two groups did not differ significantly from each other on an information test, and it was concluded that the audio-tape version was more cost-effective.

These results clearly demonstrate the effectiveness of instruction via the media, and while not lending full support to Wilkinson's suggestion that the long history of no significant difference between instructional strategies reduces the cost-effectiveness question to one of finding the cheapest method, it does lead to the inevitable conclusion that 'small is beautiful'.

References

Ackers, G.W. and Oosthoek, J.K.: 'The Evaluation of an Audio Tape Mediated Course', British Journal of Educational Technology, No. 3, 1972, pp.136-147.

Bates, T.: 'Whatever Happened to Radio at the Open University?', Educational Broadcasting International, Vol.12, No.3, 1979, pp.119-123.

Campeau, P.L.: 'Selective Review of Literature on Audio Visual Methods of Instruction', in Briggs, Campeau, Gagne and May (Eds) Instructional Media, American Institutes for Research, Pittsburgh, Pa., 1967.

Carpenter, H.A.: 'Science in the Rochester School of the Air', Science Education, No. 21, 1937, pp.77-81.

Chu, C.C. and Schramm, W.: Learning from Television: What the Research Says, National Association of Educational Broadcasters, Washington DC, 1967.

Clark, C.C.: 'Sound Motion Pictures as an Aid in Classroom Teaching', doctoral thesis, New York University, 1932.

Forsythe, R.D.: Instructional Radio: A Position Paper, Stanford University, ERIC: ED 044 933, 1970.

Foxall, A.: 'Television and Radio Vision in the Teaching of Modern Mathematics: A Comparative Study', British Journal of Educational Technology, No. 3, 1972, pp.236-244.

Greenhill, L.P.: 'Close-Circuit Television for Teaching in Colleges and Universities', Pennsylvania State University, mimeographed, 1966.

Hall, W.E. and Cushing, J.R.: 'The Relative Value of Three Methods of Presenting Learning Material', Journal of Psychology, No. 24, 1947, pp.57-64.

Hancock, A.: Planning for Educational Mass Media, Longman, London, 1977.

Heron, W.T. and Ziebarth, E.W.: 'A Preliminary Experimental Comparison of Radio and Classroom Lectures', Speech Monographs, No. 13, 1946, pp.54-57.

Hoban, C.F. and Van Ormer, E.B.: Instructional Film Research, 1918-1950, US Navy Special Devices Centre, Port Washington, NY, 1950.

Jamison, D.T., Klees, S.J. and Wells, S.J.: The Costs of Educational Media, Sage Publications, London, 1978.

Lumsdaine, A.A. and May, M.A.: 'Mass Communication and Educational Media', Annual Review of Psychology, No.16, 1965, pp.475-534.

MacDonald, R.: A Comparison of Two Methods of Presenting Occupational Information, Appalachia Education Lab, Charlestown, West Virginia, ERIC: ED 062 574, 1972.

MacLennan, D.W. and Reid, J.C.: Abstracts of Research on Instructional Television and Film: An Annotated Bibliography, Stanford University, California, 1964.

Mathur, J.C. and Neurath, P.: An Indian Experiment in Farm Radio Forums, UNESCO, Paris, 1959.

Mayo, J.K., McAnany, E.C. and Klees, S.J.: The Mexican Telesecundaria: A Cost-Effective Analysis, Institute for Communication Research, Stanford, California, 1973.

McLuhan, M.: Understanding Media, Routledge & Kegan Paul, London, 1964.

Menne, J.W., Hannum, T.E., Klingensmith, J.E. and Nord, D.: 'Use of Taped Lectures to Replace Class Attendance', Audio Visual Communication Review, No. 17, 1969, pp.42-46.

Popham, W.J.: 'Tape Recorded Lectures in the College Classroom', Audio Visual Communication Review, No. 9, 1961, pp.109-118.

Popham, W.J.: 'Tape Recorded Lectures in the College Classroom II', Audio Visual Communication Review, No. 10, 1962, pp.94-101.

Schramm, W.: Big Media, Little Media, Sage Publications, London, 1977.

Spencer, K.A.: 'An Evaluation of the Cost-Effectiveness of Audio-Taped and Video-Taped Self-Study Media Presentations', British Journal of Educational Technology, Vol.8, No. 1, 1977, pp.71-77.

Stickell, D.W.: 'A Critical Review of the Methodology and Results of Research Comparing Televised and Face-to-Face Instruction', doctoral dissertation, Pennsylvania State University, 1963.

Vernon, P.E.: 'An Experiment on the Value of the Film and Filmstrip in the Instruction of Adults', British Journal of Educational Psychology, No. 16, 1946, pp.149-162.

Wilkinson, C.L.: 'Cost Evaluation of Instructional Strategies', Audio Visual Communication Review, No. 21, 1973, pp.11-30.

5. Spencer, K.A. (1981d) AV communication: small is beautiful, Audiovisual Librarian, 7(1), 22-27.

AV communication: small is beautiful

Ken Spencer, *Lecturer in Educational Technology, University of Hull*

For several decades now attempts have been made to produce guides to help teachers, librarians and administrators select the most appropriate medium for instructional purposes. The earliest of these is the 'Cone of Experience', produced by Edgar Dale in 1946. Figure 1 shows how the 'Cone' grades experiences, starting with direct and purposeful experiences, which, Dale explains, form the unabridged version of life, such as preparing a meal or performing a laboratory experiment. As experiences become more remote from the concrete actuality of life so they are placed higher up the cone. Field trips and exhibits, in which the student is represented as a spectator, are found mid-way between the direct experiences and the abstractions of verbal symbols which form the pinnacle of the cone, represented by designations that bear no physical resemblance to the objects or ideas for which they stand.

This arrangement of pedagogical experiences has been adopted to justify the use of particular methods of instruction. Dale, for example, claimed that one of the proven contributions of audiovisual materials was that they make learning more permanent, when compared with more traditional, though unspecified, methods. This statement, it is suggested, is based on the 'distillation of a vast amount of research by many investigators'.

The underlying assumption of the 'Cone of Experience' appears to be that the more realistic and concrete a particular experience is, the more permanent will be any resulting learning. This has been dubbed the 'realism' theory when applied to audiovisual communication, and has been linked to Carpenter's (1953) assertion that aids which have a high degree of similarity to the objects which they represent will be more effective for instructional purposes than less realistic aids.

The ideas of many audiovisual enthusiasts conformed to the realism view for many years following the publication of Dale's influential text. They were first challenged by Travers (1964) who based his arguments on the results of psychological studies of man's ability to use and process sensory information. Early studies by Broadbent (1958) seemed to suggest that man can only handle information coming from one sensory input at a time, and that only a fraction of that sensory input can be processed. The implication from this research, it seemed to Travers, was that a combination of audio and visual presentations would not necessarily be beneficial because attention could only be directed to one sensory channel at a time and that, as the human processing system can handle only a fraction of information passed to it by the sense organs, simplified displays, such as black and white line drawings, would prove to be equally as useful as more complex, realistic displays.

The two opposing views have generated a vast amount of experimental data, none of which is conclusive, but some of which is pertinent to the user faced with the problem of selecting suitable audiovisual programmes. The following review of research is highly selective but does represent a balanced view of the results of many reports which I have studied in recent years.

1 Colour
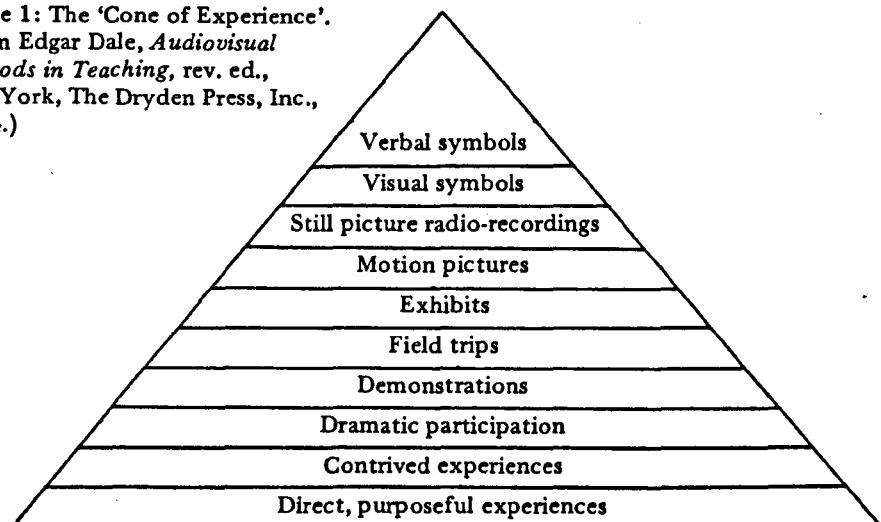This is an important attribute of pictorial displays but one that is frequently

22

over-emphasized by adults choosing materials for children. Rudisill (1952) found that children prefer an uncoloured illustration which gives the impression of reality to a coloured one that does not, although if two pictures are identical in all other respects most children prefer a realistic coloured one to an uncoloured one.

However, although we have all experienced a child's *preference* for the colourful, the realistic, it is still necessary to ask the question: is a colourful, realistic presentation more effective, in instructional terms, than a simplified monochrome version? The question is increasingly being posed in educational circles because of the expense associated with sophisticated realistic presentations. It is also necessary to point out the limitations of the question itself: namely, that measuring the effectiveness of any instructional intervention is extremely difficult, even if we are able to define such terms as 'effectiveness'. In the following analysis we must constantly remind ourselves that the relative effectiveness of the various methods of instruction is usually measured in terms of the recall of information in paper and pencil tests. Occasionally performance is used, such as the assembly of a piece of machinery, with time for completion or number of errors forming the test score. Most of the research studies quoted here have based their comparisons on the amount of information retained following exposure to one or another presentation. Other effects, such as increased motivation leading to further

Figure 1: The 'Cone of Experience'.
(From Edgar Dale, *Audiovisual Methods in Teaching*, rev. ed., New York, The Dryden Press, Inc., 1954.)

Verbal symbols
Visual symbols
Still picture radio-recordings
Motion pictures
Exhibits
Field trips
Demonstrations
Dramatic participation
Contrived experiences
Direct, purposeful experiences

study or increased involvement in other curriculum subjects, are difficult to measure and have been ignored by most researchers in this field.

Vander Meer (1954) looked at student preferences for colour and monochrome films and attempted to link these to informational learning. He quoted three common reasons for using colour in selected instructional films: (1) colour may be an important cue in learning what the film is intended to teach, e.g. a colour film to teach the identification of snakes might be expected to be more effective than a monochrome version if the colour of the snakes is an important cue in identifying them; (2) contrasting colours in graphic presentations could be used to make things stand out; (3) colour may be pleasing to the learner and this may have an indirect effect in promoting greater learning.

The results of the research, which included information tests with numerous colour items and preference rating scales, led to the conclusion that the use of colour in instructional films which may seem to 'call for colour' does not appear to be justified in terms of *greater learning* on the part of those who view the films. Students did tend to prefer the colour versions of the five films used in the study but this preference was not reflected in higher test scores for the groups viewing the colour films.

Vander Meer's study was the first of several to produce results which seem to run counter to the 'realism' theory of audiovisual presentation. A later study, conducted by Kanner and Rosenstein (1960), considered the relative effects

of colour and monochrome television when used to instruct army trainees. The results of multiple-choice tests, which included many colour-related items, e.g. showing various coloured resistors and asking for their value in ohms, showed no differences between the two presentations in ten out of the eleven programmes. The one difference noted in favour of the colour version, though it was considered unimportant in view of the overall picture and the small differences in test performance.

A summary of more than 20 research projects investigating the value of colour in pictorial displays led Cox (1976) to the conclusion that there is no marked difference in learning from colour or monochrome film or television. However, as some researchers have pointed out, this does not mean that colour has no effect; but probably that such effects are not to be found in the learning of central information, and probably work at other levels where measurement is difficult or insensitive.

2 Pictorial complexity
This term has been used by Dwyer (1978) to mean the degree of realism in a pictorial representation. Colour is one factor involved in pictorial complexity, but, according to Dwyer, the degree of realism forms a continuum from simple line drawings, through detailed shaded drawings and models, to photographs. He prepared a teaching unit on the human heart which was presented to students by synchronized tape-slides. Nine different versions of the slides were produced with varying degrees of pictorial complexity. Assessment was by four tests, two of which (drawing and identification) required the subject to draw and label a schematic diagram of

24

the heart, and two which were purely verbal, testing knowledge of terminology and comprehension. A total criterion score was obtained by summing the scores on the individual tests. Dwyer has used the materials in a very extensive series of experiments over a period of ten years. The results indicate that for pupils of widely varying ages and abilities the simple line drawings are just as effective as the more realistic drawings and photographs for the tests involving drawing or labelling of a given illustration. The surprising result is that the version *without* visuals is just as effective as those with visuals for the verbal tests. When all results are combined in the total criterion score the simple black and white line drawing is found to be as effective as the more realistic versions.

A similar result was obtained by May and Lumsdaine (1958) who compared a black and white film, made from pencil sketches produced originally as a script guide, with the final full colour version. The crude black and white version costs only a fraction of the colour version but was found to be equally as effective in terms of informational learning, much to the surprise of both authors and film sponsors.

3 Pictorial motion
It is implicit in the 'Cone of Experience' that motion films and television are more realistic than still pictures, and accordingly they have been considered to be more effective as instructional aids. This was tested as long ago as 1928 when Brown investigated the relative values of films and filmstrips for teaching factual information about physiology. His results led to the conclusion that the filmstrip,

with the greater exchange of comment it allowed, proved better. Later, Carson (1948) compared a sound motion film with a sound filmstrip and he also reported that the filmstrip version was superior in informational and conceptual learning. Such results seem to indicate that motion adds little to the effectiveness of audiovisual communication. It is certainly the case for informational and conceptual information which is usually measured by paper and pencil tests. However, for perceptual-motor tasks involving such actions as knot-tying, which use the actual performance as the criterion test, the experimental results do tend to favour the motion medium. Roshal (1949) compared film and still pictures as media demonstrating knot-tying skills and found the motion picture medium to be more effective. Two experiments using 'time-to-complete-operations' as the measure of effectiveness, showed that the motion medium enabled the tasks to be completed in a significantly shorter time (Silverman, 1958; Spangenberg, 1973), although the redesign of the slide sequences in Spangenberg's experiment eventually produced results equal to the television presentation, showing that differences were not due entirely to the motion element. Interestingly, Silverman found there to be no difference between his groups when pencil and paper tests, which included identifying parts of the equipment, were used as measures of effectiveness.

'Little media' versus 'big media'
Schramm (1977) has recently contrasted the use of the expensive 'big media' (television, films) with the less expensive 'little media' (radio, slides) and asks: are the big media always worth five times as much as the little media? There is, of course, no unequivocal answer to this, but the evidence

25

from research evaluations does suggest that the big media are frequently used in situations where simpler aids can be equally effective.

An example of this comes from a study in which introductory lectures to electronic engineering students were presented either by video or audio tape (Spencer, 1977). The audio version was essentially the same as the video version but with formulae and illustrations presented in printed handouts. Costs were estimated for both formats and performance on multiple-choice questions was used as the criterion measure. Responses to a questionnaire indicated that a majority of students (60%) preferred the expensive TV version. However, although most students preferred this version, there was no evidence from the tests that they had learned more and it was concluded that the audio presentation, produced at a fraction of the cost of the TV programmes, was more cost-effective.

A similar result was obtained by Foxall (1972), who compared the performance, on a specially designed modern mathematics test, of students who viewed either a television maths programme or a comparable tape-slide version. She concluded that radio-vision (tapes plus slides) was at least as effective as television in promoting the learning and application of mathematical concepts. Also, comments from primary school teachers suggested that radio-vision was favoured because it was more directly under the teacher's control.

Conclusion
From such research evidence it is clear that simplicity and conciseness are the watchwords in selecting media for use in direct teaching and instruction. Much of

the research has involved materials which carry the main burden of instruction, and in many cases act as substitutes for teachers, in such cases audiovisual materials will be more cost-effective if they are simple rather than complex.

References

Broadbent, D E
Perception and communication. London, Pergamon Press, 1958.

Carpenter, R R
'A theoretical orientation for instructional film research'. *Audio Visual Communication Review* 1 1953. 38-52.

Carson, D
The American way of life as portrayed in filmstrips — an experiment in visual education. Research Publication No.2. Scottish Educational Film Association, Glasgow, Scotland, 1948.

Cox, S M
'Colour in learning from film and TV: a survey of research with some indications for future investigations'. *Educational Broadcasting International* June 1976. 85-92.

Dale, E
Audiovisual methods in teaching. New York, Holt, Rinehart and Winston, Inc. 1946.

Dwyer, F M
Strategies for improving visual learning. Learning Services, State College, Pennsylvania, 1978.

Foxall, A
'Television and radiovision in the teaching of modern mathematics: a comparative.

study'. *British Journal of Educational Technology* 3 1972. 236-44.

Kanner, J H and Rosenstein, A J
'Television in army training: colour vs black and white'. *Audio Visual Communication Review* 8 1960. 243-52.

May, M A and Lumsdaine, A A
Learning from films. New Haven, Yale University Press, 1958.

Roshal, S M
Effects of learner representation of film mediated perceptual — motor learning. Technical Report No. SDC 269-7-5. Port Washington, New York, Naval Training Devices Centre, Office of Naval Research, 1949.

Rudisill, M
'Children's preferences for colour versus other qualities in illustrations'. *Elementary School Journal* April 1952. 444-51.

Schramm, W
Big media, little media: tools and technologies for instruction. London, Sage Publications, 1977.

Silverman, R E
The comparative effectiveness of animated and static transparencies. Technical Report No. Navtradevcen 78-1. Port Washington, New York, Naval Training Devices Centre, Office of Naval Research, 1958.

Spangenberg, R W
'The motion variable in procedural learning'. *Audio Visual Communication Review* 21 1973. 419-35.

Spencer, K A
'An evaluation of the cost-effectiveness of audio-taped and video-taped self-study media presentations'. *British Journal of Educational Technology* 8 1977. 73-7.

Travers, R M W
Research and theory related to audio-visual information transmission. Project No. 3-20-003, US Department of Health, Education and Welfare, Office of Education, 1964.

Vander Meer, A W
'Colour vs black and white in instructional films'. *Audio Visual Communication Review* 2 1954. 121-34.

Audiovisual equipment reviews

The Audiovisual Librarian is hoping to include regular comparative reviews of audiovisual equipment, and particularly those kinds of equipment likely to be particularly useful to libraries and librarians.

Anyone who is in a position to undertake such comparisons and would be willing to help should write to Helen Harrison, Media Librarian, The Open University, Walton Hall, Milton Keynes MK7 6AA.

6. Spencer, K.A. (1988) The Psychology of Educational Technology and Instructional Media, pp198, London: Routledge.

THE PSYCHOLOGY OF EDUCATIONAL TECHNOLOGY AND INSTRUCTIONAL MEDIA

KEN SPENCER

THE PSYCHOLOGY OF EDUCATIONAL
TECHNOLOGY AND INSTRUCTIONAL MEDIA

THE PSYCHOLOGY OF EDUCATIONAL TECHNOLOGY AND INSTRUCTIONAL MEDIA

KEN SPENCER

CONTENTS

ACKNOWLEDGEMENTS

CHAPTER 1

THE EVOLUTION OF EDUCATIONAL TECHNOLOGY

Educational technology is composed of at least two over-lapping subsets: technology in education and technology of education. The tools-technology, sometimes known as the hardware approach to educational technology, but more commonly known under the title audio-visual aids or instructional media, may be thought of as technology in education. Whereas, the educational application of knowledge from the behavioural sciences, such as psychology, forms the basis of a technology of education.

Educational technology can be seen as an umbrella term for the application of scientific knowledge in the field of education. This covers considerable ground, as can be seen in Figure 1-1, which lists some of the areas for inclusion in Hilgard's (1964) instructional technology and in Figure 1-2, which identifies crucial disciplines which impinge on Hawkridge's (1981) concept of educational technology.

Jonassen (1985) has discussed the psycho-philosophical perspective of educational technology in this century, showing the various shifts in theoretical foundations and practices associated with four distinct phases: the physical science or instructional media movement, the behaviourist and neo-behaviourist movements, and finally the constructivist or cognitive science approach.

The physical science approach gained momentum early in the century as the audiovisual movement, focusing on the machines and materials, rather than the learner (Saettler, 1968). This approach was concerned with the effects of devices and procedures, which were seen as acting as an antidote to the excessive verbalism of traditional teaching methods (Wittich and Schuller, 1953). The new media were to supply a concrete basis for conceptual thinking, make learning more permanent, develop continuity of thought, the growth of meaning and efficiency, depth, and variety of learning (Dale, 1954). It yielded to the behaviourist approach following the training effort of World War 2, culminating in the 'systems approach' to course design in the 1960's. In many ways this remains the corner-stone of educational technology. It was grounded historically in behaviourism and later developed by behavioural psychologists who explicitly excluded the study of mental processes, placing emphasis only on observable changes in behaviour. This aspect of educational technology is most obvious in the linear teaching machines devised by B.F. Skinner.

There has been a gradual reaction to this radical

1

Figure 1-1. Components of Educational Technology: Physical
and behavioural sciences. (Based on Hilgard,
1964)

Physical sciences and technologies

Optics: projected pictures (still and moving)
Chemistry: paper, ink, photography
Electronics: sound recording, television, complex
                  teaching machines
Mechanical
Engineering: printing press, movable type, movies, sound
                  recording
Electrical
Engineering: sound films, sound recording, television

Mathematics

Probability &
statistics: experimental design

Behavioural sciences

Psychology: Individual differences, psychometrics,
                  learning theory, task analysis,
                  instructional media research, teaching
                  machines

behaviourism resulting in what has been termed neo-
behaviourism, which acknowledges that behaviour is all we
can observe, but claims that the observed behaviour depends
on mental processes. It is at this stage of evolution that
educational technology is to be found in its present form,
according to Jonassen.

However, there has been a major revolution going on in
psychology for several decades, which has replaced the
mechanistic behaviourist model of human behaviour with the
constructivist model. Here the individual is seen to select
and interpret the raw materials of experience, which are
ultimately transformed into different realities for
different people. Piaget (1970) says that knowing reality is
a process of constructing systems of transformations that
model reality. This cognitive model of learning has largely
replaced behaviourism and is now in the process of being
assimilated into the technology of education. If this occurs
to any great extent the goal of replication of behavioural
responses will be replaced by the activation of covert

Figure 1-2. Components of Educational Technology: Elements
of a concept map. (Based on Hawkridge, 1981)

Methods of learning:

cognitive psychology, learning styles, interaction analysis,
gaming/simulation, communication theory, linguistics,
textual communications, artificial intelligence, information
processing

Objectives for learners:

systems theory, epistemology, politics, philosophy,
sociology

Evaluation of teaching and learning:

cost-benefit/effectiveness analysis, economics, attitude and
opinion research, social psychology, teacher evaluation,
content analysis, measurement of learning, psychological
measurement, psychology, mathematics and statistics,
computing

Environments:

group dynamics, logistics, individualised learning,
anthropology

Media for learning:

design, graphics, electronics, engineering, production
techniques

mental processes required to build an appropriate cognitive
model of reality within the learner.

Hilgard's early formulation is mainly composed of those
physical sciences which contribute to the production of
educational hardware or machines, and those behavioural
aspects which are associated with the first three stages in
the evolution of educational technology: from the physical
science approach to the neo-behaviourist approach.

Hawkridge's later drawing of the educational technology
map is much more comprehensive, covering most of Hilgard's
earlier formulation and extending it to include
anthropology, sociology, philosophy and politics. Psychology
is particularly well-represented, including cognitive
psychology in addition to those aspects traditionally

associated with educational technology.

In the following chapters the evolution of educational technology in its various guises will be considered from the psychologist's point of view, but first it is necessary to introduce the three major psychological theorists whose early work profoundly affected the development of educational technology throughout most of this century.

Thorndike, Edward Lee (1874-1949)

Although later known for his pedagogical teachings, Thorndike was initially concerned, not with human learning and its educational implications, but with animal learning and intelligence. His learning theory, one which was to dominate all others in America for a quarter of a century, was first announced in his doctoral dissertation 'Animal Intelligence' (1898). This was profoundly affected by Darwinian theory, which, according to Thorndike, provided psychology with the evolutionary point of view.

> .... the mind's present can be fully understood only in the light of its total past.
> Psychology has by no means fully mastered this lesson. Human learning is still too often described with total neglect of animal learning. But each decade since The Origin Of The Species appeared has shown a well marked increase in comparative and genetic psychology. (Thorndike, 1909, p.65-80)

He began his work in order to provide a mechanistic account of animal learning, which could be stated in terms of elementary operations and events, and was concerned to dispel the myth of animal ratiocinative intelligence. This led to his detailed study of what was later called trial-and-error learning.

In a typical experiment a hungry cat or dog was placed in a box similar to that shown in Figure 1-3. The door was pulled open by a weight attached to a string as soon as the animal loosened the bolt or bar which held it shut. A dozen or so variations of the box were used, some having loops of wire, in various positions in the cage, attached by string to latches, others having wooden buttons or catches close to the door, and others with levers or platforms inside the box operating release mechanisms. A final version had the experimenter release the door when the animals engaged in a specific activity, such as licking or scratching themselves.

Initially the animals engaged in much unsuccessful and irrelevant behaviour before the door latch was tripped in an almost accidental manner, and Thorndike provides a vivid account of that behaviour:

Figure 1-3. Thorndike's puzzle box. (Based on Thorndike, 1898)
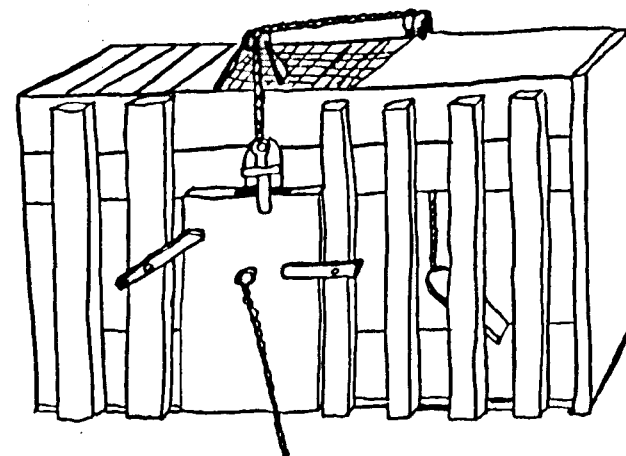
When put into the box the cat would show signs of discomfort and of an impulse to escape from confinement. It tries to squeeze through any opening; it claws and bites at the bars or wire; it thrusts its paws out through any opening and claws at everything it reaches; it continues its efforts when it strikes anything loose and shaky; it may claw at things in the box. It does not pay very much attention to the food outside, but seems simply to strive instinctively to escape from confinement. The vigour with which it struggles is extraordinary. For eight or ten minutes it will claw and bite and squeeze incessantly.... The cat that is clawing all over the box in her impulsive struggle will probably claw the string or loop or button so as to open the door. And gradually all the other non-successful impulses will be stamped out and the particular impulse leading to the successful act will be stamped in by the resulting pleasure, until, after many trials, the cat will, when put in the box, immediately claw the button or loop in a definite way.

The starting point for the formation of any association in these cases, then, is the set of instinctive activities which are aroused when a cat

feels discomfort in the box either because of confinement or a desire for food. (Thorndike, 1898, p.13)

The times taken for the animal to escape were high at first but on succeeding trials times for escape decreased, remaining relatively stable after 10-20 trials. The associative processes, which Thorndike was concerned to study, showed a gradual stamping in of correct responses and stamping out of incorrect responses, according to the commensurate pleasure (or lack of it) resulting from the response. It was from these studies with animals that the first scientific theory of learning emerged, Thorndike's theory of connectionism:

I have spoken all along of the connection between the situation and a certain impulse and act being stamped in when pleasure results from the act and stamped out when it doesn't. (Thorndike, 1898, p.103)

This 'law of effect', as it was to be called, brought motivation and reward to the foreground of experimental psychology. Rewards, or successes and failures, provided a mechanism for the selection of the more adaptive responses, and this bears much resemblance to the mechanism of 'natural selection by successful adaptation', which was the basis for Darwin's theory of evolution. This law augmented the familiar law of habit formation through repetition, for Thorndike, and the two became central to his theories of learning and instruction, when he joined the faculty of Teachers College, Columbia University, in 1899, where he shifted his emphasis from animal to human learning.

The Law of Effect: The Law of Effect is that, other things being equal, the greater the satisfyingness of the state of affairs which accompanies or follows a given response to a certain situation, the more likely that response is to be made to that situation in the future.
The Law of Exercise: All changes that are produced in human intellect, character and skill happen in accord with and as a result of, certain fundamental laws of change. The first is the Law of Exercise, that, other things being equal, the oftener or more emphatically a given response is connected with a certain situation, the more likely it is to be made to that situation in the future.... This law may be more briefly stated as: 'Other things being equal, exercise strengthens the bond between situation and response.' (Thorndike, 1912, pp.95-96)

The kinds of phenomena coming under the law of exercise are those repetitive habits such as rote memorization or the acquisition of muscular skills. The belief being simply that, as the old saying has it, 'Practice makes perfect.' From this it is predicted that learning curves plotting performance against practice trials should increase in accord with the law. Forgetting curves will show the opposite tendency with a loss of skill over various intervals.

A third law, the law of readiness, was an accessory principle which characterizes the conditions under which there is satisfaction or annoyance. It was couched in rather dubious neuro-physiological terms which can be paraphrased as follows:

1. Given the arousal of an impulse to a particular sequence of actions, the smooth carrying out of the sequence is satisfying.
2. If the sequence is blocked, that is annoying.
3. If the action is fatigued or satiated, then forced repetition is annoying.

Here, Thorndike was suggesting that satisfaction and frustration or annoyance depend on the state of readiness of the organism when a particular response occurs.

These laws of connection forming or association or habit furnished education with two obvious general rules, according to Thorndike: (1) put together what should go together and keep apart what should not go together, (2) reward desirable connections and make undesirable connections produce discomfort. In other words, exercise and reward desirable connections and prevent or punish undesirable connections. He felt that although such rules or principles may seem obvious the fact was that an examination of the literature of educational theory, at that time, would prove that they were neglected and misunderstood, and that practical use of them was never made.

However, these simple laws, derived from his early research in the animal laboratory, were placed in the context of the complexities of human learning, where four varieties of learning could be distinguished: (1) connection-forming of the common animal type, as is found when the ten-month-old baby learns to beat a drum, (2) connection-forming of ideas, as when a two-year-old learns to think of its mother upon hearing the word 'mother', (3) analysis or abstraction, as when a student of music learns to respond to an overtone in a given sound, and (4) selective thinking or reasoning, as when a language student learns the meaning of a sentence by using his knowledge of the rules of syntax.

His advice to the teacher was not limited to the application of his major laws. The active role of the

learner, who comes to the learning situation with a constellation of motivational variables was also recognized by Thorndike and he listed five aids to improvement in learning, which he believed were accepted by educators, and which will surely stand scrutiny today (1913, pp.217-226):

1. Interest in the work
2. Interest in improvement in performance
3. Significance of the lesson for some goal of the student
4. Problem attitude in which the student is made aware of a need which will be satisfied by learning the lesson
5. Attentiveness to the work

Unfortunately, his laws of learning did not stand the test of time. To Thorndike's credit, he was prepared to acknowledge his errors and, in 1929, before the International Congress of Psychology in New Haven, he admitted that he was wrong and introduced two fundamental revisions of his laws of exercise and effect.

It became apparent that exercise alone will not necessarily produce improved performance. For example, if a blindfolded subject is asked to attempt to draw a line of a given length, and to do so hundreds of times, over many days, but without information concerning the accuracy of his performance, there is no improvement in the ability to draw the line of correct length during the experiment. Feedback is clearly an essential component . Thorndike (1931) quotes an experiment in which a long series of word-number pairs were presented to the subjects. If repetition alone produced the stamping in of associations then each of the pairs (number-word, word-number) should have been recalled with equal strength. They were not, and it was clear that the subject's mental set and expectations guided the associations formed.

A number of experiments also demonstrated that the effects of reward and punishment are not equal and opposite. Instead, under several conditions, reward was found to be much more powerful than punishment. One such experiment (1932) used a simple maze which gave chicks a choice of three pathways, one of which led to 'freedom, food and company.' The wrong choices led to confinement for thirty seconds. Thorndike concluded that:

> The results of all comparisons by all methods tell the same story. Rewarding a connection always strengthened it substantially; punishing it weakened it little or not at all. (Thorndike, 1932, p.58)

Similar effects were found with human subjects learning

word translations, the rewarded responses (experimenter saying 'Right') led to repetition of the rewarded condition, but the punished condition ('Wrong') did not weaken the connections.

Thorndike went on to collect a series of testimonials about the relative effects of rewards and punishments from published biographies and other sources and found almost universal evidence of the greater beneficial effect of reward than punishment (1935, pp.135-144, 248-255).

These changes in the laws of learning occurred at a time when Thorndike's proposals for the development of a science and technology of instruction were being eclipsed by the rise of the new school of psychology, J. B. Watson's 'Behaviourism.'

His contribution to modern instructional technology, however, cannot be overestimated. He was the undisputed originator of the first scientific theory of learning and, as such, his influence has been profound and long-lasting, not least because of his prescience concerning the potential of technology in the pedagogical process. As early as 1912 he suggested that:

> Great economies are possible by printing aids, and personal comment and question should be saved to do what only it can do. A human being should not be wasted in doing what forty sheets of paper or two phonographs can do. Just because personal teaching is precious and can do what books and apparatus cannot, it should be saved for its peculiar work. The best teacher uses books and appliances as well as his own insight, sympathy, and magnetism. (Thorndike, 1912, p.167)

Saettler (1968) claims that in his three-volumed work, 'Educational Psychology' (1913), Thorndike formulated the basic principles underlying a technology of instruction. In implementing these principles Thorndike suggested that control of the learning may not be the sole responsibility of the teacher and that machines could play a part. He proposed what may be considered the earliest example of programmed learning in his influential text 'Education' (1912):

> If, by a miracle of mechanical ingenuity, a book could be arranged that only to him who had done what was directed on page one would page two become visible, and so on, much that now requires personal instruction could be managed by print. (Thorndike, 1912, p.165)

He was also an advocate of the laboratory method where students examined real things and observed them working rather than merely speculating or arguing about them. He

suggested that there are certain elements of knowledge, certain tendencies to respond, which can only be got by direct experience of real things.

> This fact may, and should, now seem axiomatic, but many teachers in practice forget it, and teachers a few centuries ago rarely thought about it. The increased use of methods whereby the realities are examined and experimented with as well as talked about, enormously improved the teaching of mathematics, science, history and even language. (Thorndike, 1912, p.176)

However, he also advised that verbal methods were not always wrong, nor that a pupil should always be provided with the direct experience itself. He anticipated the later audio-visual movement with his claim that different degrees of reality may be desirable - the actual thing or a model of it, a photograph or a rough sketch or even a map of it (1912, p.177). He also cautioned against the unbridled use of the popular discovery methods and suggested that if the doctrine is taken literally 'it becomes absurd', and that:

> It would be enormously costly thus to deprive children of the advantages of civilization, and would be the height of folly, even if it could be done at no cost. (Thorndike, 1912, p.194)

In fact, in his early work, Thorndike sees much to commend those methods embodying 'telling and showing', such as lectures and demonstrations but also including the use of printed text-books, object lessons, maps, dictionaries and tables of logarithms, all of which are 'clear cases of the lecture-demonstration or telling-showing method' (1912, p.189). He acknowledged the criticisms that the student may see or hear but not understand and that he does not learn to think but concluded that 'the chief excellence of this method is economy.... in some cases this advantage alone justifies its use' (1912, p.189).

Pavlov, Ivan Petrovich (1849-1936)

Pavlov, the father of Russian psychology, initially trained and worked as a physiologist. He gradually changed his interest in favour of psychology following the incidental observation of the phenomena we now know as the conditioned reflex. His influence on psychology has been profound and long-lasting. In fact, Kimble (1961) found that 31 terms relevant to conditioning and learning are directly attributable to Pavlov, and only 21 terms to all American and Western psychologists combined. He was awarded the Nobel Prize in 1904 for 'The Work of the Digestive Glands' (1897), the first Russian scientist, and the first physiologist in the world, to receive the award.

In his student years Pavlov was attracted to the philosophy of materialism, particularly as embodied in the work of Claude Bernard, which suggested that the science of living phenomena must have the same foundations as the science of inorganic bodies and that there is no difference between the principles of biological science and those of physico-chemical science.

A particularly strong influence on Pavlov was the physiologist I.M. Sechenov, who conducted a series of experimental investigations of the nervous system and its reflexes in Bernard's laboratory in 1862. He tried to publish the results of this work in a widely read Moscow review under the title 'An Attempt to Establish the Physiological Basis of Psychical Processes.' However, the tsarist censor realized that Sechenov's intention was to spread materialistic ideas concerning the relationship between physiology and psychic activity, and would only permit publication in a specialist medical journal under the title 'Reflexes of the Brain.' The article represented a daring challenge to the deeply entrenched Cartesian doctrine of psycho-physiological parallelism, that body and mind comprise two completely separate and materially unrelated systems, which somehow run on parallel tracks and, when it eventually appeared in book form, it again produced a reactionary outcry from the Petersburgh Censorial Committee, on the grounds that it destroyed the religious doctrine of eternal life. This was just four years after Darwin's 'Origin of the Species' (1859), which taught that all phenomena have a history including origin and development from lower forms, had shaken the foundation of the creationist religious dogma. Sechenov's book, which supported Darwin's ideas, eventually became a popular classic of the time and Sechenov passed the remaining years of his academic life as Professor of Physiology at Moscow University.

Sechenov proposed that all the immense diversity of psychical phenomena can and must be explained on the basis of the nervous system and the brain, through the mechanism of reflex arcs. He suggested that thought and emotion can be accounted for in terms of the reflex, by analogy with lower forms, but it was Pavlov who provided experimental proof of this with the discovery of the conditioned reflex.

Entering the University of St. Petersburgh to study natural science, Pavlov's early interest in physiology flourished and he received a gold medal for his experimental work on the physiology of the nerves of the pancreas. In 1875 he received his degree, with an outstanding record of achievement. By 1890 Pavlov had perfected a method for

studying the nervous regulation of the blood, with animals trained to undergo all the manipulations of a long and complex experiment without narcosis. From 1891 he was head of the Department of Physiology in the new Institute of Experimental Medicine, and it was here that he conducted his work on the digestive glands. Pavlov found that vivisectional methods were unsuitable for studying the intricate workings of the digestive system and so he perfected the 'chronic' experimental method, begun in his work on the circulatory system. This involved the creation, through highly skilled surgery, of a fistula, an opening or a window which allowed the experimenter to observe the functioning of the gland under controlled conditions, but in a normal healthy animal.

By 1901 Pavlov became engrossed in a problem which he had deferred investigating, but which was to lead to the discovery of the conditioned reflex, to which he devoted the remaining 35 years of his life.

During his experimental work Pavlov had found that gastric juice was secreted by experimental dogs not only when food was introduced into the mouth but also when they saw the food. This action at a distance, through the eyes, ears and nose was generally interpreted in terms of the so-called higher functions and was considered to be far beyond the reach of physiology.

Animal psychology was not on an objective experimental basis at this time, Thorndike's 'Animal Intelligence' had only recently been published and was not available to Pavlov. At first, Pavlov suggested that the observed anticipation and consequent salivation must be acquired with the aid of the higher functions - the animals judgement, will and desires, but he soon distanced himself from such a view:

> To understand these phenomena, are we obliged to enter into the inner state of the animal and to fancy his feelings and wishes as based on our own? For the investigator, I believe there is only one possible answer to the last question - an absolute "No."
> .... In our "psychical" experiments on the salivary glands (we shall provisionally use the word "psychical"), at first we honestly endeavoured to explain our results by fancying the subjective condition of the animal. But nothing came of it except unsuccessful controversies, and individual, personal, inco-ordinated opinions. We had no alternative but to place the investigation on a purely objective basis. (Pavlov, 1928, p.50)

The psychical experiments demonstrated that it was not just the appearance of the food, its odour etc. which

produced salivation, but 'absolutely all the surroundings in which these objects are presented to the dog, or the circumstances with which they are connected in real life'(p.52). For example, the dish, the room, the person who usually fed the animal and the noises produced by him. Pavlov recognized the adaptive value of such reactions to 'remote signs (signals).' In fact, this adaptive reflex was originally termed conditional and not conditioned by Pavlov, in his address on the occasion of receiving the Nobel Prize. However, 'conditioned' reflex has become fixed in English usage. The conditioning procedure is shown in Figure 1-4.

> A further essential difference between the old and the new reflexes is that the former are constant and unconditional, while the latter are subject to fluctuation, and dependent upon many conditions. They, therefore, deserve the name "conditional" (ooslovny). (Pavlov, 1928, p.79)

Figure 1-4. Pavlov's Conditioning Procedure.

Meat Powder (unconditioned stimulus)	-------->	Salivation (response which is an unconditioned response)
Meat Powder + Sound of bell (neutral stimulus)	-------->	Salivation
Sound of bell (conditioned stimulus)	-------->	Salivation, after repeated pairings (conditioned response)
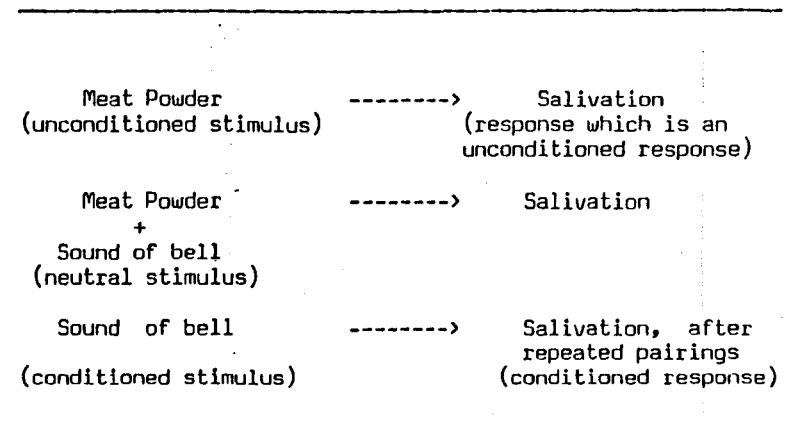
The great discovery made by Pavlov was that any stimulus, even one not remotely connected with an inborn reflex, when paired with a stimulus inextricably linked to an inborn reflex response, could take the place of the original stimulus. In this case, the sound of the bell (originally neutral, producing no salivation) is repeatedly paired with meat powder (the unconditioned stimulus) which calls forth the salivation reflex. Following repeated pairings, the sound of the bell acquires the power to evoke the salivation reflex, even in the absence of meat powder.

It is, in effect, acting as a signal for the meat powder, producing an anticipatory response.

By 1904, Pavlov was discussing the importance of the signals which stand in place of the unconditioned stimulus in the reflex. These signals can have only a conditional significance and are readily subjected to change, for example, when the sound of the bell is not accompanied by meat powder the salivation response is gradually weakened, a process termed 'extinction.' And this is precisely how adaptive advantage is conferred. In the wild a startle or fear response may be evoked in an animal by any number of stimuli presented by a predator, those which are repeatedly present are strengthened as signals, while others of an incidental nature are extinguished.

Pavlov decided that organisms must possess two sets of reflexes. A fixed set of simple, inherited reflexes and a set of acquired, conditioned reflexes formed by 'the pairing of previously neutral stimuli with a stimulus which triggers off a simple reflex. The conditioned reflexes were found to be routed through the cerebral hemispheres, causing Pavlov to suggest that:

> The central physiological phenomenon in the normal work of the cerebral hemispheres is that which we have termed the 'conditioned reflex.' This is a temporary nervous connection between numberless agents in the animal's external environment, which are received by the receptors of the given animal, and the definite activities of the organism. This phenomenon is called by psychologists 'association'....The basic physiological function of the cerebral hemispheres throughout the.... individual's life consists in a constant addition of numberless signalling conditioned stimuli to the limited number of the initial inborn unconditioned stimuli, in other words, in constantly supplementing the unconditioned reflexes by conditioned ones. Thus, the objects of the instincts exert an influence on the organism in ever-widening regions of nature by means of more and more diverse signs or signals, both simple and complex; consequently, the instincts are more and more fully and perfectly satisfied, ie., the organism is more reliably preserved in the surrounding nature. (Pavlov, 1955, p.272-273)

The organism is sensitized to signals in the environment which help in the avoidance of pain and difficulties and lead to the things needed for the preservation of the individual and perpetuation of the species. Pavlov felt that all complex learned behaviour is composed of orchestrated simple reflexes, which are physiological, rather than mental, processes. From this it followed that objective investigations would lead, ultimately, to the prediction of all human and animal behaviour. In this, Pavlov paved the way for an objective science of behaviour which was to appeal to Watson and the behaviourists later in the century.

Watson, John B. (1878-1958)

John B. Watson is generally accorded the distinction of being the founder of 'Behaviourism.' His early work was concerned with ethological studies of the behaviour of birds in the wild, and laboratory learning experiments with white rats in mazes, which earned him his doctorate in 1903. His major impact was made with the so-called manifesto for behaviourism: Psychology as the Behaviourist Views It (1913), in which he was not proposing a new science as such, but was arguing that psychology should be redefined as the 'study of behaviour' rather than the science of the 'phenomena of consciousness.'

> Psychology as the behaviourist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behaviour. Introspection forms no essential part of its methods, nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness. The behaviourist, in his efforts to get a unitary scheme of animal response, recognizes no dividing line between man and brute. The behaviour of man, with all of its refinements and complexity, forms only a part of the behaviourist's total scheme of investigation. (Watson, 1913, p.158)

Just as Thorndike had sought to rid psychology of speculations concerning animal consciousness and thought, so Watson wished to circumvent human consciousness in favour of an objective study of behaviour alone. He wished psychology to turn away from the introspectionist psychology of Wundt and Titchener. Wundt had set up the world's first formal laboratory of psychology in Leipzig and had started the first effective journal for experimental psychology, but his work and that of his followers, including Titchener at Cornell University's laboratory of experimental psychology, failed to live up to Watson's expectations:

> It was the boast of Wundt's students, in 1879, when the first psychological laboratory was established, that psychology had at last become a science without a soul. For fifty years we have kept this pseudo-science, exactly as Wundt laid it down. All that Wundt and his

students really accomplished was to substitute for the word "soul" the word "consciousness." (Watson, 1925, p.5)

Watson felt that one can assume either the presence or absence of consciousness anywhere in the phylogenetic scale, without it influencing the study of behaviour or the behaviouristic experimental method. The emphasis on human consciousness as the centre of reference for all behaviour caused him to draw the analogy with the Darwinian movement and its initial concern with material which contributed to an understanding of the origin and development of the human race. However, the moment zoology undertook the experimental study of evolution and descent, the situation immediately changed and data was accumulated from the study of many species of plants and animals, with the laws of inheritance being worked out for the particular type under investigation. Such studies were of equal value when compared with those dealing with human evolution.

The new brand of psychology was also to provide an answer to the question 'What is the bearing of animal work upon human psychology?' which, Watson admitted, had caused him some embarrassment. He suggested that for a fusion of animal and human studies to occur some kind of compromise was necessary: either psychology would have to change its viewpoint so as to take into account facts of behaviour, whether or not they had bearings upon the problems of consciousness; or else behaviour would have to stand alone as a wholly separate and independent science. Watson declared that should there be a failure of 'human psychologists' to accommodate, the behaviourists would be driven to using methods of investigation comparable to those employed in animal work, with human subjects.

Watson's quarrel was not with the systematic and structural introspectionists alone. It was also with the new styled functionalist school, which was supposed to throw light on the biological significance of consciousness instead of its analysis into structural elements. Watson found, having done his best to understand the difference between structural and functional psychology, that 'instead of clarity, confusion grows in me.'

He was concerned, ultimately, to unite such diverse studies as the paramecium response to light, learning problems in rats and plateaus in human learning curves, but in each case by direct observation and under experimental conditions. This psychology would be undertaken in terms of stimulus and response, habit formation, habit integrations and the like. It would take as a starting point, first, the observable fact that organisms, man and animal alike, adjust to their environments by two means: hereditary and habit. Secondly, that certain stimuli lead organisms to make

responses. Such a system of psychology, Watson believed, when completely worked out would enable the stimulus to be predicted, given the response; or, given the stimulus, the response could be predicted. Watson makes clear what it is he really requires of psychology:

> In the main, my desire in all such work is to gain an accurate knowledge of adjustments and the stimuli calling them forth. My final reason for this is to learn general and particular methods by which I may control behaviour. (Watson, 1913, p.168)

If psychology were to take such a path there would be practical benefits, which Watson put into effect in 1920, when he was forced to leave the academic life following a scandalous affair with a former pupil: he became an advertising man. He felt that, with such a psychology, the educator, the physician, the jurist and the businessman could utilize the data in a practical way. In pedagogy, for example, the psychologist may endeavour to find out 'by experimentation whether a series of stanzas may be acquired more readily if the whole is learned at once, or whether it is more advantageous to learn each stanza separately and then pass to the succeeding one' (p.169). The psychologist would not attempt to apply the findings, they would merely be made available to the teacher, application of the principle being purely voluntary.

But who would fail to apply the principles and powerful techniques which Watson had in mind for his new psychology? Conditioning was to play a significant role in this psychology as Watson made clear two years after his 'somewhat impolite papers against current methods in psychology' (1916, p.89). He felt it incumbent upon himself to suggest some method which might be used in place of introspection, following his criticisms. Conditioned reflexes were claimed to be suitable for providing the possibility of objectively approaching many problems in psychology. Using such methods Watson claimed that he gave no more instruction to the human subjects than he gave to the animal subjects; nor did he care what language the subject spoke or whether he spoke at all. The conditioned reflexes enabled Watson to 'tap certain reservoirs which have hitherto been tapped only by the introspective method.'

It was, however, Watson's conditioning of Little Albert which has attracted most attention, and has been enshrined in the history of psychology (Watson and Watson, 1921). Albert was eleven months old when he first entered Watson's laboratory. A small, white rat was shown to the boy, whose natural tendency was to stretch out his hands to touch and play with the animal. Watson documented this response, identifying it as that of a normal child, who is curious and

reacts positively to the rat. A few days later the boy and the rat were brought together in the same room, but with a rather bizarre difference. Each time the child reached for the rat Watson banged a long steel bar very loudly with a carpenter's hammer, causing the infant to jump violently and fall forward 'burying his face in the mattress.' After a short interval the boy again reached out to the animal - and Watson repeated his routine with the hammer. 'Again the infant jumped violently, fell forward and began to whimper.' Little Albert was so disturbed that no further tests took place for a week.

At first, when boy and rat were reunited, Albert did not reach out for the rat. Eventually, when the rat was placed near to him, he began to reach out towards it, very gingerly, then he began playing with blocks in a normal way. Watson now showed Albert the rat three times and accompanied each presentation with the same loud noise. The child puckered his face, whimpered and withdrew his body. This procedure was repeated on a further three occasions, until as soon as the child saw the rat he began to cry and crawl away so rapidly that he was 'caught with difficulty before he reached the edge of the mattress.'

Watson claimed that this demonstrated an explanatory principle of how our emotions originate. The conditioned response generalized to other animals, including a rabbit and a dog, and even a fur coat, and survived intact for more than a week. Watson did intend de-conditioning the child, but Little Albert was adopted by a family living some distance from the laboratory, and may, even at the time of writing, survive with his fear intact!

Building on his success with Little Albert and other similar conditioning and de-conditioning experiments, and following on from the claims made by Pavlov, which suggested the powerful influences of the environment on an organism's behaviour through the agency of conditioning, Watson proposed a daring challenge, epitomizing the empiricist tradition:

> Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select - doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors. (1925, p.82)

The work of Thorndike, Pavlov and Watson made psychology a respectable science in which phylogenetic continuity was an essential ingredient, permitting extrapolation from the results of animal experimentation to the complex domains of human behaviour. And nowhere was this more apparent than in the emergence of B.F. Skinner's radical behaviourism, and its educational manifestation: the linear teaching machine.

REFERENCES

Cohen, D. (1979) J.B. Watson: The Founder Of Behaviourism. London: Routledge and Kegan Paul.

Dale, E. (1954) Audiovisual Methods In Teaching. London: Holt, Rinehart and Winston.

Hawkridge, D.G. (1981) 'The telesis of educational technology', British Journal of Educational Technology, 12(1), 4-18.

Hilgard, E.R. (1964) Theories Of Learning and Instruction. Chicago: University of Chicago Press.

Jonassen, D.H. (1985) 'Learning strategies: a new educational technology', Programmed Learning And Educational Technology, 22(1), 26-34.

Kimble, G.A. (1961) Hilgard And Marquis' Conditioning And Learning. New York: Appleton-Century-Crofts.

Pavlov, I.P. (1897/1902) The Work Of The Digestive Glands. London: Griffin.

Pavlov, I.P. (1928) Lectures On Conditioned Reflexes. (Translated by W.H. Gantt) London: Allen and Unwin.

Pavlov, I.P. (1955) Selected Works. Moscow: Foreign Languages Publishing House.

Piaget, J. (1970) Genetic Epistemology. New York: Columbia University.

Saettler, P. (1968) A History Of Instructional Technology. New York: McGraw-Hill.

Thorndike, E.L. (1898) 'Animal Intelligence', Psychological Review, Monograph Supplement, 2, No. 8.

Thorndike, E.L. (1909/1962) 'Darwin's contribution to psychology', University of California Chronicle, 12, 65-80. In G.E. Joncich (Ed) Psychology And The Science Of Education. New York: Columbia University.

Thorndike, E.L. (1912) Education: A First Book. New York: Macmillan.

Thorndike, E.L. (1913) Educational Psychology. New York: Teachers College, Columbia University.

Thorndike, E.L. (1931/1966) Human Learning. Cambridge: MIT Press.

Thorndike, E.L. (1932) 'Reward and punishment in animal learning', Comparative Psychological Monographs, 8(9).

Thorndike, E.L. (1935) The Psychology Of Wants, Interests And Attitudes. New York: Appleton-Century-Crofts.

Watson, J.B. (1913) 'Psychology as the behaviourist views it', Psychological Review, 20, 158-178.

Watson, J.B. (1916) 'The place of the conditioned reflex in psychology', Psychological Review, 23(2), 89-117.

Watson, J.B. (1925) Behaviourism. London: Kegan Paul, Trench and Trubner.

Watson, J.B. and Watson, R. (1921) 'Studies in infant psychology', Scientific Monthly, Dec., 493-515.

Wittich, W.A. and Schuller, C.F. (1953) Audio-visual Materials: Their Nature And Use. N.Y.: Harper.

CHAPTER 2

THE ART OF TEACHING AND THE SCIENCE OF LEARNING

B.F. Skinner took up the ideas of behaviourism in the 1930's and after extensive experimental research, much of it based on the performance of laboratory rats and pigeons, extended his conclusions to the higher reaches of the phylogenetic chain, to human behaviour.

Skinner's success is attributable to his isolation of a few highly repeatable phenomena in conditioning - many of which Pavlov and Thorndike had studied and named before him (Hilgard and Bower, 1975) - which were then used as a basis for analyzing more complex forms of behaviour. Following on from Watson's psychology, Skinner continued to reject mentalistic or cognitive explanations of behaviour and was interested only in the observable and repeatable.

Operant Conditioning

His early work made extensive use of a piece of apparatus which is now commonly termed the 'Skinner box', shown in Figure 2-1. In such a box, the caged animal (usually a rat or pigeon) presses a lever or pecks at a key and receives food pellets as a reward. It is from this experimental base that Skinner developed the linear teaching machine. A close look at the procedures involved in this 'operant conditioning' is required to appreciate the evolution of such machines, and subsequent criticisms of their use.

The type of conditioning studied by Skinner is similar to the trial and error learning discussed by Thorndike (1898). This differs from Pavlov's procedure in that the confined animal emits the response (presses the lever) which is, according to Skinner, under 'voluntary' control. Whereas classical or Pavlovian conditioning is 'involuntary' and the response is under the control of the autonomic system; it consists of elicited behaviour, such as salivary secretion or pupillary constriction to light.

At the start of the operant conditioning procedure for a rat an unconditioned, hungry animal is placed in the Skinner box. Initially, the rat is encouraged to associate the sound of the food magazine with the delivery of food pellets in the food tray. Pellets are discharged from the magazine periodically by the experimenter to encourage this. During the preliminary training the lever is disconnected from the magazine and at this stage the delivery of food is not contingent upon any behaviour exhibited by the rat.

Figure 2-1. The Skinner Box.
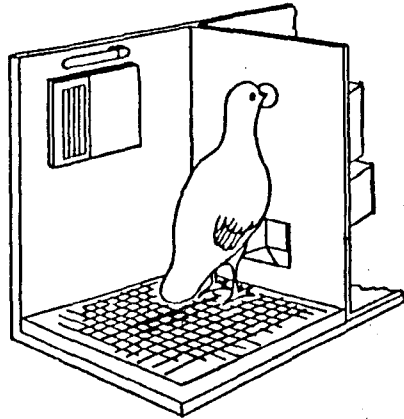
When the rat has been adapted to the environment and moves to the food tray each time it hears the magazine in operation, operant conditioning can begin. The lever is connected to the food magazine so that each time the lever is depressed by the rat a pellet of food is dispensed at the food tray. The rat can then be left to explore the environment. It will show a rather low rate of lever pressing as it explores, but not a zero rate. It will press the lever a few times during an hour at a rather irregular rate (Skinner, 1938) and in doing so will receive the pellet of food, which will act as a reinforcer of the behaviour which immediately preceded it (the pressing of the lever), making it more likely to be repeated. Eventually, the relationship between lever pressing and dispensing of food is made and the rat presses the lever at an increasing rate, as described by Skinner:

> On the first day of conditioning a first response was made five minutes after release. The reinforcement had no observable effect upon the behaviour. A second response was made 51.5 minutes later, also without effect. A third was made 47.5 minutes later followed by a fourth 25 minutes after that. The fourth response was followed by an appreciable increase in rate showing a swift acceleration to a maximum. The intervals elapsing

before the fifth, sixth, and following responses were 43, 21, 28, 10, 10 and 15 seconds respectively. (Skinner, 1938, p.67-68)

Shaping Behaviour

Not all behaviours can be considered to be discrete units and for some behaviours the experimenter may have to wait hours, days or even weeks before they emerge. The original probability of the response in its final form is very low; in some cases it may even be zero. Fortunately, there is a procedure for producing the response more quickly, the experimenter takes control of the food magazine by disconnecting the lever from the food magazine and rewards successive approximations to the desired response.

> To get the pigeon to peck the spot as quickly as possible we proceed as follows: We first give the bird food when it turns slightly in the direction of the spot from any part of the cage. This increases the frequency of such behaviour. We then withhold reinforcement until a slight movement is made towards the spot.... We continue by reinforcing positions successively closer to the spot, then by reinforcing only when the head is moved slightly forward, and finally only when the beak actually makes contact with the spot. We may reach this final response in a remarkably short time. (Skinner, 1953, p.92)

This procedure is termed 'shaping' the behaviour and it enables the experimenter to bring a rare response to a very high probability in a short time. Having established the response the pigeon will continue pecking the key and receiving food until sated. The animal is usually reduced to 80% of its normal body weight in order to provide the motivation necessary to maintain the lever pressing or key pecking.

Extinction of a Response

If the lever is disconnected from the food magazine the animal's response rate will decline, a property noted by Pavlov, which he termed 'extinction' of the response. This may be used to produce a differential response, as is the case in discrimination training. For example, a pigeon can be trained to peck at a key only when it is illuminated by light of a particular colour (green), but not when it is another colour (red). Initially , it is trained to peck at the unilluminated key until a stable rate is observed. The

training then proceeds with a food pellet being delivered for each peck only when it is green. When it is red the animal receives no reinforcement. This soon produces a differential rate of pecking: a high rate for green and an increasingly lower rate to red as the response extinguishes. The special stimulus which must be present in order for a reinforcement to occur (green key) is called a Discriminated Stimulus or S^D.

Schedules of Reinforcement

When each response receives a reinforcement the procedure is termed 'continuous reinforcement'. Of course, under such a schedule of reinforcement the animal soon becomes sated and refrains from responding at a high rate until it is again hungry. Skinner discovered that a high rate of responding could be maintained under a variety of reinforcement schedules, without the necessity to reward each and every response, a technique that 'gets more responses out of an organism in return for a given number of reinforcements' (Skinner, 1953, p.99).

He distinguished between schedules which are controlled by the system outside the organism and those controlled by the animal's behaviour. An example of the first is a schedule based on the passage of time, eg. the animal receives a reinforcement every 5 minutes. When a reinforcement is given every fiftieth response, it is an example of the second type of control. Both types of these 'intermittent schedules' can be either fixed or variable. Thus, a reinforcement in a fixed interval schedule is given after the set time has elapsed, eg. every five minutes; with a variable-interval schedule the reinforcement is given on average after a set time, eg. on average every five minutes, with the intervening interval varying between a few seconds or as long as ten minutes. This type of schedule produces behaviour which is extraordinarily persistent and very difficult to extinguish. A fixed ratio schedule also produces a very high rate of response, but the most powerful of all the schedules is the variable-ratio, where the reinforcement occurs after a given number of responses on average, the pigeon or rat becoming a victim of an unpredictable contingency of reinforcement.

Cultural Engineering

The results of Skinner's early work were published in his 'Behavior of Organisms' (1938), but his next major work is a somewhat startling contrast, until it is realized that Skinner's first career was journalism. 'Walden Two' (1948),

a novel, was written as a programme that outlined the utopia attainable by adopting a behaviouristic, experimental approach toward social problems. The scientists in 'Walden Two' dispensed with traditional 'trial-and-error' freedom because there were efficient ways of attaining desired goals, based mainly on positive reinforcement. Cultural engineering was not a pejorative term in this brave new society, although the book was frequently read as a dystopia. A more technical treatment of the 'Walden Two' ideas is to be found in 'Science and Human Behaviour' (1953), dedicated to his colleague Fred Keller, who later developed the Personalized System of Instruction (see Chapter 4), based on Skinnerian principles. In the section dealing with education, he refers to 'knowledge' as 'the entity which is traditionally said to be maximized by education' (p.408). He acknowledges that it is a complex term, but continues with the assertion that most knowledge acquired in education is verbal. He claims that the acquisition of such knowledge does not mean that education is merely rote learning, because the student also comes to understand the facts. For example, in the study of history this is because:

> The individual agrees with a statement about a historical event in the sense that he shows high probability of making the statement himself. The growing understanding with which he reads and re-reads a passage describing a period in history may also be identified with the growing probability that he will emit verbal responses similar to those which comprise the passage. (Skinner, 1953, p.409)

The Appliance of Science to Teaching

With such a view, linking verbal behaviour to knowledge as the entity to be maximized in education, the stage is set for one of the most important announcements in the history of educational technology. 'The Science of Learning and the Art of Teaching' (1954) was first read at a conference at the University of Pittsburgh, in March, during which there was a demonstration of an experimental teaching machine. It was published later in the year in the influential 'Harvard Educational Review.'

Skinner claimed that it was a great shock to turn from the exciting prospect of an advancing science of learning to that branch of technology which is most directly concerned with the learning process, education. In the first place he questioned the reinforcements used, though he did acknowledge that they had changed since the early years of the century and were not all based on major aversive

methods of control, now they were based on 'minor' aversive events such as the teacher's displeasure, low marks etc. He then asked how the contingencies of reinforcement were arranged, for example, when is an arithmetical operation reinforced as 'right'? The conclusion was that many minutes and in many cases many hours or even days may intervene between a child's response and the teacher's reinforcement, even though it could be demonstrated that, unless explicit mediating behaviour has been set up, the lapse of only a few seconds between response and reinforcement destroys most of the effect. He calculated that during the first four years of education 50,000 reinforcements would be necessary to attain efficient mathematical behaviour, but that in a traditional class situation it would be possible for the teacher to provide a pupil with only a few thousand. In order to provide sufficient reinforcements the only possible conclusion was that human learning would require instrumental aid, and the simple fact was that 'as a mere reinforcing mechanism, the teacher is out of date' (Skinner, 1954, p.94).

With the introduction of instrumental aid complex patterns of behaviour could receive gradual elaboration, with the whole process of becoming competent in any field being divided into a very large number of very small steps. The maintenance of the behaviour in strength at each stage could be accomplished by applying the techniques of scheduling which had been developed in the studies of other animals, but more probably would be most effectively arranged through the design of the material to be learned.

> By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the aversive consequences of being wrong are reduced to a minimum. (Skinner, 1954, p.94)

Skinner demonstrated two prototype machines at Pittsburgh. The first taught arithmetic, with the material to be learned appearing in a square window. Missing numbers were made to appear by moving sliders with numbers printed on. When the problem was completed the child turned a knob on the front of the machine. The machine 'senses the composed answer, and if it is correct, the knob turns freely and a new frame of material moves into place.' It would not turn if the answer was incorrect and the child could then re-set the sliders until a correct answer was obtained. Errors were recorded automatically. The second machine was similar, but with more sliders and was designed to teach both arithmetic and spelling.

Skinner foresaw objections to the use of such devices in the classroom, including the challenge that children were being treated as mere animals and that an essentially human intellectual achievement was being analysed in unduly mechanistic terms. He countered this by suggesting that the behaviours elicited by the instrumental aids were the 'very behaviours which are taken to be the evidences of such mental states or processes' and that the 'behaviour in terms of which human thinking must eventually be defined is worth treating in its own right as a substantial goal of education' (1954, p.96).

With regard to possible costs, he argued that a country which annually produces millions of fridges, automatic dishwashers and washing machines can certainly afford the equipment necessary to promote high standards of education.

Echoing Thorndike he stated that the teacher has a more important function than to say 'right' or 'wrong' and that 'it is beneath the dignity of any intelligent person' to mark a set of papers in arithmetic. The more important work, the teacher's relations with her pupils, cannot be duplicated by a machine, and instrumental help would improve these relations by freeing the teacher. He concluded:

> There is a simple job to be done. The task can be stated in concrete terms. The necessary techniques are known. Nothing stands in the way but cultural inertia. But what is more characteristic of the modern temper than an unwillingness to accept the traditional as inevitable? We are on the threshold of an exciting and revolutionary period, in which the scientific study of man will be put to work in man's best interests. Education must play its part. (Skinner, 1954, p.97.)

With the new technology, which was to apply the newly emerging science of learning to the age-old art of teaching, we can see how Skinner drew parallels between his work with animals and the action of an ideal teaching environment. As with the animals in the Skinner box the human learner must make a response which is observable and can be described in operational terms. The learner's behaviour is shaped by the programme, which consists of discriminated stimuli (programme frames), with the correct responses being immediately reinforced (knowledge of results), on a continuous schedule. Incorrect responses are extinguished, because they receive no reinforcement.

Within a few years there was such interest in programmed learning that Schramm (1964) in his review of the research concluded that 'no method of instruction has ever come into use surrounded by so much research activity.' By 1957 Skinner was demonstrating a machine in which a frame of material appeared in a window near the centre of the machine and the student wrote his answer in a window to one side of the frame. By moving a lever the response made by the student moved under a transparent cover (it could be seen,

but not altered, making the device cheat-proof) and at the same time the correct response was uncovered in the upper corner of the frame. If the student recognized that his response was correct he moved another lever which punched a hole next to the correct response and shifted the programme on to the next frame. Such machines were used to teach a course in human behaviour to Skinner's Harvard University students, with nearly 200 students completing the course.

Skinner compared the machine to a private tutor because there is constant interchange between programme and student, unlike lectures and textbooks. The programme 'insists' that a given point be thoroughly understood and presents just sufficient material for which the student is ready, thus ensuring that the student comes up with the correct answer, which receives positive reinforcement and holds the student's interest, according to Skinner.

It was eventually realized that the essential feature of the new method of teaching was the 'programme' rather than the machine and this led to the development of the programmed linear text in which 'frames' were printed on one side of the page and answers on the reverse, a form used in Holland and Skinner's influential text 'The Analysis of Behavior: a program for self-instruction' (1961).

The Research: Does It Work?

The major research question in the decade that followed Skinner's announcement of the forthcoming revolution in education was 'Do students learn from programmed instruction?' Claims were made that student performance could be increased, giving 90% of students scores of 90% on evaluation tests, and the commercial stakes were high, particularly for companies with educational interests. Schramm (1964) attempted to answer this question and located 190 reports of original research on programmed instruction, which included some experiments with programmed films and television. Most of the research (40%) was conducted with college students as subjects, with smaller numbers of secondary (20%) and adult/military (20%) subjects and a minority of primary/pre-school (10%) subjects. From this mixture of research Schramm concluded that many different kinds of students - college, high school, secondary, primary, pre-school, adult, professional, skilled labour, clerical employees, military, deaf, retarded, imprisoned - do indeed learn from linear programmes either on machines or as texts. Using programmes such students learned mathematics and science at different levels, foreign languages, spelling, electronics, computer science, psychology, statistics, business skills, reading, and many other subjects. For each of the groups of students and the

different kinds of subject matter the early experimental evidence demonstrated that 'a considerable amount of learning can be derived from programmes' (Schramm, 1964, p.4) either by comparing pre- and post-tests or the time and trials to reach a given criterion of performance.

When the question is changed to 'How well do students learn from programmes as compared to how well they learn from other kinds of instruction?' Schramm could not answer quite so confidently. Of the 36 studies which compared programmed instruction with traditional teaching, 18 showed no significant statistical difference when the two groups were measured on the same criterion test. But 17 showed a statistically significant superiority in favour of the programmes, with only one showing a difference in favour of classroom students. Schramm cautiously concluded that 'the results should not discourage us about the amount of learning derived from programmes' (p. 5).

In addition to the question of comparative effectiveness researchers were also investigating the nature of the components of the teaching programme. There are six components of an efficient linear programme (1) an ordered sequence of items, through which the student works, (2) at his own pace, in (3) short steps (4) making few errors in his (5) written, constructed responses, to which he receives (6) immediate knowledge of results. Each has received attention in experimental evaluations of the effectiveness of programmed instruction.

A Logical Sequence?

Five interesting experiments compared logically sequenced programmes with a presentation of the items in random order and, surprisingly, three of the experiments showed no difference, one showed no difference on a retention test and the remaining comparison showed a difference in favour of the logical structure. The results were obtained with short programmes and it was concluded that longer programmes would require a more ordered sequence. Gagné and his collaborators (Gagné, 1962a; Gagné and Paradise, 1961; Gagné et al, 1962) analyzed a number of tasks (solving algebraic equations) and found evidence for hierarchies of subordinate learning sets, thus providing a basis for logical structure within programmes. A subordinate learning set being defined as what the subject needed to know in order to be able to perform a new task. However, Mager (1961a) complicated matters by demonstrating that when students in an electronics course were given a free hand to learn by questioning an instructor, they took a path which bore little resemblance to the logical sequence in which the course was taught. The instructors tended to move from parts

to wholes, whereas the students typically moved from smaller wholes to larger wholes. A seemingly straightforward issue was not resolved by these competing results. Mager captured their essence when he claimed that:

> Where it is necessary to teach one thing before another, do so. But be careful! There isn't as much reason for this kind of sequencing as instructors like to think. (Mager, 1961b)

Later, McKeachie (1974a) also commented on the ability of learners to withstand quite considerable distortions of the original programme sequences, suggesting that the motivational effect of 'surprisingness' is sometimes neglected and that 'randomness may not make sense conceptually but it may be more fun than a sequence so logical and with steps so small that one knows exactly what is coming next.'

A Self-paced System?

It was assumed that a major advantage of the new technology would be that the student could proceed at her own pace. Surprisingly, the early research failed to confirm this. Seven of the studies found no difference between external and individual pacing, with students taught by machine, text or television, compared with two studies in favour of individualised pacing. Carpenter and Greenhill (1963) varied the pace from 20 percent below to 10 percent above the average of class self-pacing without decreasing the overall amount of learning. Hartley (1974), in a later review of the research from 1954-1974, also found a wealth of non-significant differences in test performance between individual and group learning in either paced or unpaced situations, and concluded that the research indicated three main points:

(i) Self-pacing produces considerable administrative difficulties, because even like-ability groups spread out enormously in terms of time taken to complete a programme.
(ii) Group-pacing methods are often technically complex to set up: the gains (if any) may not be worth the technical effort involved.
(iii) Some learners have difficulty in judging what is an appropriate pace at which to go. (Hartley, 1974, p.281)

The last point has implications for programme writers because, although the learner can control the speed at which

he works through the programme, the overall pace is largely determined by the writer.

Short Steps?

Skinner proposed that programmes should have short steps, which would lead to few errors in the learning sequence and, initially, this was controlled by the physical size of the frame window in a Skinner machine. Much of the early evidence supported short step size. Evans, Glaser and Homme (1960) tested programmes teaching conversion to unfamiliar number bases, with 30, 40, 51 and 68 steps and found the 51 and 68 step versions to be superior to the others on both immediate and delayed retention. Coulson and Silberman (1960) obtained similar results for programmes teaching psychology items. But, Smith and Moore (1961) found no difference when teaching spelling by means of programmes of 1, 128, 830 and 546 steps.

One of the most fascinating series of early experiments employed training films, such as the assembly of an automobile ignition distributor (Maccoby and Sheffield, 1958; Margolius and Sheffield, 1961; and Weiss, Maccoby and Sheffield, 1961). The length of film viewed before permitting practice was varied in each experiment and it was found that more learning came from gradually increasing step size rather than from maintaining short or long steps. When the students were given control of the film sequence, they also gradually increased the length. Students only permitted to practice short steps showed a gradually deteriorating performance, adding to a body of evidence suggesting student impatience and boredom with long programmes having only short steps.

Although the balance of the early studies was in favour of short steps, Hartley suggested that the following issues were also important:

(i) the ability of the learners involved (university students found small steps irritating)
(ii) the age of the learners (small steps were more appropriate for small children)
(iii) the pre-knowledge of the learners (small steps were more suitable than large steps for learners with little relevant background)
(iv) the confidence of the learners (small steps were more suitable for learners who were afraid of the task in hand - eg. statistics for female arts undergraduates)
(v) the kind of subject matter (small steps might be useful for statistics, or for subjects with their own built-in logic, but not for more over-

inclusive subjects such as, for example, literary appreciation)

(vi)   the language employed (complex language can still be confusing even in small steps).
(Hartley, 1974, p. 282)

An Active Response?

One of the major differences between operant and classical Pavlovian conditioning was claimed by Skinner to be the active, 'voluntary' response of the animal to a given stimulus. This aspect of the early animal research is included in programmed instruction by the demand for an overt, observable, constructed response, usually a written response to a stimulus item in the programme. Unfortunately, this characteristic of the linear programme was by no means unequivocally supported by the early research. The great majority of the studies showed no statistically significant differences between the amount learned from overt and covert responses. Schramm identified 16 research investigations demonstrating this result. Since there was no difference, and since covert or 'thinking' the response took less time, the covert mode could be considered more efficient. A minority of cases demonstrated some difference in favour of the overt, constructed response, mainly for complex subject matter (Cummings and Goldstein, 1962).

Pressey (1963) offered a direct challenge to the Skinnerian programme ideal when he tested the first unit of the Holland-Skinner psychology programme (54 frames) against the same material re-written in good prose paragraphs. There was no difference in the final performance, but the prose version took less time for the students. Schramm concluded that the results 'are going to cause researchers in this field to do a great deal of thinking about the principles behind programmed instruction' (p. 10). Of course, when the step size is such that error levels are reduced to a minimum, overt responding may become superfluous.

Leith (1968) demonstrated that the need for overt responding depended on the type of task and the background of the learner. Conceptual tasks, for example, are clearly different from response-learning tasks. English spelling was found to be better accomplished by writing the responses, while co-ordinate geometry or the structure of genetic materials were learned more successfully when covert responses were required. Even recall, retention and transfer of complex circuit diagrams were carried out just as efficiently by thinking the responses as by drawing them if the students had a background of 'O' level Physics. Those students with less than 'O' level knowledge of Physics did poorly when thinking the responses but did as well if allowed to make overt responses.

Hartley (1974) concurred, adding that overt responses may be superior when the learners were young children, when the material was difficult and the programmes lengthy and when novel or specific terminology was taught.

Immediate Knowledge of Results?

The majority of the early studies did support another of the central tenets of Skinnerian programming, the provision of immediate knowledge of results (Schramm, 1964, p.10). Meyer (1961) found significantly more learning from immediate knowledge when compared with a group that waited for the next meeting to receive results. However, not all the research demonstrated advantages for immediate knowledge of results. Glaser and Taber (1961) did not find any difference when the knowledge of results was varied from 100 percent continuous knowledge of results to 50 or 25 percent knowledge, or when it was received in a variable ratio. They concluded that if the probability of error is low, as in typical linear programmes, knowledge of results is less important than when the probability of error is high.

Later studies confirmed the complexity of the issue. Grundin (1969) found that overt responding interacted with frequency of feedback, making the knowledge of results superfluous, or even detrimental. When learners were required to work out each stage of a correction for themselves, receiving feedback stage by stage, Tait, Hartley and Anderson (1973) found that they performed better than learners receiving more global feedback. McKeachie (1974b) commented that

Programmed instruction proponents were understandably aghast to find that immediate knowledge of the correct response (expected to be a reinforcer) did not facilitate learning in programmed instruction. (p.8)

Intrinsic Programming

One major change in programming styles, which contested the mode of responding, size of step and type of feedback, was introduced by Crowder in the late 1950's. This intrinsic or branching style of programming presents larger amounts of information, usually a paragraph, for the student to study. There is an immediate test, usually a multiple-choice question, and the test result is used to determine the content of the next frame in the teaching sequence. If the answer is correct, she is automatically given the next unit

of information and the next question. But if the answer is incorrect the preceding unit of information is reviewed, the nature of the error is explained and she is retested (Crowder, 1960). There is a separate set of correctional materials for each wrong answer that is included in the multiple-choice test.

This 'intrinsic' programming technique was devised to operate with a sophisticated device for handling programmed materials on microfilm, but could equally be applied to what became known as 'scrambled' texts, in which each answer has a different page number associated with it. The student turns to the page number associated with his answer, which may contain the next unit of information, if the answer is correct, or the correctional information followed by the original question, if the answer is wrong.

Crowder claimed that human learning takes place in a variety of ways and that these ways vary with the abilities and knowledge of the students, the nature of the subject matter etc. The intrinsic programming method would provide the necessary feedback control in this complex series of interactions. This is different to simply providing knowledge of results to the student because the test result is used to control the behaviour of the teaching machine, the primary purpose being to determine whether the communication was successful, in order that corrective steps may be taken if it has failed (Crowder, 1960, p.288).

The term 'intrinsic' programming was used to indicate that the necessary programme of alternatives was built into the material itself such that no external programming device was required, in contrast to the early experiments which were using computers as 'extrinsic' programming devices.

Schramm found few research evaluations of the intrinsic methods, but those which were available showed no differences between the two methods. Hartley (1974) agreed with Schramm's overall conclusion concerning the lack of differences between linear and branching programmes, but observed that some studies demonstrated a saving in time taken to complete the programme, in favour of branching programmes, with older and intelligent learners. This strikes at the heart of Skinnerian psychology, because Skinner had argued most strongly for maximizing the positive consequences of learning and a minimizing of the negative aspects.

Few Errors?

The larger frames of information and the method of responding in Crowder's method naturally allows for more errors in the programme, but in this case this is seen to be a positive virtue. Leith (1968) demonstrated that for 10 and

16-year-olds there was no relationship between success on test performance and numbers of errors made. This was confirmed by Elley (1966). A further study by Leith showed that making and overcoming errors may sometimes help learning. Children were taught to calculate in bases other than ten. One group learned four different bases, another learned two (with twice as much practice on each) and a third group learned only one base, but with four times the practice. Each group read the same number of frames. The group with two bases made most errors when they changed to the second base, but were best on a transfer test to an entirely new base.

Having determined that research had indicated that each of the original characteristics was not a sine qua non of programmed learning, Hartley concluded that 'Skinner's techniques have not been shown to be universally valid. They have not, however, been shown to be valueless' (1974, p.286).

The Art of Scientific Application

The results of these early research studies do seem to cast doubt on most of the characteristics of linear programmes, which were derived from Skinner's psychological theories. This may well be because the translation of psychological principles into educational practice has not always been as precise or accurate as Skinner claimed (Bugelski, 1971). For example, linear programmes embody the Skinnerian principle of shaping behaviour, the student being led step by step to the final goal. But in normal laboratory usage shaping refers to a sequential act. The pigeon is trained to turn clockwise in a circle and then to peck at a key and then climb on a perch. When training is completed the acts are carried out in succession by the animal, but this is not so for the student who ends the programme by carrying out a completely new act, one which he may never have carried out before. The student does not go through the entire sequence of steps in the programme when performing the last operations.

Thorndike had used the words 'right' and 'wrong' as rewards and punishments, with the consequence that knowledge of results was adopted as a perfectly satisfactory secondary reinforcer in teaching machines. To Bugelski (1971) the use of knowledge of results as a reinforcer makes it impossible for anyone to determine whether the teaching machine really represents an application of a psychological principle. Indeed, it must be remembered that the animals in Skinner's experiments were reduced to 80% of their normal weight; they were in a deprived state before engaging in the required behaviour and once satisfied stopped performing.

There is also evidence that different groups of students react in different ways to the type of reinforcement or feedback. McKeachie (1974b) reports a number of studies showing that informing a child of the correctness of his response increases achievement for middle class children, while other children may learn more effectively when given praise or tangible rewards. Means and Means (1971) found that below average students did well on a mid-term exam when told that they had previously performed well on an aptitude test; whereas above average students did better when told that they had done poorly on the aptitude test.

Later Research: Meta-analysis

Three decades and more further on there is little discussion concerning the merits of one form or another of programmed instruction. Research peaked in 1967 and has been declining steadily since that date, although there continues to be interest in individualised modes of learning and the old techniques of programming have been given a new lease of life in some applications of computer-based learning.

There are, however, new techniques for evaluating the results of the many early and varied studies of programmed instruction, which attempt to provide estimates of the magnitude of the benefits to be gained from such educational innovations and interventions. Kulik, Cohen and Ebeling (1980) have used a statistical procedure called 'meta-analysis' (Glass, 1976) for the statistical analysis of a large collection of results from individual studies for the purpose of integrating findings. In a traditional review an innovatory method may be shown to be better than a more conventional approach in most cases, but such a review does not say whether the new method wins 'by a nose or a walkaway' (Glass, 1976). Kulik, Cohen and Ebeling calculated the size of the effect of the programmed instruction intervention, using as the index Glass's Effect Size (ES), defined as the difference between the treatment (programmed instruction) and control (traditional classroom) mean scores, divided by the standard deviation of the control group. An effect size of 1.0 would indicate that the difference achieved by the new method is equal to an increase of one standard deviation of the control group's score, shifting the position of an average pupil in the control group from the 50th percentile point to the top 20% of the class. An effect size of 0.8 and above demonstrates a large effect of the experimental treatment, which certainly has educational significance; an ES of 0.2 and less is small and educationally insignificant; an ES of 0.5 is a medium effect and the experimental treatment warrants further investigation and consideration.

Kulik, Cohen and Ebeling (1980) found 56 studies comparing programmed instruction with conventional methods in higher education. A majority of these favoured programmed instruction (40), the remaining 16 favouring conventional methods. Only 25 studies reported statistically significant differences between methods, with 21 favouring programmed instruction. At first sight this seems to produce a very favourable result for programmed instruction. However, when Glass's Effect Size is calculated it is only 0.24. In other words, the effect of programmed instruction in a typical class is to raise student achievement by about a quarter of a standard deviation unit. This is a small effect in educational terms, according to Cohen (1977), but it does mean that in a programmed class 60% of pupils attain at least the average score of the conventional class, compared with 50% of conventional pupils. Medium to large effects were observed in a third of the studies.

When examination scores were analyzed the average score of the conventional class was 64.8%; the average of the programmed class was 67.1%. The two groups differed by 2.3% points on average, based on scores taken from 56 studies.

The same technique was applied by Kulik, Schwalb and Kulik (1982), but with results from secondary school rather than higher education. A total of 47 studies were analysed, with 23 of the studies favouring programmed instruction and 24 favouring conventional classes. However, only 19 studies reported statistically different results, with a majority (12) favouring programmed instruction. The average value of the Effect Size was 0.08, the typical programmed instruction student gaining an advantage of less than one tenth of a standard deviation unit. Such an effect is trivial. It implies that 53% of programmed students perform at least at the average level of the conventional group, which has 50% of students at this level. The ES did vary from strongly negative to strongly positive and when study features were analyzed size of effect was found to be significantly related to the subject matter taught, with social sciences showing a particularly high effect size of 0.57. For mathematics it was surprisingly small and negative (-0.01) and it was small for the science studies. The conclusion was that:

in general programmed instruction did not improve the effectiveness of secondary school teaching. In the typical study, programmed instruction failed to raise student achievement on final examinations. It did not make students feel more positively about the subject matter they were studying or about the quality of teaching at their schools. Nor did it reduce the role that aptitude plays in determining how much students

learn in secondary school classes. (Kulik, Schwalb and Kulik, 1982, p.137)

Finally, the results of studies investigating computer-based learning also contribute to the overall picture of small or trivial effects. Kulik, Bangert and Williams (1983) examined the use of computers in several educational modes: managing, tutoring, simulation, programming and 'drill and practice'. The results for 'drill and practice' are relevant given Skinner's early claim for automation of this aspect of teaching. However, they do not show a great advantage for the computer in this role, with a small average ES of 0.27, when computer and conventional classes were compared.

The results of thirty years of experimental evaluations demonstrate that the expectations set by Skinner for his technological revolution have not been fulfilled. Programmed instruction was supposed to make learning more efficient and enjoyable. The research, however, shows that student reactions to programmed instruction were not discernibly different from reactions of students taught conventionally and that differences in attainment were trivial or small, in educational terms, in most cases. Clearly the job of improving education was more difficult than Skinner had predicted from his animal learning theory.

It may well be that learning theories are irrelevant to the solution of such problems, as Gagné (1962b) and McKeachie (1974b) have suggested. After all, the history of automated instruction goes back at least as far as 1860, before Thorndike's work was published, when Halcyon Skinner developed and patented a device for teaching spelling; and Pressey began building machines, which performed all the essential tasks carried out by later programmed devices, as early as 1915, without a supporting learning theory. Perhaps Leith (1968) is correct in suggesting that 'we have been bamboozled in our interpretation of programmed learning by the colourful exploits of a few flamboyant cheer-leaders.'(p.1)

REFERENCES

Bugelski, B.R. (1971) The Psychology of Learning Applied to Teaching. Indianapolis: Bobbs-Merrill.

Carpenter, C.R. and Greenhill, L.P. (1963) Comparative Research on Methods and Media for Presenting Programmed Courses in Mathematics and English. University Park, Pa.: Division of Instructional Services, Pennsylvania State University.

Cohen, J. (1977) Statistical Power Analysis for the Behavioural Sciences. (Revised edition) New York: Academic Press.

Coulson, J.E. and Silberman, H.F. (1960) 'Effects of three variables in a teaching machine', Journal of Educational Psychology, 51, 135-143.

Crowder, N.A. (1960) 'Automatic tutoring by intrinsic programming'. In A.A. Lumsdaine and R. Glaser (Eds) Teaching Machines and Programmed Learning. Washington, D.C.: Department of Audio-Visual Instruction, National Education Association.

Cummings, A. and Goldstein, L.S. (1962/1964) 'The Effect of Overt and Covert Responding on Two Kinds of Learning Tasks.' In W. Schramm (Ed) The Research on Programmed Learning: an annotated bibliography. (OE-31034) Washington, D.C.: U.S. Department of Health, Education and Welfare.

Elley, W.B. (1966) 'The role of errors in learning with feedback', British Journal of Psychology, 36, 296-300.

Evans, J.L., Glaser, R., and Homme, L.E. (1960) 'An investigation of variation in the properties of verbal learning sequences of the teaching machine type.' In A.A. Lumsdaine and R. Glaser (Eds) Teaching Machines and Programmed Learning. Washington, D.C.: Department of Audio-Visual Instruction, National Education Association.

Gagné, R.M. (1962a) 'The acquisition of knowledge', Psychological Review, 69(4), 355-365.

Gagné, R.M. (1962b) 'Military training and principles of learning', American Psychologist, 17, 83-91.

Gagné, R.M. and Paradise, N.E. (1961) 'Abilities and learning sets in knowledge acquisition', Psychological Monographs, 75(15), No.518.

Gagné, R.M., Mayor, J.R., Garstens, H.L. and Paradise, N.E. (1962) 'Factors in acquiring knowledge of a mathematical task', Psychological Monographs, 76(7), No.526.

Glaser, R. and Taber, J.I. (1961/64) Investigations of the Characteristics of Programmed Learning Sequences. In W. Schramm (Ed) The Research on Programmed Learning: an annotated bibliography. (OE-31034) Washington, D.C.: U.S. Department of Health, Education and Welfare.

Glass, G.V. (1976) 'Primary, secondary, and meta-analysis of research', Educational Researcher, 5, 3-8.

Grundin, H.U. (1969) 'Response mode and information about correct answers in programmed instruction: a discussion of experimental evidence and educational decisions'. In A.P. Mann and C.K. Brunstrom (Eds) Aspects of Educational Technology, III. London: Pitman.

Hartley, J. (1974) 'Programmed instruction 1954-1974: a review', Programmed Learning and Educational Technology, 6(11), 278-291.

Hilgard, E.R. and Bower, G.H. (1975) Theories of Learning. Englewood-Cliffs, N.J.: Prentice-Hall.

Holland, J.G. and Skinner, B.F. (1961) *The Analysis of Behaviour*.   New York: McGraw-Hill

Kulik, C.C., Schwalb, B.J. and Kulik, J.A. (1982) 'Programmed instruction in secondary education: a meta-analysis of evaluation findings', *Journal of Educational Research*, 75(3), 133-138.

Kulik, J.A., Bangert, R.L. and Williams, G.W. (1983) 'Effects of computer-based teaching on secondary school students', *Journal of Educational Psychology*, 75(1), 19-26.

Kulik, J.A., Cohen, P.A. and Ebeling, B.J. (1980) 'Effectiveness of programmed instruction in higher education: a meta-analysis of findings', *Educational Evaluation and Policy Analysis*, 6(2), 51-64.

Leith, G.O.M. (1968) *Second Thoughts on Programmed Learning*. Occasional paper No.1. London: National Council For Educational Technology.

Maccoby, N. and Sheffield, F.D. (1958) 'Theory and experimental research on the teaching of complex sequential procedures by alternative demonstration and practice'. In G. Finch and F. Cameron (Eds) *Symposium on Air Force Human Engineering, Personnel and Training Research*. Washington, D.C.: National Academy of Sciences, National Research Council.

McKeachie, W.J. (1974a) 'Instructional psychology', *Annual Review of Psychology*, 25, 161-193.

McKeachie, W.J. (1974b) 'The decline and fall of the laws of learning', *Educational Researcher*, 3, 7-11.

Mager, R.F. (1961a) 'On the sequencing of instructional content', *Psychological Reports*, 405-413.

Mager, R.F. (1961b) 'On the sequencing of instructional content'. In I.K. Davies and J. Hartley (Eds) *Contributions to an Educational Technology*. London: Butterworths.

Margolius, G.J. and Sheffield, F.D. (1961) 'Optimum methods of combining practice with filmed demonstration in teaching complex response sequences: serial learning of a mechanical assembly task'. In A.A. Lumsdaine (Ed) *Student Response in Programmed Instruction. A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media*. Washington, D.C.: National Academy of Sciences, National Research Council.

Means, R.S. and Means, G.H. (1971) 'Achievement as a function of the presence of prior information concerning aptitude', *Journal of Educational Psychology*, 62, 185-187.

Meyer, S.R. (1961) 'Report on the initial test of a junior high school vocabulary program'. In A.A. Lumsdaine (Ed) *Student Response in Programmed Instruction. A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media*.

Washington, D.C.: National Academy of Sciences, National Research Council.

Pressey, S.L. (1963/64) 'A Puncture of the Huge "Programming" Boom?' In W. Schramm (Ed) *The Research on Programmed Learning: an annotated bibliography*. (OE-31034) Washington, D.C.: U.S. Department of Health, Education and Welfare.

Schramm, W. (1964) *The Research on Programmed Learning: an annotated bibliography*. (OE-31034) Washington, D.C.: U.S. Department of Health, Education and Welfare.

Skinner, B.F. (1938) *The Behaviour of Organisms*. N.Y.: Appleton-Century-Croft.

Skinner, B.F. (1948) *Walden Two*. N.Y.: Macmillan.

Skinner, B.F. (1953) *Science and Human Behaviour*. N.Y.: Macmillan.

Skinner, B.F. (1954) 'The science of learning and the art of teaching', *Harvard Educational review*, 24, 86-97.

Skinner, B.F. (1968) *The Technology of Teaching*. N.Y.: Appleton-Century-Croft.

Smith, W. and Moore, J.W. (1961/1964) 'Size-of-step and Achievement in Programmed Spelling.' Lewisburg, Pa.: Bucknell University. In W. Schramm (Ed) *The Research on Programmed Learning: an annotated bibliography*. (OE-31034) Washington, D.C.: U.S. Department of Health, Education and Welfare.

Tait, K., Hartley, J.R. and Anderson, R.C. (1973) 'Feedback procedures in computer assisted arithmetic instruction', *British Journal of Educational Psychology*, 43, 161-171.

Thorndike, E.L. (1898) 'Animal Intelligence', *Psychological Review, Monograph Supplement*, 2, No.8.

Weiss, W., Maccoby, N. and Sheffield, F.D. (1961) 'Combining practice with demonstration in teaching complex sequences: serial learning of a geometrical construction task'. In A.A. Lumsdaine (Ed) *Student Response in Programmed Instruction. A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media*. Washington, D.C.: National Academy of Sciences, National Research Council.

CHAPTER 3

BEHAVIOURAL OBJECTIVES AND SYSTEMATIC INSTRUCTION

Perhaps Skinner was a flamboyant cheer-leader, as Leith suggested. His work, however, did have profound effects on educational thought in a number of ways related to the general concept of programmed learning. Attention was focussed first on the objectives of instruction in a general sense (Bloom et al., 1956), and then on the specific objectives of the behavioural objectives movement (Mager, 1962). The idea of mastery learning, popular in the 1930's but failing because of lack of a technology to sustain it (Block, 1971, p.4), was also resurrected as a corollary of programmed instruction, and with it evolved an interest in learning hierarchies and the structure of knowledge (Gagné and Paradise, 1961). These, in turn, have been influential in the development of systems of learning and instruction (Keller, 1968; Bloom, 1968) which have achieved the improvements in standards of learning which Skinner had hoped to bring about by his technological revolution. Here the emphasis has turned from technology in education, with its emphasis on hardware, to the technology of education.

BLOOM'S TAXONOMY OF OBJECTIVES

The 'Taxonomy of Educational Objectives: Handbook 1: The Cognitive Domain' (1956) arrived at a time when the demand for improvements in the efficiency of education was being made by Skinner and his colleagues. The idea for such a taxonomy came during a meeting of examiners attending the 1948 American Psychological Association Convention. Bloom and thirty-three colleagues met over a five year period to discuss their taxonomy, which was then organised and written by a select committee of five members. The taxonomy was to provide a 'theoretical framework which could be used to facilitate communication among examiners.'

It is intended to provide for classification of the goals of our educational system. It is expected to be of general help to all teachers, administrators, professional specialists, and research workers who deal with curricular and evaluation problems. It is especially intended to help them discuss these problems with greater precision. (Bloom et al., 1956, p.1)

It was hoped that such ambiguous terms as 'really understand', 'internalize knowledge' and 'grasp the core or

Figure 3-1. Major Classes of Taxonomies of Educational Objectives. (Based on Bloom et al., 1956; Krathwohl et al., 1964; Harrow, 1972)

Bloom et al.: Cognitive Domain

    1.00 KNOWLEDGE
    2.00 COMPREHENSION
    3.00 APPLICATION
    4.00 ANALYSIS
    5.00 SYNTHESIS
    6.00 EVALUATION

Krathwohl et al.: Affective Domain

    1.00 RECEIVING (attending)
    2.00 RESPONDING
    3.00 VALUING
    4.00 ORGANISATION
    5.00 CHARACTERIZATION BY A VALUE OR VALUE COMPLEX

Harrow: Psychomotor Domain

    1.00 REFLEX MOVEMENTS
    2.00 BASIC-FUNDAMENTAL MOVEMENTS
    3.00 PERCEPTUAL ABILITIES
    4.00 PHYSICAL ABILITIES
    5.00 SKILLED MOVEMENTS
    6.00 NON-DISCURSIVE COMMUNICATION

---

essence', would be redefined as a set of standard classifications, making exchange of information about curricular development and evaluation more precise. Equally important, the psychological relationships within the taxonomy were seen as forming a basis for psychological investigations to shed light on changes in the learner's behaviour. This aspect of the taxonomy has been pursued in various forms throughout Bloom's extensive involvement in the study of education.

The educational objectives described in the taxonomy were in a behavioural form and were seen as being consistent with relevant and accepted psychological principles and theories (p.6). They were concerned with the changes produced in individuals as a result of educational experiences. In other words, the taxonomy was a classification of behaviours which represented the intended

outcomes of the educational process (p. 12). However, it was not intended to evaluate the actual behaviours acquired by the students. The emphasis was on obtaining evidence on the extent to which desired and intended behaviours were learned by students. Determining the appropriate value to be placed on the different degrees of achievement of the objectives of instruction was considered to be a matter of grading or evaluating the performance and as such it was considered to be outside the scope of the taxonomy.

One major principle of the taxonomy, that it should be consistent with the then current understanding of psychological phenomena, was reflected in the behaviourist overtones that it should be applied only to those educational programmes which could be specified in terms of changes in 'intended student behaviours.' It was assumed that these changes in behaviour were changes from the simple to the complex.

> If we view statements of educational objectives as intended behaviours which the student shall display at the end of some period of education, we can then view the process as one of change. As teachers we intend the learning experiences to change the student's behaviour from a simpler type to another more complex one which in some ways at least will include the first type.
> (Bloom et al., 1956, p.16)

This view led to the taxonomy being organised into six major classes (Figure 3-1), with a hierarchical order, the objectives in one class making use of and being built on the behaviours in the preceding classes (p.18). According to Bloom, research showed that there was 'an unmistakable trend pointing toward a hierarchy of classes of behaviour' which was in accordance with the 'Taxonomy Of Educational Objectives: Cognitive Domain.' Similar arrangements were to be found in the second part of the taxonomy dealing with the affective domain (Krathwohl et al., 1964) and the tradition was maintained in the taxonomies of the psychomotor domain developed separately by Simpson (1966) and Harrow (1972).

The cognitive domain was further divided into subclasses and subcategories, giving a more comprehensive picture of each level of the hierarchy. For example, the class 'Knowledge' was divided into two subclasses, each with several subcategories (Figure 3-2).

Davies (1976) maintains that there is also a fundamental division in the cognitive domain, which has been largely overlooked in subsequent literature. It was divided into two broad areas, with the lower area of knowledge dealing with products, where little more than recall and recognition is required. The higher areas deal essentially with intellectual processes and require more complex intellectual

Figure 3-2. Taxonomic Structure of 'Knowledge'. (Bloom at al., 1956)

1.00	KNOWLEDGE
1.10	Knowledge of specifics
1.11	Knowledge of terminology
1.12	Knowledge of specific facts
1.20	Knowledge of ways and means of dealing with specifics
1.21	Knowledge of conventions
1.22	Knowledge of trends, sequences
1.23	Knowledge of classifications and categories
1.24	Knowledge of criteria
1.30	Knowledge of the universals and abstractions in a field
1.31	Knowledge of principles, generalizations
1.32	Knowledge of theories and structures

abilities, such as determining relationships between and within the knowledge associations acquired in the lower levels of the domain.

Validity of Bloom's Taxonomy

Kropp and Stoker (1966) tested the validity of the hierarchical arrangement for the cognitive domain. The theoretically expected relative order for the categories Knowledge, Comprehension, Application and Analysis was obtained, but not for Synthesis and Evaluation, which both fell between Knowledge and Comprehension. Madaus, Woods and Nuttall (1973) suggested that the taxonomy had a Y-shaped structure, with the stem formed from Knowledge, Comprehension and Application and subsequently divided into one branch of Analysis and another branch from Synthesis to Evaluation.

Ormell (1974) has also found contradictions in the taxonomy. He believes that certain demands for the Knowledge category are more complex than the demands for Analysis or Evaluation. He has suggested that the hierarchical organisation based on increasing complexity of operations should be abandoned and the taxonomy should be split into

six parallel categories, but maintaining the different levels within each category.

The linear assumption has been heavily criticized by a number of philosophers of education. For example, Pring (1971) argues that even 'knowledge of specifics' has a variety of intellectual abilities and skills implicitly associated with it:

.... for something to be recognized as a fact requires some comprehension of the concepts employed and thus of the conceptual framework within which the concepts operate. Similarly with regard to the knowledge of terminology, it does not make sense to talk of the knowledge of terms or of symbols in isolation from the working knowledge of these terms or symbols, that is, from a comprehension of them and thus an ability to apply them. (Pring, 1971, p.90)

Seddon's (1978) appraisal of the cognitive domain dealt critically with the educational issue of communicability and the psychological issue of the cumulative structure of the hierarchy, but also recognized the success of the taxonomy in terms of the considerable impact it has made on educational thought and practice and concluded that it had achieved one of Bloom's criteria for its success:

The taxonomy must be accepted and used by workers in the field if it is to be regarded as a useful and effective tool. (Bloom et al., 1956, p.24)

Green (1964) was critical of the assumption that the goal of teaching is to change students' behaviour. He argued that the goal was to transform behaviour into action, in other words the goal is the capability, competence or understanding that makes rational action possible, an idea since exploited by Gagné and Briggs (1974) in 'Principles of Instructional Design.'

The fact that the taxonomy chose to classify objectives on the basis of intended behaviours without regard for the particular content has also received wide criticism, as Sockett (1971) makes clear:

"remembering" is unintelligible just as a psychological process (even if we lay aside its counterpart, forgetting) for we remember something, cases of remembering are cases of being right about what was or is the case. We cannot posit remembering in any sense apart from content. If remembering is thought of as content-free we have an empty concept which could not be even part of an educational objective. (p.20)

Furst views as rather more serious the omissions stemming from the separation of the cognitive, affective and psychomotor domains. This artificial separation prevents some desired outcomes from being encompassed within the cognitive scheme, and he gives examples of those excluded: receptivity and sensitivity, skill in observing and data gathering, 'perceptual' or 'motor' activities, moral concepts, basic democratic values and rationality. But perhaps the most serious omission, for Furst, is the lack of an 'understanding' category. Ormell (1979) also objects to this and does not accept the argument that it was omitted as a major category because it was imprecise and has proposed both a definition and a research strategy for identifying this cognitive element.

Travers (1980) does not see it as a true taxonomy at all, and claims that although the major categories give the appearance of a continuum of complexity, it is more akin to an 'inventory of educational customs.' He feels that it lacks the theoretical under-pinnings of true taxonomies.

However, despite such criticisms, even Furst (1981) acknowledges that:

A handbook that has had over a million copies sold, been translated into several languages, used worldwide, and cited thousands of times hardly needs extensive documentation on its usefulness. (p.448)

OBJECTIVES: A HISTORICAL PERSPECTIVE

The idea for educational objectives was not a new one, and Bloom et al. acknowledged their debt to many authors, particularly Ralph Tyler, to whom the first volume was dedicated. Davies (1976) traces the influences back to Cicero's six divisions of a speech, which were used by Herbart (1924) in his five stages of preparation for lessons. Earlier, Herbert Spencer (1910) had called for a curriculum based on the 'things it most concerns us to know', given that we do not have sufficient time to master all subjects.

Franklin Bobbitt (1924), professor of education at the University of Chicago, took the idea of objectives a stage further, insisting that, in an age of science, exactness and particularity were demanded. He even went so far as to report a list of objectives for each of the major occupations in Indianapolis, which would be needed by children when they left school. These included, for example, the ability to sharpen, adjust, clean, lubricate, replace worn or broken parts, and otherwise keep household and garden tools and appliances in good order and good working condition.

Werrett Charters (1924) extended this scientific perspective, using techniques of job analysis, to argue that individual tasks should be broken down to enable the student to learn without assistance. He suggested seven steps in the construction of a curriculum. First, the major objectives were to be determined; then, these were to be analysed to the level of working units; the working units were to be arranged in order of importance, first for adults and then for children; the number of items capable of being learned in school were selected; the best methods for teaching these items were to be determined; and, finally, the items were to be assembled in an order according to the nature of the children.

Tyler's Principles of Objectives

Bobbitt and Charters were recommending a behavioural approach to curriculum design, founded on the specification of educational objectives. These ideas were later applied, by Ralph Tyler, to the realms of test construction when he was faced with the unscientific diagnostic tests used in the biological sciences at Ohio University. He was in favour of having teachers:

> formulate the course objectives, define the objectives in terms of student behaviour, collect situations in which students are to indicate the presence or absence of each objective, and provide the method of evaluating the student's reactions in the light of each objective. (Tyler, 1932)

Tyler was, in effect, proposing an early description of what was later to be called 'criterion reference' testing (see Chapter 4). He went on, in his classic text 'Basic Principles of Curriculum and Instruction' (1949), to detail a series of techniques by which 'materials are selected, content is outlined, instructional procedures are developed and tests and examinations are prepared.' And, as a first step in the systematic and intelligent study of educational programmes, he suggests that 'we must be sure as to the educational objectives aimed at' (1949, p.3). The objectives are not to be confused with things which the instructor is to do, they must be expressed in terms of changes in student behaviour:

> Since the real purpose of education is not to have the instructor perform certain activities but to bring about significant changes in the students' patterns of behaviour, it becomes important to recognize that any

statement of the objectives of the school should be a statement of changes to take place in students. (Tyler, 1949, p.44)

He claims that the most useful form for stating objectives is to include both the kind of behaviour to be developed in the student and 'the content or area of life in which this behaviour is to operate' (p.46). The objectives 'to write clear and well-organized reports of social studies projects', 'familiarity with dependable sources of information on questions relating to nutrition' and 'to develop an appreciation of the modern novel', are given by Tyler as models of clear objectives including both behavioural and content aspects, although he does acknowledge that 'to develop an appreciation' may require some further definition.

The 'behavioural objectives' must be defined clearly because they will form the basis for the evaluation procedures, which are an integral part of Tyler's scheme. Choosing and formulating educational objectives and the learning experiences do not complete the planning cycle for Tyler. Evaluation does this because it 'becomes a process for finding out how far the learning experiences as developed and organized are actually producing the desired results' (p.105). There are two aspects to evaluation. The first implies that evaluation must look at the behaviour of the student, and the changes in behaviour as a result of the process of education. The second implies that evaluation must involve more than a single appraisal. At least, it must consist of both an assessment at the beginning of the course of instruction and an appraisal at the end of the programme in order to determine the changes which have taken place. This is not enough for Tyler who expresses a desire for more testing during the educational programme in order to determine the changes in behaviour step by step. His suggestion finds fertile ground in later years in the form of summative and formative testing.

Mager's Behavioural Objectives

While Tyler advocated the use of objectives in the context of evaluation procedures Bloom and his colleagues concentrated on the precise specification of educational objectives without such an emphasis. It was taken up again in 1962, with the accelerating interest being shown in the programmed learning movement, when Robert Mager published a short book aimed at helping instructors to answer the question 'What is worth teaching?' by demonstrating the steps involved in 'Preparing Instructional Objectives.' Mager claimed that, having read the book:

Given any objective in a subject area with which you are familiar, in all instances be able to identify (label) correctly the performance, the conditions, and the criterion of acceptable performance, when any or all those characteristics are present. (Mager, 1962/1975, p.3)

This statement presents an immediate outline of the three essential components in a Mager objective: first, the behaviour or performance to be accepted as evidence that the learner has changed; second, the conditions under which the behaviour will be demonstrated; and third, the standards or criteria of acceptable performance against which the success or failure of the learner will be judged. Davies (1976) is correct in assuming that the first component in Mager's scheme includes both of Tyler's dimensions for an objective, the kind of behaviour and the content area. The third incorporates Tyler's demand for evaluation as part of the complete cycle. Thus, Mager is taking on board that aspect explicitly excluded in Bloom's Taxonomy, and as such can be seen as a major advance in the application of behavioural objectives.

In one of the asides in his book (it was a scrambled programmed text, with 'optional material' scattered throughout, which the reader could 'save for a rainy day') he vividly describes what can happen when objectives are not stated clearly. It concerns students who were learning how to operate and repair a large, complex electronic system. The course goal was 'to be able to operate and maintain' the system. Time on the actual machine was limited and in order to increase the practice time students were taught in the classroom. The instructor would point out a component on a diagram and ask what effect it would have on the system if it failed. The students traced back through their individual diagrams and listed the symptoms. Of course, this is the exact opposite to what was expected of the operator, who would observe a symptom and then deduce the possible causes. As Mager said, the students were learning to run forwards by being taught to run backwards!

The correct approach was demonstrated in his analysis of the separate skills and objectives required by the general statement that students were 'able to read electrical meters.'

1. Given a meter scale, be able to identify (state) the value indicated by the position of the pointer as accurately as the construction of the meter will allow.
2. Be able to identify (state) the value indicated by the pointer on meter scales that are linear, non-linear, reversed, or bi-directional.

3. Given a meter with a single scale and a range switch, be able to identify (state) the value indicated by the pointer for each of the ranges shown by the range switch.
4. Given a meter with several scales and a range switch, be able to identify (state) the scale corresponding to each setting of the range switch.
(Mager, 1962/75, p.53)

There are improvements that can still be made to these objectives, for example, by adding to the criterion for reading the value in objective 1. The objectives can also be made too specific and in doing so may lose their value. Adding lighting and room temperature conditions to the objective 'will type a specimen letter, in accordance with office procedures, on an IBM PC word-processor and obtain one paper copy, with no mistakes, in less than five minutes', will not add a great deal to the usefulness of the objective. Adding details of the type of software running on the machine or the type of printer or even paper (single sheet, continuous roll) will improve its value.

In answering the performance question 'What should the learner be able to do?' Mager suggests that action verbs are going to be more useful than 'fuzzies' or words open to a wide range of interpretation. 'Fuzzies' include such words as: to know, to understand, to really understand, to fully appreciate, to grasp the significance of. The action words open to fewer interpretations include: to write, to recite, to construct, to solve.

The conditions under which the performance will be demonstrated should clarify what the student brings to the situation and what is provided. Mager suggests a number of examples: given any reference of the learner's choice... ; given a standard set of tools... ; without the aid of a slide rule.... Thus, the statement 'be able to solve problems in algebra' is changed to something along the lines 'given a linear algebraic equation be able to solve (write the solution) for the unknown without the aid of references, tables, or calculating devices.'

Finally, the criterion must be made explicit because:

if we can specify the acceptable performance for each objective, we will have a standard against which to test our instruction; we will have the means for determining whether our instruction is successful in achieving our instructional intent. If, for example, our best experience and wisdom tell us that we must not consider a student competent until that student can perform within a strict time limit, then we know that we will

have to instruct and assist that student until the desired performance level is reached. (Mager, 1962/1975, p.71)

For Mager the criterion is essentially a measure of the correctness of instruction for a particular student. When the student fails to reach the required level, the instruction is failing as well as the student. The criterion should not be a minimum or barely tolerable criterion, it should describe the desirable performance, although this will depend on the specific circumstances, as Mager says, it may be acceptable for a shipping clerk to occasionally tie a knot that slips, but that level of performance would not be acceptable for a surgeon. The criterion can be expressed as speed, accuracy or quality. A speed criterion is usually given as a time limit within which the performance must occur and accuracy usually indicates the limits eg. be able to tell the time shown on a clock to within one minute. Quality may be reduced to descriptions of accuracy, or it may include a quantitative specification eg. the number of examples to be given in a definition.

Mager's contribution to the development of specific behavioural objectives has been perfectly summarized by Davies:

> Rarely during the history of education has such a small book (62 pages), with so much blank paper, been as influential on both the thinking and practice, loves and fears of both British and American teachers. (Davies, 1976, p.57)

Objectives and the Principles of Instructional Design

Objectives have continued to occupy a central role in systems of instructional design and the components of operational definitions of objectives have been further extended by Gagné and Briggs (1979) from three to five, although they are not widely different from those of Mager and do not 'differ from them in any crucial respect.' Rather, there are differences in emphasis and distinctions which previously had not been highlighted, in particular, the distinction between verbs for 'action' and verbs used to identify the 'learned capability' implied by the behaviour under observation. There are five components in the scheme: action, object, situation, tools and constraints, and capability to be learned. The example of the poor objective 'types a letter' is used by Gagné and Briggs to illustrate the five components.

1. ACTION. The objective is fine with respect to the action component.
2. OBJECT. The object of the performance is clear ie. a letter is to be produced.
3. SITUATION. Clearly this requires some improvement: is she given a letter written in long hand to type, or is it on a tape, or even in note from?
4. TOOLS AND OTHER CONSTRAINTS. This aspect needs to be developed. What sort of typewriter is to be used, is a carbon required, what is the length of the letter? Obviously, different limits will be set for students at different levels.
5. CAPABILITY TO BE LEARNED. 'Types a letter' does not say much about the different human capabilities that may be involved in the process of producing the letter. It could mean that the letter is to be copied, or it could be that the letter is to be composed. These will require very different capabilities.

The poor objective is up-graded along the following lines: given a letter inquiring about the shipping of an order (situation) the student will generate (the learned capability, implying a problem-solving approach) a letter in reply (object) by typing (action), using an electric typewriter, making one carbon of a one-page letter (tools and other constraints).

Hierarchies, Capabilities and Systems

The new component 'capability to be learned' is related to Gagné's hierarchy of learning categories (Figure 3-3) which originally arranged learning into eight different types ranging from the simple to the complex. The most important class of conditions that distinguished one type from another in the hierarchy was the initial state of each. Gagné (1965) called this its prerequisites. At the higher end of the hierarchy each type has the previous types as prerequisites.

The arrangement of these different types of learning into a hierarchical structure reflects Gagné's behaviouristic S-R origins, which became less relevant with the rise of more sophisticated models of human cognition, particularly the information processing model which he later adopted. Gagné acknowledges this change in emphasis in his preface to the third edition of 'The Conditions of Learning':

> The considerable shift in orientation from its previous edition requires a very different organisation for this

Figure 3-3. Gagné's Hierarchy of Learning Types, based on Conditions of Learning (1965, First Edition).

Type 8. Problem solving, requires the internal events usually called thinking, which requires as prerequisites:

Type 7. Principle or rule learning, is the forming a chain of two or more concepts, which requires as prerequisites:

Type 6. Concept learning, requires a response that identifies an entire class of objects or events, which has as prerequisites:

Type 5. Multiple discriminations, requires different responses to similar stimuli, which has as prerequisites Type 4 or Type 3.

Type 4. Verbal association or the learning of chains that are verbal, which has as prerequisite Type 2.

Type 3. Chaining of 2 or more S-R connections, which has as prerequisite Type 2.

Type 2. Stimulus-response learning, as found in the work of Thorndike and Skinner, which Gagné felt did not have Type 1 as a prerequisite.

Type 1. Signal learning, the classical conditioning of Pavlov, the most basic type of learning.

one, and has occasioned its being largely rewritten to include much new material. This book, I feel, deserves a preface which is both frank and more than usually detailed. (Gagné, 1976, p.iv)

The hierarchy of learning types was removed and its place was taken by the 'learned dispositions or capabilities' which form the varieties of learning outcomes. There are now five of these (Gagné, 1985): intellectual skills, cognitive strategies, verbal information, motor skills and attitudes and, unlike the learning types, they do not have 'simply ordered relationships with one another.' Their order of presentation is mainly for convenience. However, the higher levels of the old hierarchy are incorporated into the fabric of the new scheme, while the

Figure 3-4. Verbs Describing Human Capabilities. (Based on Gagné and Briggs, 1979; Gagné, 1985)

CAPABILITY	ASSOCIATED VERB
Intellectual Skill	
Discrimination	DISCRIMINATES
Concrete Concept	IDENTIFIES
Defined Concept	CLASSIFIES
Rule	DEMONSTRATES
Higher-order Rule	
(Problem Solving)	GENERATES
Cognitive Strategy	ORIGINATES
Verbal Information	STATES
Motor Skill	EXECUTES
Attitude	CHOOSES

lower half of the hierarchy is ultimately relegated to a single category. Type 8, problem solving, becomes associated with both intellectual skills and cognitive strategies capabilities, while Type 7 (principle or rule learning), Type 6 (concept learning) and Type 5 (multiple discrimination) are incorporated in the intellectual skills capability. Each capability has 'standard verbs' associated with it to describe what will be present when the capability is learned (Figure 3-4).

Intellectual skills involve 'knowing how', or procedural knowledge. They include making discriminations, as when a child learns to make the correct sound associated with a given letter, rather than another sound. Identification, the capability verb for concrete concepts, is in evidence when an individual can respond to a collection of things by distinguishing among them and is able to name the individual items. Putting things into a class and responding to any instance of the class as a member of that class is evidence of a defined concept, and the ability to respond to any instance of a class of situations with an appropriate instance of a class of performances is evidence of a rule being used. Gagné (1985) feels that learning of rules is of vast importance because rules make up the bulk of what is learned in schools. One aspect of problem solving is found when the learner discovers a combination of previously learned rules and combines them to generate a solution in a novel situation.

A further aspect of problem solving is associated with the cognitive strategies capability, which is represented by the ways that individuals focus their knowledge and skills

on problem situations that may not have been encountered before. Gagné calls these strategies ways of 'using one's head.' The 'problem finding' aspect of cognitive strategies, involving both the identification of novel problems and the translation of known rules into forms which provide the solutions, is of major importance.

Verbal information is also known as declarative knowledge and refers to knowledge that is 'verbalizable.' It can take the form of names or labels, single propositions or facts or collections of meaningfully organised propositions.

Motor skills require the execution of muscular movements, although a more comprehensive term is 'psychomotor' skills because of the necessary co-ordination of bodily systems. The extent of this category can be seen from the examples Gagné (1985) gives, including executing a dive, using a typewriter and adjusting the internal mechanism of a watch.

Learning results in the creation of internal states that influence choices of an individual's action. These outcomes are called attitudes and are made up of three components: a cognitive component, an affective component and a behavioural component. According to Gagné this capability is demonstrated when a person chooses to behave in one manner rather than another.

Gagné and Briggs have attempted to build a framework which takes account of the three areas of educational objectives (cognitive, affective and psychomotor) and integrates them with the various types of learning to form the human capabilities component, thus producing a more coherent scheme.

The addition of this may be seen as an improvement, but there is one glaring omission; the criterion component does not appear in the scheme. There is no mention of the number of mistakes permitted nor the time limit for completion of the job. Gagné and Briggs take issue with Mager on this point:

> Instructional objectives describe the class of performances that may be used to determine whether the implied human capability has been learned. However, they do not state in quantitative terms what criteria will be used to judge whether any particular performance class has been learned.
> (Gagné and Briggs, 1979, p.127)

Two reasons are given for this. First, the criteria should be linked to the different types of human capabilities. Second, criteria of performance are related to the techniques of performance assessment, and it could be confusing to become concerned with assessment procedures when objectives are being described. Statements of

objectives always imply 'mastery', but the criterion of mastery needs to be separately decided when considering assessment methods. If the proper conditions are provided the frequently used course design criterion of 90 percent of the students achieving 90 percent of objectives is recommended by Briggs and Gagné. This can have the effect of monitoring both student performance and the adequacy of the instruction.

The objectives movement has given rise to what is now known as the 'systems approach' to instructional design, in which educational technology is related to the process of planning, although the term 'systematic approach' is preferred by some authors because of its cyclical nature (MacDonald-Ross, 1973). Rowntree (1982) sees this as consisting of four stages:

1. Identifying purposes, which includes analysing aims, describing the students, determining the objectives and evaluation procedures.
2. Design of learning, during which objectives and subject matter are analysed, learning sequences and strategies are devised, and media, materials and experiences are selected.
3. Evaluation, when the materials are tried out and monitored.
4. Improvement, during which phase the results of the evaluation stage are reviewed and revisions are made of the products of the previous activities.

Gagné and Briggs (1979) have a similar scheme, but the analysis of objectives is separated from the design of measures of learner performances.

This approach to the design of instruction, with its emphasis on the early specification of objectives, is an essential part of the production of programmed instruction and other behaviourally oriented methods of teaching, including Bloom's 'Learning for Mastery' and Keller's 'Personalized System of Instruction.' The success of such systems must be attributable, in part, to this aspect of the design process.

SUPPORT, CRITICISM AND RESEARCH

Instructional objectives attracted widespread criticism and fervent support in equal measure throughout the early years. Popham (1969) reviewed ten reasons which he found to be used by educators to 'escape the practice of stating their objectives behaviourally.' He saw each of the reasons as carrying its own appeal to different sorts of educators,

each with its own degree of reasonableness or emotionality, but each being essentially invalid.

To the suggestion that 'trivial learner behaviours are the easiest to operationalize, hence the really important outcomes of education will be under-emphasised', he did acknowledge that too many examples of behavioural objectives dealt with the picayune, especially those following Mager's examples, most of which dealt with the lowest level of Bloom's Taxonomy. But the fact that it was possible to make such behaviours explicit should result in their identification as unworthy of educational effort and hence their elimination.

It was also suggested that pre-specification of explicit goals prevents the teacher from taking advantage of opportunities unexpectedly occurring in the classroom. Popham argued that serendipity is always welcome in the classroom, but 'it should always be justified in terms of its contribution to the learner's attainment of worthwhile objectives', otherwise it may fall into the category of ephemeral entertainment for pupils.

There are other educational outcomes which are as important as changes in pupil behaviour, such as changes in parental attitudes, community values, etc. and to this Popham responded that all such modifications should be justified in terms of their contribution toward the pupils' achievements and the desired changes in their behaviours.

It was also suggested that the emphasis on behaviour which can be objectively and mechanistically measured must result in a de-humanizing approach to education. This view was based on the erroneous views of measurement as exclusively examination-based, which ignore the 'extremely sophisticated ways of securing qualitative as well as quantitative indices of learner performance.'

Another reason for not specifying educational objectives was that it was somehow undemocratic to plan in advance exactly how the learner should behave after instruction. Popham retorted that teachers generally have an idea of how they wish to behave, and they promote these goals with more or less efficiency. In any case, 'if schools were allowing students to "democratically" deviate from societal mandated goals' they would cease to receive society's support.

The fact that teachers do not specify their goals in terms of measurable behaviours and since that is the way of the real world perhaps it is unreasonable to expect such a dramatic change, was given as another reason for avoiding educational objectives. Not so, replied Popham, they may not, but they should because 'The way teaching really is at the moment just isn't good enough.'

Certain subjects were seen to offer more scope for the specification of student behaviours than others. In the fine arts and humanities it would be particularly difficult.

'Tough but not impossible', was Popham's reply, and he gave an example that many students must have sympathized with:

> Any English teacher, for example, will tell you how difficult it is to make a valid judgement of a pupil's essay response. Yet criteria lurk whenever this teacher does make a judgement, and these criteria must be made explicit.
> (Popham, 1969, p.50)

Another reason for not adopting objectives was that if most educational goals were stated precisely, they would be revealed as generally innocuous. Popham acknowledged this potential threat to school people and declared that much of what goes on in the schools is indefensible. If what is happening in schools is trivial, educators should know it and so should those who support it. He called upon teachers to abandon the ploy of 'obfuscation by generality' and clarify what is being done.

'Measurability implies accountability' and as such represents a good reason for many teachers not using the objectives approach. Teachers might actually be judged on their ability to bring about the desired changes, and so they should be, according to Popham.

He sympathized with the final reason, that it is very difficult and time consuming for teachers to generate precise objectives. The answer to this was to reduce the teacher's load to enable her to become a professional decision-maker and not merely a custodian. Alternatively, funds could be diverted to enable objectives to be generated by external agencies, which would produce alternative objectives for all fields and at all grades.

MacDonald-Ross (1973) looked at both the advantages and disadvantages claimed for behavioural objectives. In their favour he concluded that they formed the basis of the only well-worked out method of rational planning in education, encouraging detailed planning and making explicit previously concealed values. They were seen as providing a basis for evaluation and the choice of instructional means, within a self-improving system which achieves internal consistency and the attainment of its initial aims. Objectives act as a medium of communication and can form the basis for individualising instruction.

However, there were certain disadvantages. The most serious were that there was no view as to the origins of objectives and there were no well-defined prescriptions for deriving them. Although it is claimed that behavioural objectives provide the only objective basis for evaluation, MacDonald-Ross felt that they cannot prescribe the validity of test items because they are inherently ambiguous and various items can usually be written for each objective. He

Figure 3-5. Major Objections to Behavioural Objectives,
according to MacDonald-Ross (1973)

Major Objections

1. No consistent view exists as to the origins of
   objectives.
2. In the educational domain no well-defined prescriptions
   are available for deriving objectives.
3. Weak prescriptions lead to cycling. This can be costly.
4. Lists of behaviours do not adequately represent the
   structure of knowledge.
5. The use of behavioural objectives implies a poverty-
   stricken model of student-teacher interaction.
6. The behavioural objectives scheme suffers from many
   weaknesses of any operational dogma.
7. Objectives do not prescribe the validity of test items.
8. The level of specificity problem has never been solved.

Minor Objections

9. Defining objectives before the event conflicts with
   voyages of exploration.
10. Advocates do not show how teachers can use objectives to
    guide unpredictable classroom events.
11. There are an extremely large number of paths through any
    body of knowledge, thus reducing the effectiveness of
    objectives in design.
12. In some disciplines criteria can only be applied after
    the event.
13. Objectives are inherently ambiguous.
14. Objectives do not communicate intent unambiguously,
    especially to students.
15. Trivial objectives are the easiest to operationalize,
    and this is a problem.
16. The relevance of goal-referenced models of education can
    be questioned.

---

also questioned the level of specificity of objectives: just
how specific is it necessary to be to define the required
behaviours in a course of instruction? He calculated that to
move to a level suggested by Mager, if applied to an average
educational psychology text, would require 10,000 objectives
in a volume of 150,000 words.

A further disadvantage expressed by MacDonald-Ross was
the cost involved in the cyclical nature of testing and
modifying instruction to attain preset objectives. He
supported the common criticism of objectives, that they
represent a poverty-sticken view of student-teacher

interaction and, at a more philosophical level, agreed with
Pask (1972) that lists of behaviours do not adequately
represent the structure of knowledge. His views were
summarized as a list of 16 objections to objectives,
including many that were to be found in Popham's discussion
(Figure 3-5).

The Research

It has been suggested that objectives contribute to
student successes because they provide a detailed
description of precisely what it is that the student must
learn during instruction. Duchastel and Merrill (1973)
reviewed the evidence accumulated to answer the question:
Does communicating behavioural objectives to students
facilitate learning? In studies which investigated the
straight comparison between students with and students
without objectives they concluded that five studies
demonstrated a significant effect due to the availability of
behavioural objectives, and five found no difference, for
tests of immediate retention. For delayed retention the
results were also inconclusive.

Melton (1978) points out that a substantial number of
researchers have recorded experiments that lend support to
the claim that behavioural objectives enhance relevant
learning when presented to students, whilst others show no
such enhancement. One aspect, explored by Engel (1968), was
that students must be aware of the objectives and must read
them for there to be a positive effect. The difficulty of
the objectives may also be a factor, according to Brown
(1970), for if the objectives are extremely difficult the
majority of students will fail to master them and it will be
difficult to discriminate between the performance of
students who do and do not receive the objectives. These and
other results led Melton to the conclusion that there were
five conditions under which objectives would be ineffective:

1. If students ignore the objectives provided, either
   because they are unaware of them, or because prior
   experience suggests that it is not important to take
   note of them.
2. If the objectives are too general, or too ambiguous,
   to be of particular assistance.
3. If the objectives are of extreme facility or
   difficulty. (The structure of readability of
   instructional material may be closely related to this
   condition.)
4. If the objectives of particular interest are only a
   small proportion of those provided to students.

> 5. *If students are conscientious, or so highly*
> *motivated, that they achieve the objectives*
> *regardless of whether or not they are specified.*
> (Melton, 1978, p.294)

Although most research had involved the effects of objectives on relevant learning, Melton pointed to the controversy concerning the effects of objectives on incidental learning. Morse and Tillman (1972) had found that behavioural objectives enhanced relevant learning without detracting from incidental learning against unspecified objectives. Whereas, Duchastel (1972, 1977) demonstrated that providing objectives had a detrimental effect on incidental learning. Confounding these observations, Rothkopf and Kaplan (1972, 1974) found the provision of behavioural objectives enhanced both relevant and incidental learning. Melton reasoned from the experimental details in these studies that the effect of the objectives will depend on when they are presented to the student and how they are used during study. When objectives are given prior to study and are read by the students before commencing their study the objectives would act as orienting stimuli, enhancing relevant learning but depressing incidental learning. Objectives read after the period of study would act as reinforcing stimuli, enhancing relevant learning but not at the expense of incidental learning.

When the interaction of objectives and the type of learning was investigated only one study out of seven found objectives to be more effective with one type of learning (knowledge) than with others, and this was not sustained on the retention test.

When learner characteristics were considered, there was some evidence of interactions between such characteristics and provision of objectives, with middle ability students benefitting more from objectives than either high or low ability students. Kueter (1970) found behavioural objectives to be less effective for personality traits of submissiveness, self-control, considerateness, conscientiousness, or low ergic tension. Objectives appear to reduce the level of anxiety for some students (Merrill and Towle, 1971).

Investigations of the hypothesis that students who receive objectives will take less time in learning were few but two demonstrated a great reduction in learning time if the learners had control over how and when they learned. Mager and Clark (1963) reported a reduction in training time of 65 percent, with the learners being better equipped than graduates of the traditional course. Learners were provided with 24 pages of detailed objectives, classes were cancelled and the students were told that they would have complete control over their learning and could ask for instruction

when they required it. In a further experiment a saving of 50 percent in learning time was observed. Smith (1970), in a more controlled experiment, did not support previous findings and Duchastel and Merrill caution that studies reported by Mager and Clark are heavily confounded by the student control variable.

Although the research has demonstrated some benefits for courses in which objectives are specified for the benefit of both students and teachers, it must be remembered that the specification of objectives is usually one part of what is now known as the 'systems approach' to course design. Evaluations of the benefits of objectives in isolation can only measure the effects of one of several components, whereas the basis of the systems approach is its synergism. As such, it makes sense to evaluate the total system, as well as the contributions of the individual parts. The programmed instruction 'system', which gave the initial impetus for behavioural objectives, was discussed in the previous chapter. In the following chapter the major mastery 'systems' will be evaluated.

REFERENCES

Block, J.H. (1971) *Mastery Learning*. London: Holt, Rinehart and Winston.

Bloom, B.S. et al. (1956) *Taxonomy of Educational Objectives Handbook I: Cognitive Domain*. N.Y.: McKay.

Bloom, B.S. (1968/81) 'Learning for Mastery', (UCLA-CSEIP) *The Evaluation Comment*, 1(2). In B.S. Bloom (Ed) *All Our Children Learning*. London: McGraw-Hill.

Bobbitt, F. (1924) *How To Make a Curriculum*. Boston: Houghton Mifflin.

Brown, J.L. (1970) 'The Effects of Revealing Instructional Objectives on the Learning of Political Concepts and Attitudes in Two Role-playing Games.' Unpublished Doctoral dissertation, University of California, Los Angeles.

Charters, W.W. (1924) *Curriculum Construction*, N.Y.: MacMillan.

Davies, I.K. (1976) *Objectives in Curriculum Design*. London: McGraw-Hill.

Duchastel, P.C. (1972) *Incidental and Relevant Learning with Instructional Objectives: Technical Memo 66*. Tallahassee: Florida State University, CAI Centre.

Duchastel, P.C. (1977) *Functions of Instructional Objectives: Organization and Direction*. Paper presented at the meeting of the American Educational Research Association, New York, April.

Duchastel, P.C. and Merrill, P.F. (1973) 'The effects of behavioural objectives on learning: a review of empirical studies', _Review of Educational Research_, 43(1), 53-69.

Engel, R.S. (1968) 'An Experimental Study of the Effect of Stated Behavioural Objectives on Achievement in a Unit of Instruction on Negative and Rational Base Systems of Numeration.' Unpublished Master's thesis, University of Maryland, College Park.

Furst, E.J. (1981) 'Bloom's taxonomy of educational objectives for the cognitive domain: philosophical and educational issues', _Review of Educational Research_, 51(4), 441-453.

Gagné, R.M. (1965) _The Conditions of Learning_. (1st Edition) London: Holt, Rinehart and Winston.

Gagné, R.M. (1976) _The Conditions of Learning_. (2nd Edition) London: Holt, Rinehart and Winston.

Gagné, R.M. (1985) _The Conditions of Learning_. (4th Edition) London: Holt, Rinehart and Winston.

Gagné, R.M. and Briggs, L.J. (1979) _Principles of Instructional Design_. (2nd Edition) London: Holt, Rinehart and Winston.

Gagné, R.M. and Paradise, N.E. (1961) 'Abilities and learning sets in knowledge acquisition', _Psychological Monographs_, 75(14), Whole No. 518.

Green, T.F. (1964) 'Teaching, acting and behaving', _Harvard Educational Review_, 34, 507-524,

Herbart, J.F. (1924) _The Science of Education_, London: Routledge and Kegan Paul.

Harrow, A.J. (1972) _A Taxonomy of the Psychomotor Domain_. N.Y.: McKay.

Keller, F.S. (1968) 'Goodbye, Teacher', _Journal of Applied Behaviour Analysis'_, 1, 79-87.

Krathwohl, D.R., Bloom, B.S. and Masia, B.B. (1964) _Taxonomy of Educational Objectives Handbook II: Affective Domain_. N.Y.: McKay.

Kropp, R.P and Stoker, H.W. (1966) _The Construction and Validation of Tests of the Cognitive Processes Described in The Taxonomy of Educational Objectives_. Washington, D.C.: U.S. Office of Education.

Kueter, R.A. (1970) 'Instructional Strategies: The Effect of Personality Factors on Recognition Learning Using Statements of Behavioural Objectives as Opposed to No Statements of Behavioural Objectives Prior to Learning.' Unpublished Doctoral dissertation, Indiana University.

MacDonald-Ross, M. (1973) 'Behavioural Objectives: A Critical Review', _Instructional Science_, 2, 1-52.

Madaus, G.F., Woods, E.M. and Nuttall, R.L. (1973) 'A causal model analysis of Bloom's taxonomy', _American Educational Research Journal_, 10, 253-262.

Mager, R.F. (1962/75) _Preparing Instructional Objectives_. Belmont: Pitman Learning.

Mager, R.F and Clark, C. (1963) 'Explorations in student controlled instruction', _Psychological Reports_, 13, 71-76.

Melton, R.F. (1978) 'Resolution of conflicting claims concerning the effect of behavioural objectives on student learning', _Review of Educational Research_, 48(2), 291-302.

Merrill, P.F. and Towle, N.J. (1971) 'Interaction of abilities and anxiety with availability of objectives and/or test items on computer-based task performance', _Proceedings of the 79th Annual Convention_, American Psychological Association, 539-540.

Morse, J.A. and Tillman, M.H. (1972) _Effects on Achievement Of Possession of Behavioural Objectives and Training Concerning Their Use_. Paper presented at the meeting of the American Educational Research Association, Chicago.

Ormell, C.P. (1974) 'Bloom's taxonomy and the objectives of education', _Educational Research_, 17, 3-18.

Ormell, C.P. (1979) 'The problem of analysing understanding', _Educational Research_, 22, 32-38.

Pask, G. (1972) 'A fresh look at cognition and the individual', _International Journal of Man-Machine Studies_, 4, 211-216.

Popham, J.W. (1969) 'Objectives and instruction'. In W.J. Popham, E.W. Eisner, H.J. Sullivan and L.L. Tyler (Eds) _Instructional Objectives_. Chicago: Rand McNally.

Pring, R. (1971) 'Bloom's taxonomy: a philosophical critique (2)', _Cambridge Journal of Education_, 1, 83-91.

Rothkopf, E.Z. and Kaplan, R. (1972) 'Exploration of the effect of density and specificity of instructional objectives on learning from text', _Journal of Educational Psychology_, 63, 295-302.

Rothkopf, E.Z. and Kaplan, R. (1974) 'Instructional objectives as directions to learners: effect of passage length and amount of objective-relevant content', _Journal of Educational Psychology_, 4, 448-456.

Rowntree, D. (1982) _Educational Technology in Curriculum Development_. London: Harper and Row.

Seddon, G.M. (1978) 'The properties of Bloom's taxonomy of educational objectives for the cognitive domain', _Review of Educational Research_, 48, 303-323.

Simpson, E.J. (1966) _The Classification of Educational Objectives: Psychomotor Domain_. University of Illinois, Research Project No.: OE 5-85-104.

Smith, J.M. (1970) 'Relations among Behavioural Objectives, Time of Acquisition, and Retention.' Unpublished Doctoral dissertation, University of Maryland.

Sockett, H. (1971) 'Bloom's taxonomy: a philosophical critique (1)', _Cambridge Journal of Education_, 1, 16-25.

Spencer, H. (1910) _Education: Intellectual, Moral and Physical_, London: Appleton.

Travers, R.M.W. (1980) 'Taxonomies of educational objectives and theories of classification', _Educational Evaluation and Policy Analysis_, 2, 5-23.

Tyler, R.W. (1932) 'The construction of examinations in botany and zoology', _Service Studies in Higher Education_, Ohio State University, Bureau of Educational Research Monographs, 15, 49-50.

Tyler, R.W. (1949) _Basic Principles of Curriculum and Instruction_. Chicago: University of Chicago Press.

CHAPTER 4

MASTERING LEARNING

Although early attempts at developing a scientific approach to learning were not always as successful as their proponents anticipated, they have led to an approach to learning that does produce consistent benefits in terms of achievement, benefits that are educationally significant.

Mastery learning is embodied in Bloom's (1968) 'Learning For Mastery' and Keller's (1968) 'Personalized System of Instruction.' Bloom's approach was developed to exploit mastery learning in the schoolroom, whereas Keller developed his system for higher education. However, both have been applied in many different contexts and have been found to be very powerful methods for increasing student performance in a wide range of activities.

BLOOM'S LEARNING FOR MASTERY

Bloom (1968) indicated that a learning strategy for mastery may be derived from the work of Carroll (1963), with support from the ideas of Morrison (1926), Skinner (1954) and Bruner (1966). Block (1971) traces the influences back to Washburne's (1922) Winnetka Plan and Morrison's approach at the University of Chicago's Laboratory School of 1926. In both systems mastery was defined in terms of particular educational objectives and mastery of each unit was required of students before they proceeded to the next unit. An ungraded diagnostic-progress test was administered at the completion of each unit to provide feedback on the adequacy of the student's learning. The test either indicated mastery, usually set at a level of 80-90%, or it highlighted the material the student still had to master. On the basis of the results of the diagnostic test each student's original instruction could be supplemented with appropriate remedial materials so that he could complete his unit, by obtaining a score reaching the mastery criterion. Although popular well into the 1930's these ideas eventually disappeared due primarily to a lack of the technology to sustain them (Block, 1971). They didn't emerge again until the late 1950's and 1960's, as a corollary of programmed instruction. The basic idea underlying programmed instruction being that the learning of any behaviour, no matter how complex, rested upon the learning of a sequence of less complex component behaviours. Theoretically, by breaking down complex behaviour into a chain of component parts and by ensuring mastery of every link in the chain, it

would be possible for any student to master even the most complex skills.

The research in programmed instruction showed that it worked very well for some students, especially those requiring small steps, drill and frequent reinforcement, but it was not so effective for all students, and it did not provide a useful mastery model in itself.

Bloom (1968) based his 'Learning For Mastery' (LFM) on the model proposed by J. Carroll (1963). This requires 'that the task can be unequivocally described and that means can be found for making valid judgement as to when the learner has accomplished the learning task - that is, has achieved the learning goal which has been set for him.'

A Model of School Learning

Carroll made the assumption that the work of the school can be broken down into a series of learning tasks. Although he admitted that this can be called into question because, in the school, the various tasks to be learned are not necessarily treated as being separate and distinct, and the process of teaching is often organised so that learning will take place 'incidentally' and in the course of other activities.

Some activities are not included as being appropriate for the model of learning that Carroll has in mind. For example, it is not intended to apply to those goals of the school which do not lend themselves to being considered as learning tasks, such as those having to do with attitudes and dispositions.

The model says that the learner will succeed in learning a given task to the extent that he spends the amount of time that he needs to learn the task. 'Spending time' means actually spending time on the act of learning. It is not 'elapsed time', but the time during which the person is oriented to the learning task and actively engaged in learning.

The variables in the model are considered under two headings: 1) determinants of time needed for learning, and 2) determinants of time spent in learning.

Time Needed in Learning

There are three factors affecting the time needed in learning, according to Carroll: Aptitude, Quality of Instruction, and Ability to Understand Instruction.

1. Aptitude

Common experience, as well as abundant research evidence, suggests that the amounts of time needed by the children even under ideal conditions will differ widely. Carroll suggests that the amount of time each pupil will need to learn a task under optimal conditions is the primary measure of a variable he calls 'aptitude for learning this task.' Learners who need only a small amount of time are said to have high aptitude; learners who need a large amount of time are said to have low aptitude. Some learners, Carroll admits, will never learn even under these optimal conditions; and these learners would need an indefinitely large (or an infinite) amount of time to learn the task.

The measure of aptitude is specific to the task under consideration. It may be regarded as a function of numerous other variables, such as the amount of prior learning or it may also depend upon a series of traits or characteristics of the learner which enter into a wide variety of tasks. These traits may be accounted for solely on the basis of generalized prior learnings, or they may reflect genetically determined individual characteristics.

Bloom (1968) contrasts this view with a common view that high levels of achievement are possible only for the most able students. Implicit in Carroll's formulation is the idea that all students can achieve mastery of a given task if they are provided with sufficient time. Therefore, mastery is available to all, if the right means can be found to help each individual. Bloom quotes evidence to support this, based on the grade norms for many standardized achievement tests. Such norms show that certain scores attained by the most able students at one grade level are achieved by the majority of students at a later grade.

Bloom believes that there is likely to be a small percentage of students (1%-5%), with special talents for a given subject, who are able to learn and use a subject with greater fluency than other students. These may be inherited, or may be developed by special training or interests. There is another small percentage of individuals who have extreme difficulty in learning particular subjects. Between these extremes are approximately 90% of the individuals for whom, Carroll and Bloom believe, aptitudes are predictive of rate of learning rather than the level or complexity of learning that is possible.

This means that 95% of students are capable of learning a subject to a high level of mastery, with up to 95% of a class being capable of attaining the equivalent of an A-grade, previously associated with the top 5-10%. Bloom assumes that for some students more time, effort and help may be required to reach this level, and for some the time and help required may make such achievement prohibitive.

Thus, a major problem for mastery-learning strategists is to find a way of reducing the time and effort that the less able students are required to make, to a point where it is not prohibitive. Bloom points out that students' aptitudes can be changed by appropriate learning experiences in the school and in the home, and that such changes can be most effectively brought about in the elementary years of schooling and during the early years in the home. Even if such changes cannot be effected in the early years it is highly probable that more effective learning conditions can reduce the amount of time required for learning to mastery.

2. Quality of Instruction

Carroll states that one job of the teacher is to organize and present the task to be learned in such a way that the learner can learn it as rapidly and as efficiently as he is able. This means that the learner must be put into adequate sensory contact with the material to be learned and that the various aspects of the learning task must be presented in such an order and with such detail that every step of the learning is adequately prepared for by a previous step. It may also mean that the instruction must be adapted for the special needs and characteristics of the learner, including his stage of learning. This variable applies not only to the performance of a teacher but also to the characteristics of textbooks, workbooks, films, and teaching programmes.

If the quality of instruction is anything less than optimal, it is probable that the learner will need more time to learn the task than he would otherwise need, although some learners will be more handicapped by poor instruction than others.

Bloom (1968) is concerned that we have fallen into the trap of looking at the class of 30 or so students and asking the questions 'What is the best teacher for the group?'; 'What is the best method of instruction for the group?'; 'What is the best instructional material for the group?' He maintains that it is necessary to start from a very different perspective in which each individual student may require very different types and qualities of instruction. In other words 'the same content and objectives of instruction may be learned by different students as the result of very different types of instruction.' For example, some students may need more examples, some may require more approval or reinforcement, or more concrete illustrations or explanations.

The example of home tutoring of middle-class students is quoted by Bloom to illustrate the benefits of the quality of instruction suited to the student's needs. A third of the middle-class students in one study received home tuition in algebra when parents saw the school as failing to provide the necessary tuition. These students received relatively high grades for algebra, with the relationship between mathematics aptitude tests at the start of the course and their achievement at the end of the year being almost zero. With no extra tuition the relationship between aptitude and achievement was highly correlated $(+0.9)$.

3. Ability to Understand Instruction

A further factor affecting the time needed in learning is the ability to understand instruction. Carroll suggests that this variable interacts with the method of instruction in a special and interesting way. It could be measured as some combination of 'general intelligence' and 'verbal ability'; the former of these two would come into play 'in instructional situations where the learner is left to infer for himself the concepts and relationships inherent in the material to be learned', while the latter would come into play 'whenever the instruction utilized language beyond the grasp of the learner.'

Bloom comments that in schools there is frequently a single teacher and a single set of instructional materials for all students in a given class. The student who has facility in understanding the teacher's communications has little difficulty in learning the subject. However, the student with the same basic aptitude for the subject but lacking the ability to understand the teacher or the text will experience considerable difficulty in learning.

This is the point at which the student's abilities interact with the instructional materials. For Bloom this ability will primarily be determined by verbal ability and reading comprehension because schools place such great stress on these abilities.

Bloom (1964) in 'Stability and Change in Human Characteristics' presents evidence that it is possible to produce changes in verbal ability early in life by appropriate training, but that such changes are more difficult later on in life, although vocabulary and reading ability may be improved to some extent at all ages. Therefore, the greatest immediate pay-off in this area will come from modifications in the instruction. Bloom suggests that given help and various types of aids, individual teachers can find ways of modifying their instruction to fit the differing needs of their students. He gives examples of how to overcome this problem within the school context:

1) Group Study: Small groups of students meeting regularly to go over points of difficulty were found to

be most effective, especially if this could be arranged in a non-competitive atmosphere, with more able students being provided with opportunities to strengthen their understanding while helping less able students to grasp ideas presented through alternative ways of explanation.

2) Tutorial Help: This is very costly, but should be made available to students as they require it, especially where difficulties cannot be corrected in any other way.

3) Text Books: The provision of various text books providing instruction for different abilities, rather than adopting one set book, can obviously be of benefit.

4) Workbooks and Programmed Texts: Drill and practice or small steps with frequent reinforcement may·help some students and programmed texts or workbooks can provide these.

5) Audio-visual Methods and Academic Games: Bloom suggests that some students may learn a particular idea from concrete illustrations and vivid explanations which AV methods (video tapes, filmstrips etc.) can provide. Other students may require direct laboratory experiences, games and computer simulations. In other words the highest priority need not always be given to abstract and verbal ways of instruction. The suggestion is not for particular materials to be used by particular students throughout the course, but for such materials as are appropriate to be used in helping students to overcome their difficulties.

Time Spent in Learning

In Carroll's model there are two factors which directly affect the time which the student actually spends in learning.

1. Time Allowed for Learning or 'Opportunity'

This is the first factor in the time actually spent in learning. Carroll comments that schools may allow less than adequate time for many students learning many tasks. Schools respond to differences in learning rates in many ways, in particular, they ignore these differences; a certain amount of time is provided for everybody to learn, and no more. At the other extreme is the case where each student is allowed to proceed exactly at his own rate; private instruction in

music or foreign languages and self-instruction by teaching-machine or other means are approximations to this case. The middle position is occupied by learning situations in which there is some kind of 'ability grouping.'

Obviously, failure to allow enough time for learning produces incomplete learning. If a person needs two hours to learn something and is allowed only one hour, and if we assume that learning proceeds linearly with time, the degree of learning is only 50%. Carroll feels that one of the most aversive things which a school can do is not to allow sufficient time for a well-motivated child to master a given learning task before the next step is taken up. Children meet such frustrations by indifference or the more extreme avoidance reactions and are, in any case, handicapped in undertaking the next task.

This is the key to mastery. Aptitude and other factors determine the rate of learning and all students can achieve mastery if they devote the amount of time needed, and for this they must be provided with sufficient time.

However, it is not just the amount of time that accounts for the level of achievement, but the amount of time in relation to other factors such as quality of instruction, aptitude and verbal ability and also the student's ability to persevere with the task in hand.

2. Perseverance

The second factor is the amount of time the learner is willing to spend in the learning process. Carroll describes this as consisting of three attitudes: (1) a marked willingness to spend time beyond the ordinary schedule, in a given task; (2) a willingness to withstand discomfort, which includes adjusting to shortened lunch hours, or no lunch hours, working without holidays and withstanding fatigue; (3) a willingness to face failure, which brings with it a realization that patient work may lead to successful termination of the task in hand.

A learner who needs a certain amount of time to learn a task may or may not be willing to persevere for that amount of time in trying to learn. The learner may not be motivated to learn at all, or may regard the task as something too difficult for him to learn; in either case, he may spend no time at all in trying to learn. He may, of course, be so highly motivated that he would be willing to spend more time than he needs in order to reach a specified criterion of mastery.

This variable, if it is not sufficiently great to allow the learner to attain mastery, operates in the conceptual model to reduce the degree of learning. It is a function partly of 'motivation' or the desire to learn and may also

be a function of what are ordinarily called emotional variables. For example, one may desire to learn but be unable to endure frustrations caused by difficulties in the learning task.

Students approach different tasks with different degrees of perseverance and, if frustrated in learning, will reduce the time devoted to a subject. Bloom suggests that it may be difficult to manipulate the perseverance of a student and that changes in the type of instruction and learning materials may help students to master material for a given level of perseverance. Then, as a corollary of this action, the reward and perceived success can raise the student's level of perseverance. As students attain mastery they are likely to raise their level of perseverance in related areas.

Bloom found that with appropriate instructional resources the demands for perseverance may be reduced. Frequent feedback accompanied by specific help in instruction and materials can reduce not only the time taken but also the perseverance required. Improvement in the quality of instruction (explanations, illustrations, etc.) may also reduce the required perseverance. He concludes that:

> There seems to be little reason to make learning so difficult that only a small proportion of the students can persevere to mastery. Endurance and unusual perseverance may be appropriate for long-distance running - they are not great virtues in their own right. The emphasis should be on learning, not vague ideas of discipline and endurance. (Bloom, 1968/81, p.164)

Thus, the model involves five elements - three residing in the individual and two stemming from external conditions. Factors in the individual are (1) aptitude - the amount of time needed to learn the task under optimal conditions, (2) ability to understand instruction, and (3) perseverance - the amount of time the learner is willing to engage actively in learning. Factors in external conditions are (4) opportunity - the time allowed for learning, and (5) the quality of instruction - a measure of the degree to which instruction is presented so that it will not require additional time for mastery beyond that required in view of aptitude.

Assessment Criteria

Carroll's model advocates a 'criterion referenced' assessment procedure, as suggested by Tyler (1932), in which the student's performance is judged according to how well he

has done by comparison with some predetermined criterion. This contrasts with the traditional 'norm referenced' assessment in which a given student is judged against the performance of his colleagues.

Criterion referenced tests provide a standard against which all student performances can be measured, and can ensure a minimum standard of competence. The driving test is a good example of this kind of assessment: all learners take the test and if they don't reach the criterion level they fail but may repeat the test until they reach the set level of competence. Norm referencing of a group at a driving school may result in a good distribution of grades from fail to A-grade, but this would not necessarily mean that even the A-grade students were able to reach the criterion level. It would mean that they could out-perform their colleagues, but the overall standard could still be below that required for safe driving on the highway.

The programmed learning movement popularized criterion referencing, with the programme writer producing a criterion test related to the objectives of his programme. The aim of the programme was to teach in such a way that 90% of the students obtained 90% on the criterion test. To attain this the programme writer would test the programme and revise areas of the text which produced low scores in the test. He would have to:

> resist the pressure from psychometricians who have had him insert some extra-hard questions and cut out easy ones - 'in order to spread students out a bit and show that some have achieved more than others.' He would reply, 'No, my intention is not to show whether some have learned more than others but to sow whether or not they have learned well enough to have attained the objectives. If all score 100% I shall be well pleased.' (Rowntree, 1977, p.180)

With norm referenced tests the aim is to produce the sort of achievement distribution shown in Figure 4-1, whereas the criterion referenced test should produce the achievement results shown in Figure 4-2.

Bloom (1968) took these ideas and transformed Carroll's conceptual model into an effective working model for mastery. He argued that if aptitudes were predictive of the rate at which (and not the level to which) a student could learn a given task, then it should be possible to fix the degree of learning expected of students at some level of mastery and to systematically manipulate the variables in Carroll's model so that all or almost all students attained it. He also reasoned that if aptitude for a subject was normally distributed in the student population and they were provided with uniform instruction in terms of quality and

Figure 4-1. Aptitude-Achievement for Uniform Instruction per Learner.

UNIFORM INSTRUCTION
PER LEARNER

aptitude   achievement
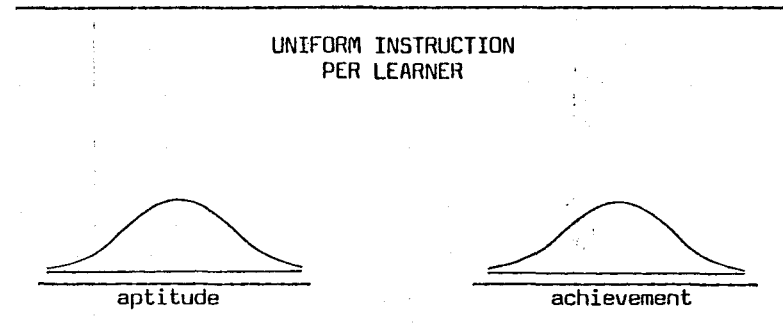
learning time, then achievement on completion of instruction would be normally distributed (Figure 4-1).

However, if the learner were to receive optimal quality of instruction and learning time then a majority of students could expect to attain mastery and there would be little or no relationship between aptitude and achievement (Figure 4-2).

Learning Hierarchies

Bloom also incorporated Gagné's theory of learning hierarchies in his learning for mastery approach. This theory states that the learning of any intellectual skill is important because it supports the learning of more complex skills. The support is only there if the previously learned skills are readily available when the new learning is taking place. Gagné suggests that mastery means 'readily accessible to recall at the time of learning of the more complex skill.'

Learning hierarchy research started with Gagné's early work with programmed learning of mathematical skills. In the first investigation a programmed book was used to teach a hierarchy of twenty-two elements leading to the ability to solve linear equations. Gagné felt that each element in the hierarchy could not be acquired unless the learners possessed those prerequisite skills found at lower levels in the hierarchy. These prerequisites could be determined by asking 'What would the individual have to be able to do in order that he can attain successful performance on this task, provided he is given only instructions?' (Gagné, 1962, p.358). He proposed that transfer from one learning set to another standing above it in the hierarchy will be zero if the lower one cannot be recalled, and will be as much as

Figure 4-2. Aptitude-Achievement for Optimal Instruction per Learner.

OPTIMAL INSTRUCTION
PER LEARNER

aptitude   achievement
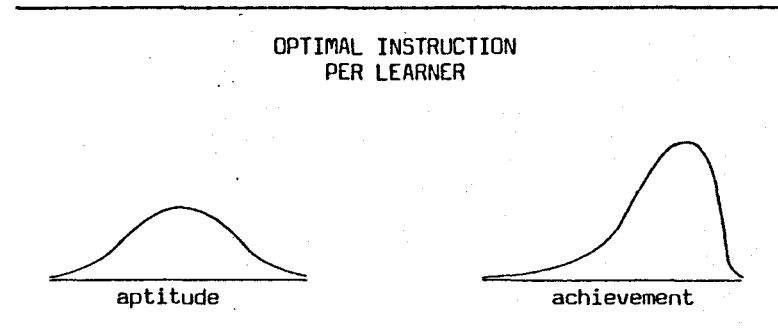
100% if it can be recalled. From this it follows that attaining each new learning set is dependent on recall of previous subordinate sets interacting with the effects of instructions.

> In order for learning to occur at any point in the hierarchy, according to this theory, each of the learning sets subordinate to a given task must be highly recallable, and integrated by a thinking process into the solution of the problem posed by the task. The attainment of the final task is thus conceived to be a matter of successive attainment and "integration" of a series of lower level learning sets, beginning with those which are already available to the learner.
> (Gagné and Paradise, 1961, p.2)

The results tended to support Gagné's theory and the numbers of subjects who learned the higher elements without possessing the relevant subordinate ones were small but not zero, as they should have been if the hierarchy was correct (Gagné and Paradise, 1961). A similar result, but with fewer exceptions to the postulated hierarchy, was found in the next experimental investigation of mathematical subject matter (Gagné, Mayor, Garstens and Paradise, 1962).

Despite the importance of the implications of Gagné's early work, few investigations of the validity of hierarchies were performed in the subsequent decade (White, 1973). Those that did take place did not always confirm the existence of the postulated hierarchy. For example, Kolb (1967-8) found that many subjects passed tests of higher elements dealing with basic ideas of ratio and line-segment graphs, having failed at the lower ones. Although in later experiments subjects behaved in ways which did not fit the proposed hierarchies and consequently brought the theory

into disrepute, White (1973) suggested that the studies were at fault because of their unsatisfactory design. Bergan (1980), however, disagreed with White and suggested that the results arose from the limitations in the learning hierarchy itself. One limitation is that Gagné's model does not include a set of equations describing the relationships among variables in the model. Such equations could be beneficial in estimating the magnitude of influence of variables in the model. The major limitation seems to be the restriction of the description of the relationships depicted in the model to two conditions: a skill is either prerequisite or it is not, no other possibilities are allowed. This fails to take account of the varieties of relationships that could exist.

> There could be reciprocal causal relationships between skills on the same level in a hierarchy. Unspecified variables outside the model could be exerting a significant influence on skill learning. There could be relationships among variables that have no simple causal interpretation. Finally, a skill placed in a hierarchy by virtue of one prerequisite relationship could influence several other skills at different levels in the hierarchy without being prerequisite to any of them. (Bergan, 1980, p.628)

There is nevertheless great appeal in the idea of learning hierarchies and it is an integral part of Bloom's approach to learning for mastery. Bloom (1976) does, however, acknowledge that not all learning tasks contain hierarchical arrangements of elements.

Learning For Mastery Strategies

Bloom (1968) acknowledged the fact that there were many different strategies for mastery learning, but that each strategy should deal effectively with the variables described by Carroll (1963).

He suggested that one ideal strategy would be to have a good tutor for each individual student, but acknowledged that this is not possible on a large scale on economic grounds. Nevertheless, the tutor-student relationship is a useful model to consider when working out the details of a more economically viable strategy.

Other strategies suggested include allowing students to go at their own pace, guiding students as to the most suitable courses, and providing different tracks or streams.

Bloom and his colleagues, following in Morrison's footsteps at the University of Chicago, experimented with various mastery strategies using diagnostic procedures and

alternative methods and materials as supplements to regular classroom instruction. With such an approach the aim was to bring all students up to mastery standard within the regular term or period of calendar time in which the course was usually taught.

The strategy required specification of the 'preconditions' necessary, the development of 'operating procedures' and evaluation of the 'outcomes' of the strategy.

Pre-conditions

To develop mastery learning it is necessary, according to Bloom, to define what mastery is and whether or not a student has achieved it. Specification of objectives and content of instruction are, therefore, necessary preconditions. The translation of these into evaluation procedures is also necessary. Implicit in defining outcomes and evaluation procedures is the distinction between the teaching-learning process and the evaluation process. The teaching-learning is intended to prepare a student for achievement of a specific level of competence in an area of learning. The summative evaluation at the end of a course is intended to determine the extent to which this has been achieved. This is not a competitive approach to education because Bloom sees much of learning and development being destroyed by primary emphasis on competition.

The LFM approach relies on setting standards of mastery and excellence, followed by a strategy to bring as many students as possible to this standard. One method employed was to base the standards for the new approach on those achieved by the top group prior to the introduction of the mastery approach. In this way students were informed of the performance required but were not in competition for grades, they were to be judged on the basis of levels of mastery actually obtained by students in the previous year. This enabled a more co-operative approach, students helping each other without the fear that special advantage was being given.

Operating Procedures

Bloom's mastery approach was based on materials and methods used in previous years, within a similar timetable, in the belief that requiring extensive training of teachers would limit the acceptance of the new approach.

The operating procedures were used as supplements to the regular teaching. They were devised to ensure mastery of each learning unit in such a way as to reduce time required

by altering quality of instruction and ability to understand the instruction. Two main operating procedures were adopted:

1. Formative and Summative Evaluation

The course was broken down into small units of learning, such as a chapter in a textbook, a time unit of the course or a well defined portion of a course, usually a week or two of learning activity.

Based on the ideas of Bloom (1956) and Gagné (1965) each unit was analysed into a number of elements ranging from specific terms or facts, to more complex abstract ideas such as concepts or principles and finally to applications of principles and the analysis of complex theoretical statements. Much of this initial work was conducted with such subjects as algebra and science topics, which fitted in well with Gagné's theoretical position on hierarchies of learning tasks.

Following from this breaking down of subject matter brief diagnostic progress tests were devised, which were then used to determine the extent of student mastery of a given unit. Such tests also indicated what the student still had to accomplish when he failed to reach the mastery criterion. These diagnostic tests were referred to as Formative Evaluation.

Frequent formative evaluation paces the students and helps motivate them, according to mastery theory, ensuring that each unit of learning is mastered before subsequent learning tasks are started. Early parts of the course may require more frequent formative testing than later sections, but there will be some variation in testing in various parts of a course.

Bloom believed that, for students who thoroughly mastered the unit, formative evaluation would reinforce the learning and reduce course anxiety about achievement.

Students lacking mastery of a particular unit had the specific deficits identified and particular instructional materials or processes were recommended to correct the learning. A very specific prescription was seen as being essential for students failing a formative test.

The test was not graded other than as showing mastery or non-mastery because Bloom believed that giving different point-grades (A, B, C, etc.) could be counter-productive.

Thus, the main purpose of the formative evaluation was to determine the degree of mastery of a given learning task and to identify those aspects of the task which were not mastered.

The purpose is not to grade or certify the learner; it is to help both the learner and the teacher focus upon the particular learning necessary for movement toward mastery. (Bloom, Hastings and Madaus, 1971, p.61)

The term 'summative evaluation' was given to the much more general assessment of the degree to which the 'larger outcomes have been attained over the entire course.' Such summative evaluation could be used to grade pupils and to report to parents or administrators. It was felt that several skills or concepts which combine to make a broader ability should be presented before a summative examination is administered. Summative tests should not be thought of only as end of course or final examinations, they could equally be used every four or six weeks to assess mid-term attainment of the various abilities taught in a course.

Bloom, Hastings and Madaus (1971) suggest that it is the level of generalization which differentiates summative from formative evaluation, and they indicate that summative assessment is linked to Tyler's more general descriptions of desired behaviours, whereas formative evaluation is more closely allied to Gagné's detailed prerequisite capabilities for each large aim.

2. Alternative Learning Resources

Bloom found that students do attempt to work on their difficulties if they are given specific suggestions of what to do. The best procedure identified was to have small groups of students (two or three) meet regularly for up to an hour each week to review the results of formative tests and to help each other overcome difficulties. Tutorial assistance was found not to be as popular as peer group meetings for secondary and higher education students. Other types of learning resource prescriptions, recommended by Bloom, were:

1. Rereading pages of the original instructional materials.
2. Reading alternative materials.
3. Use of specific pages of workbooks or programmed texts.
4. Use of selected a-v aids.

Ryan and Schmidt (1979) identified the most successful corrective strategies as being those which include objectives plus a problem testing the objectives of the previous lesson, discussion of the problem, specific prescriptions for using the text, class notes and handouts, and alternative resources, such as texts, workbooks, games and kits. When correctives consisted of objectives or

problems only their effectiveness was considerably diminished.

Affective and Cognitive Consequences of LFM

Bloom suggested that if a system of formative and summative evaluation informs the student of his mastery of the subject, he will come to believe in his own mastery and competence and this will change his view of himself and the world.

He will certainly begin to respond more positively to the subject he has mastered. Mastery will provide the necessary reassurance and reinforcement to help the student to view himself as adequate thus contributing to positive mental health:

> Finally, modern society requires continual learning throughout life. If the schools do not promote adequate learning and reassurance of progress, the student must come to reject learning - both in the school and later in life. Mastery learning can give zest to school learning and can develop a life-long interest in learning. It is this continual learning which should be the major goal of the educational system. (Bloom, 1968/81, p.174)

Bloom's initial description of the cognitive outcomes described results for an experiment conducted from 1965 to 1967. Using the standards of grading adopted in 1965, prior to the introduction of a mastery approach, the aim was to use a variety of techniques during ensuing years to turn all students into 'A' students (ie. all students in the years 1966 and 1967 would obtain the same grades, using the same criteria, as were obtained by the top students of 1965). The results indicated that for 1967 80% of students obtained the level of mastery previously associated with 'A' grade students and by 1967 this had risen to 90% of students. The techniques certainly appeared to work in these initial studies and led to further investigations.

Learning For Mastery: The Research

Bloom (1976) summarized the early research on the mastery approach in 'Human Characteristics and School Learning.' These were 'a series of small experimental and school-based studies' in which students from the University of Chicago contrasted mastery with traditional methods.

Cognitive entry behaviours were measured at the beginning of each learning task, using a formative diagnostic test, and students were tested again at the end of the teaching session. Those students in the mastery groups received additional tuition, as indicated by the formative diagnostic test, if they failed to reach the set mastery criterion. They were then tested again after these 'corrective procedures' using a parallel form of formative test.

If Bloom's theory is correct we would expect a student's performance on one learning task to be highly correlated with performance on the next task, in the traditional group. The best students in this group would perform well on the first task and this would enable them to perform at an equally high level on the next task.

This would not be the case in the mastery group because those students who were failing to reach mastery on the initial task would receive appropriate materials and time to enable them to reach a position previously associated with the more able students. Having gained this position they would be just as likely to perform at a higher level in the next task. Thus, their performance on the original task would not be predictive of their performance on the second task. That at least is the theory, and it also predicts that there would be high correlations between one task score and the next task score for traditional courses, but much smaller (or even zero) correlations between one task and the next for mastery groups.

Bloom (1976) quotes studies covering subjects such as second language acquisition, matrix algebra, probability and an 'imaginary' science. The correlations between achievement on the first learning task in the series and the final (summative) achievement are:

	Mastery	Non-mastery
Anderson (1973)	0.31	0.68
Arlin (1973)	0.19	0.49
Block (1970)	0.44	0.78
Levin (1975)	0.59	0.72
Ozcelik (1974)	0.36	0.48
Median	0.36	0.68

These results demonstrate an overall agreement with Bloom's prediction from his theory, although the mastery correlations do not reach the zero figure that Bloom had hoped for. There is a definite trend for a much reduced correlation between initial performance and final performance under mastery conditions. The larger correlation of 0.68 for non-mastery students indicates the extent to which performance on the initial learning task will influence the final achievement when there are no correctives for defective learning.

Table 4-1. Mean performance scores for Mastery and Non-mastery groups.

	LEARNING TASK 1		LEARNING TASK 2		LEARNING TASK 3		
	Orig. Ach.	Corr. Ach.	Orig. Ach.	Corr. Ach.	Orig. Ach.	Corr. Ach.	Summ. Ach.
Mastery Groups	63	85	69	87	71	86	76
Non-mastery Groups	63		55		54		44

Orig. Ach. - Original achievement after instruction on learning task.
Corr. Ach. - Corrected achievement (mastery group only) after receiving additional corrective instruction.
Summ. Ach. - Summative achievement on completion of three tasks.

Of further interest is the average cumulative performance of the five experimental groups reported by Bloom (1976) given in Table 4.1. The non-mastery group does not receive correctives and, therefore, is represented by one score only for each of the three hierarchically arranged learning tasks. This is called the 'Original Achievement' score. The mastery group receives correctives after each learning task, for those failing to reach the criterion level, and takes a parallel form of the test giving a 'Corrected Achievement' score in addition to the 'Original Achievement' score. The difference between the 'Original Achievement' score and the 'Corrected Achievement' score represents the improvement in the mastery group following the correctives procedure. This improvement in performance for the mastery group results in a larger difference between the groups on the next learning task. With each set of correctives we see the gap between the mastery and non-mastery group increasing: initially, there is no difference; on the second task in the hierarchy there is a difference of 14%, in favour of the mastery group; and on the third task this has further increased to 17%. The final summative average difference is 32%, in favour of the mastery group.

These results were indicative of the scale of improvements possible with a mastery learning approach, and

Table 4-2. Student achievement in mastery and non-mastery conditions. (Based on Kim, 1971)

Instructional Treatment	Percentage of Students Achieving Mastery			
	English	Maths	Physics	Biology
Mastery	50	48	30	22
Non-Mastery	44	26	8	3

were confirmed by the research studies reviewed by Block (1974).

Block quotes the work of Kim (1971) and his colleagues (Lee et al., 1971) as showing the power of LFM. In these studies, using a group-based, mastery learning strategy, with pupil-teacher ratios of 70:1, Kim and his colleagues converted much of the Korean elementary and middle-school system into a mastery model. The pilot study taught 272 seventh grade students an eight-session unit on geometry. Kim found that 75% of the mastery group scored at the mastery criteria of 80% correct, compared with only 40% of the traditionally taught students reaching this level.

On the basis of this successful experiment Kim expanded his next study to 5,800 students, who studied maths and english over an eight week period. In English, 75% of mastery students reached the 80% criterion, compared with 28% of the traditional group; in maths, 67% of mastery students reached the criterion, against 39% of the traditional group.

The third experiment was much larger in scope. Rural and urban pupils learned maths, English, physics and biology over a full school year. There were 25,887 students in the experiment. The results were quite impressive (Table 4-2), with 2,000-3,000 more students gaining mastery in maths, physics and biology when compared with traditional methods.

Similar results were obtained by Lee et al. (1971), for arithmetic and science when taught to 12,504 pupils in 5th and 6th grades, with 27-32% more pupils reaching the 80% criterion under mastery conditions.

These differences in performance demonstrate the scale of achievement possible for students who are not reaching their maximum potential under normal instructional circumstances. Block (1974) applauds the magnitude of the successful application of mastery learning:

Here a difference of a few percentage points, let alone differences of 20 to 30 points, between the percentage of students reaching a high performance level under mastery conditions and the percentage reaching the same level under nonmastery learning conditions translates into differences of thousands *of students who are achieving levels in their learning they might never ordinarily have reached.* (Block, 1974, p.34)

Further research on mastery learning in the classroom was summarized by Block and Burns (1977) in 'Review of Research in Education' (vol. 4), edited by L.S. Shulman. In this later paper the authors reported a meta-analysis on the research, giving an average effect size for the LFM-taught students of 0.83, which would move the average student in the non-mastery group from the 50th position in class to the top 20% of the class, a large educational effect. Keller studies were also included in this review and before discussing the overall results a description of Keller's 'Personalized System Of Instruction' is necessary.

KELLER'S PERSONALIZED SYSTEM OF INSTRUCTION

Keller describes the the first steps towards his system as having taken place during an evening brain-storming session involving Keller and three other psychologist colleagues, who had been invited to set up a new department of Psychology in the University of Brazilia, in 1963. They were given absolute freedom to follow their own dictates and fancies in carrying out the task.

Whereas Bloom had adapted the work of Carroll and Gagné, Keller's group was overtly 'operant' in its orientation, or as Keller put it 'tarred with the brush of reinforcement theory.' They were excited by Skinner's Natural Science Course at Harvard University, and all were convinced that traditional teaching methods were hopelessly out-of-date, and were impressed by the teaching machine and programmed-instruction movement.

They were not the only ones disillusioned with traditional methods of instruction. In the late 1950s and early 1960s educational researchers raised some unsettling questions about the effectiveness of most college teaching. In a number of major studies, investigators found that they could reduce by more than two-thirds the amount of time students spent in the classroom without affecting end-of-course student achievement (Kulik and Jaksa, 1977). Dubin & Taveggia (1968) went even further and demonstrated that such typical classroom activities as lectures and discussions might be eliminated completely without altering student achievement. Pooling data from 91 studies conducted between 1924 and 1965 they showed that students learned as much from independent study as they did from more conventional approaches to teaching. In fact, one of the chapters in Dubin and Taveggia's 'The Teaching-Learning Paradox' is titled "THERE ARE NO DIFFERENCES", in very large capital letters.

On the basis of these research results many innovators believed that students could learn as much through the newer approaches as they had learned from the more traditional approaches. And some even believed that they might actually improve learning.

According to the behaviouristic approach of Keller and his colleagues students are more likely to perform well if they find satisfaction in their studies or in behaviouristic terms 'The consequences of learning are important instructional contingencies.'

This means that positive consequences (instructor praise, good grades, feelings of achievement) are considered to be much more effective facilitators of learning than negative consequences (boredom, failure, or other forms of punishment).

External rewards are important, but it was thought that for adult learners they were not nearly as important as the general internal rewards (feelings of achievement, satisfaction or accomplishment).

Keller's 'Personalized System of Instruction' (PSI), which is also known as the 'Keller Plan', is based on these principles. It specifies objectives and provides reinforcement for their successful achievement, as well as giving the student more options and opportunities for personal interaction than traditional instructional systems.

The system was first publicly announced in 1968 in Keller's article 'Goodbye, teacher', which appeared in the new 'Journal of Applied Behaviour Analysis'. It begins with the chant:

> Goodbye scholars, goodbye school
> Goodbye teacher, darned old fool

and goes on to describe Keller's reasons for being disillusioned with the traditional classroom and formal education. He also outlines some of his observations made during his time in a military training centre. This is what he says about that centre:

> I should have seen many things that I didn't see at all, or saw very dimly. I could have noted, for example, that instruction in such a centre was highly individualised, in spite of large classes, sometimes permitting students to advance at their own speed throughout a course of study. I could have seen the clear specification of

terminal skills for each course, together with the carefully graded steps leading to this end. I could have seen the demand for perfection at every level of training and for every student; the employment of classroom instructors who were little more than the successful graduates of earlier classes; the minimizing of the lecture as a teaching device and the maximizing of student participation. I could have seen, especially, an interesting division of labour in the educational process, wherein the non-commissioned, classroom teacher was restricted to duties of guiding, clarifying, demonstrating, testing, grading, and the like, while the commissioned teacher, the training officer, dealt with matters of course logistics, the interpretation of training manuals, the construction of lesson plans and guides, the evaluation of student progress, the selection of non-commissioned cadre, and the writing of reports for his superiors. (Keller, 1968, p.80)

Much of the Personalized System of Instruction can be seen as a development of Keller's observations of military training - perhaps one reason for the emphasis on instruction, rather than 'education' or 'teaching'. There are five main elements in PSI as described in the 1968 paper:

1. The unit-perfection requirement for advance, which lets the student go ahead to new material only after demonstrating mastery of the preceding unit.
2. The go-at-your-own pace feature, which permits a student to move through the course at a speed commensurate with his ability and other demands upon his time.
3. The stress upon the written word in teacher-student communications.
4. The use of proctors, which permits repeated testing, immediate scoring, almost unavoidable tutoring, and a marked enhancement of the personal-social aspect of the educational process.
5. The use of lectures and demonstrations as vehicles of motivation, rather than sources of critical information.

Mastery

Mastery learning proposes that all or almost all students can master what they are taught. It enables 90-95% of students to achieve the same high level as the top 25% learning under typical group-based instructional methods. Under such circumstances it is assumed that students will learn more material in less time and there will be greater student interest in and attitude toward the subject learned than is found in traditional classrooms.

Keller indicates that the mastery requirement means an essentially perfect performance. It is acceptable to interpret this as each student answering nine out of ten questions correctly, provided the single error is corrected. Any further compromise with mastery is unwise because:

> the requirement of eighty-five percent will automatically raise questions as to why eighty-three percent isn't 'close enough.' There will be requests for 'partial credit.' An adversary system with 'bargaining' about grades quickly returns to the classroom. The mutual respect and high morale upon which the course depends deteriorates. The course policy statement should specify that the criterion for a pass is perfection. (Keller and Sherman, 1974, p.37)

Self-pacing

One of the essential characteristics of mastery learning is the time spent in study and this leads to the emphasis, in the Keller Plan, placed on a completely student-paced programme of study.

It is important to stress that the system places no restraints on the student's study time - thus a student may elect to take a unit quiz at her convenience or when she is most ready to demonstrate mastery, rather than at a time for all students which is dictated by the instructor.

The self-pacing allows for individual differences in rate of study and is essential if the level of achievement is held constant. This is because in essence it is not possible to hold both level of individual achievement and rate of completion constant. Traditional lecture methods, for example, hold rate of completion constant and, because of this, level of achievement varies widely. However, in PSI courses tests can be taken at any time and can be repeated as required, and the result is that while the average student evidences a steady progress, at any one time individual students will be working on different units and taking different tests. Here we have mastery level of achievement held constant and so the rate of completion varies.

Written Materials

Keller in 'Goodbye, Teacher' summarizes his use of written materials by stating that PSI places stress upon the

written word in teacher-student communication. Although programmed texts are used, Keller emphasizes that the steps in a PSI programme are not frames in a set; they are more inclusive and can be better described as reading assignments or laboratory exercises. Also, he states in a later description of PSI that television, computers and teaching machines may all be used in the system but should not be equated with the system.

PSI courses are based upon a standard textbook, journal articles, other readings and are supplemented by a study guide. This guide breaks the course into units. Each guide consists of an introduction, statement of objectives, procedure, study questions, and supplementary materials as dictated by the specific course.

The study guide is designed to allow the student to work through the material independently and this function over-rides the form the guide might take. A typical guide has four sections:

1. Introduction: in which the instructor comments upon the textbook and supplements it. It may point out differing views to those expressed in the text and correct deficiencies.
2. Statement of objectives: which indicates the goals of the study unit in behavioural terms. It tells the student what he will be able to do on completion of the unit. The objectives should be very specific, but make the work interesting at the same time.
3. Study questions: these should help the student to detect his progress and success, and may include concepts or technical terms that require defining and problems for completion.
4. Procedure: this section specifies the activities necessary for completion of the unit objectives. It may consist only of the assignment to be read, or it may be more detailed, providing an indication of specific pages to be read and the order for working through different articles. A page-and-line key of special instructions may be included where the textbook is found to be inadequate.

Keller and Sherman (1974) indicate that a well-written procedure section tells the students what to do, how to self-test comprehension, how to decide whether to proceed or review and how to decide when they have finished.

The study guide may include any further information that helps the students to complete their work successfully, the main criterion for a study guide being that it works.

Student Proctors

These are usually students who have successfully completed the course. Their job is to score quizzes, discuss answers, tutor students who are experiencing difficulties, and provide feedback to the instructor on the course in general.

According to Keller, the proctor is a vital component of the system, being a primary source of external reward for the mastery of each unit; they constitute the personal contact between student and system and provide immediate feedback to the students. The proctor allows interaction between the student and the system and a degree of flexibility that can only come from human interaction: credit can be given to a student if he appears to understand the item tested, even though the answer to the item was incorrect due to some minor error. Also the proctor can discuss correct responses in order to determine whether the answer is guessed or arrived at for invalid reasons. This is what Keller says of the 'proctor':

> The proctor is the new link in the chain. His principal job is that of decreasing the gap of understanding between the student and teacher - a gap that is sometimes very wide.
> He can do this because he has a repertoire of behaviour that is intermediate between the two with respect to the subject matter of the course. He may find logic in an answer that an instructor would never have perceived; he may restate a question in a way that brings out evidence of knowledge; and he may by example or parallel statement strengthen a student's grasp of a concept. He provides the individual consideration that a student may never have had before.
> (Keller and Sherman, 1974, p.20-1)

The proctor is helped in his work by 4 factors:

1. his experience with a similar course in which he was recently successful;
2. a weekly proctor's meeting in which every question on every test may be discussed;
3. a handy list of acceptable answers to the questions, to jog his memory if needed;
4. usually, relevant knowledge from advanced courses within the same subject-matter field.

Those proctors, described above, are called by Keller 'grading proctors', and he estimates that ten or more would be required for a class of one hundred students. Two further

proctors may be involved in the general scheme: (1) the testing proctor; and (2) the study hall proctor.

1. Testing proctor: is responsible for all record-keeping and actually administering the tests. Keller says that this non-educational role may be undertaken by a reliable clerk or by grading proctors on a rotational basis.

2. Study hall proctor: this position is so important that it may be filled by the instructor or his assistant. He stands between the textbook and the student (just as the grading proctor stands between student and the instructor). He clarifies obscure passages, helps with difficult operations, explains unintelligible references etc. The study hall proctor generally serves a small percentage of the class (usually its weakest members) and the number of cries for help will decrease as study habits are improved and students gain in self-reliance.

Keller says that the study-hall proctor may be selected on the basis of course history, achievement or special aptitudes but also cautions that if he acts too much like a teacher he may not be the most effective promoter of learning in the classroom.

In addition to the proctors there is the classroom assistant whose tasks include distribution of assignments, study questions, checking of supplies, maintenance of progress charts etc. Later, he may take the place of the instructor at various meetings and in general give the instructor time to alter the course as and when necessary.

The instructor's role is that of a manager of learning. He selects and analyses the material to be mastered, he decides how to present it, and constructs the various questions based upon it. Keller emphasizes that the textbooks must be carefully read before selection to determine their suitability for students who will be questioned upon it in detail - errors, confusions and contradictions that might go unnoticed in a conventional system will stand out as 'unsightly blemishes' in PSI.

The instructor is also responsible for dividing material into suitable units, devising suitable study questions to help students and test questions for the mastery test. He will receive a great deal of information on everything he has devised for the course and as a result he will be responsible for the continual modification of the system, as he aims for perfection.

Lectures

Lectures are used for enrichment rather than to transmit information, which in PSI is done through the use of study guides and the text. Lectures are given to demonstrate the excitement of the discipline; and may be used to show films or for other special events such as visiting speakers. They serve a purely supplemental function. As Keller and Sherman point out:

A few lectures are useful. The instructor can serve as a model, impart style, even provide inspiration. In fact we may expect a very polished and challenging presentation now that such events are infrequent. Such lectures must be fine indeed; otherwise they will not attract an audience. Since the lecture is given for motivational reasons rather than for the transmittal of essential information, PSI lectures are 'extras' and not required. (Keller and Sherman, 1954, pp.41-42)

Personalized System of Instruction: The Research

In the 'Review of Research in Education' (1975) McKeachie & Kulik summarized the results of the available evaluations of PSI:

1. Keller plan is an attractive teaching method to most students. In published reports students rate Keller plan more favourably than teaching by lecture.
2. Self-pacing and interaction with tutors seem to be features most favoured by students. There are reports of higher-than-average student drop-out rates for Keller courses.
3. Content learning, as measured by final examinations, is adequate: in published studies final examination performance always equals and usually exceeds performance in lecture sections.
4. Students almost invariably report that they learn more from PSI and report expending more time and effort.

In addition to these generally favourable reports evidence was also accumulating which showed the relative importance of the various components of PSI in contributing to the system's overall successful performance.

Student Ratings of the Course

Gallup (1970) found favourable comments on a psychology

PSI course from 98% of students and Born & Herbert (1971) found that only 7 of 145 students would not recommend the course to other students. When looking at Library Science, Knightly & Sayre (1972) found that the self-paced mode of instruction was considered better than lecture courses by 100% of students; the course being rated as 'above average' or 'one of the best' by 93%.

Perhaps it should be noted here that in the early days many academics tended to describe PSI as a de-humanizing, programmed approach to instruction, and stress its links with Skinner's behaviourist orientation. Students who studied such courses, however, liked its freedom, self-determination and personal interaction!

Features Rated Highly by Students

Green (1971) categorized student comments about Keller courses and reported that the feature most often praised is the freedom of pace. In another study, Nelson & Scott (1972) had students rank 12 features of PSI. Most important were (1) self-pacing; (2) interactions with teaching assistant or instructor; (3) small unit steps; (4) discussions with proctors. Self-pacing was also seen to be the feature rated most highly in a similar study by Myers (1970). Clearly, for the students, the most attractive features of PSI are self-pacing and personalized interactions.

Green (1971) reported that 90% of students on Keller plan took further Keller options when offered; and Gallup (1974) reported that he had to limit student numbers on his course when it was introduced as a Keller type option.

This overall positive response to PSI was not consonant with McKeachie & Kulik's (1975) suggestion that student drop-out was higher than the average for conventional lecture courses. For example, Keller presented data that showed withdrawal rates of about 20% - no comparison data was available, but 20% does seem to be very high. Sheppard & MacDermot (1970) reported a drop-out rate of 17% for a Keller course on introductory course, compared with 6% for a traditional class. Born (1971) also presented data showing a withdrawal rate of 14%, 25% and 14% for Keller sections compared with 5%, 6% and 5% for conventional courses.

However, it is not the whole story. McMichael & Corey's (1969) results showed a high drop-out rate for the PSI course, but a slightly higher one for the conventional lecture course: 12% compared with 13%. Kulik, Kulik and Cohen (1979) in their meta-analysis of PSI research found that of 27 investigations which reported comparisons of course completion, the differences between PSI and conventional courses were trivial.

Procrastination did seem to be a major disadvantage of

such a self-paced system of instruction in the early days and Green (1971) and others suggested ways in which it could be prevented:

1. publishing a recommended schedule of dates for passing units, which will yield a steady rate of work,
2. making admittance to 'fun' lectures contingent upon passing units,
3. offering early final exams for students finishing the course early.

Content Learning

Kulik, Kulik and Smith (1976) presented details of 39 studies which satisfied their basic requirements of sound experimental design. In 38 of the 39 studies final exam performance was better in the PSI course. In 34 of the comparisons the difference was considered to be statistically reliable. The authors state that the differences were large enough to have practical significance: they show that for a typical result a student, who is average on a standard test, scoring 50% in a typical lectures-based course, would score 75% if she takes the PSI courses, indicating an Effect Size of 0.66. A later meta-analysis (Kulik, Kulik, and Cohen, 1979) found 61 studies which compared PSI and conventional teaching. PSI final examinations were 8 percentage points higher, giving an average effect size of 0.49. Also, there was 35% less variation in the scores for the PSI group. These results refer to final examinations. But what about the long-term impact PSI has on students? Again, Kulik, Kulik and Cohen (1979) indicated that PSI comes out with better results. In fact the differences were greater after the retention period (which varied from 3 weeks to 15 months), suggesting that final exams, if anything, underestimate the magnitude of PSI's effect.

Another way in which the effects of PSI can be investigated is the transfer study, where students study, for example, MATHS 1 by either PSI or conventional methods. They then study MATHS 2 together in a conventional class. How does PSI affect performance in MATHS 2 ? The transfer studies all indicate that those students who take the PSI option for the initial course out-perform conventionally taught students on the second course, even though the second course is a conventionally taught course.

Component Analysis

There can be no doubt that PSI is effective, the question is, why is PSI so effective? Various studies have attempted to throw some light on this question, and have looked at the various components of PSI.

The research studies tend to show that small units are important to the effectiveness of PSI - if the units become too long PSI's effectiveness deteriorates (Calhoun, 1976). Semb (1974) also found better performance in major exams for students working on smaller units. Nelson and Bennett (1973) and O'Neill et al (1975) found more consistent progress in shorter unit PSI courses.

Feedback also seems to be important and there are several studies in which groups receiving delayed feedback were compared with groups receiving immediate feedback (Calhoun, 1976; Farmer et al, 1972; and Johnson and Sulzer-Azaroff, 1974). The achievement of the delayed feedback groups was significantly lower than groups receiving immediate feedback.

Relaxation of the mastery criterion leads to a decline in the effectiveness of PSI, as was anticipated by Keller. Semb in 1974 showed that a high mastery criterion (100%) produced better final test performance than a low mastery criterion (60%) and this was confirmed in a later study (Phillips and Semb, 1975).

Surprisingly, the proctors don't seem to be an essential feature of PSI. Blackburn, Semb and Hopkins (1975) compared a Keller PSI course with a section which included self-grading (the proctors only acting as administrators). In the final exam both groups performed equally well. Calhoun (1976) compared written feedback with feedback provided by the proctor and again there were differences, but they were not statistically significant.

Research studies have also failed to show that self-pacing is important. Several studies show no difference on final examination or student evaluation for teacher-paced and self-paced groups (Beneke and Taylor, 1975; Calhoun, 1976; Lewis, 1972).

Although Keller recommends lectures in PSI courses as motivational agents, some instructors have built in conventional lectures in addition to the individualised work. Calhoun (1976) found that twice-weekly lectures did not improve the performance of PSI students and Minke and Carlson (1973) found that adding lectures did not change the percentage of students passing, the withdrawal rate or the overall evaluation of students passing the course.

LFM AND PSI: A REVIEW OF THE RESEARCH

The research on mastery learning as found in the approaches of Bloom and Keller was summarized in a comprehensive review by Block and Burns (1977). In this paper the authors reported a meta-analysis on the research, giving the effect size for LFM achievement comparisons and a combined retention effect size for PSI and LFM.

A total of 97 comparisons were identified, involving different types and numbers of students. Mastery students scored higher than non-mastery students 89% of the time, and statistically significantly higher 61% of the time. For retention (combining PSI and LFM) the mastery students were almost always superior to the non-mastery groups, significantly so 63% of the time. The average effect size for the LFM-taught students was 0.83. A similar large effect size was obtained from studies of LFM and PSI which included a retention test, with an average retention effect size of 0.67.

In addition, Block and Burns looked at the variability in learning. According to Bloom's theory the mastery approach will lead to greater uniformity in learning when compared to traditional methods, which should show increasing variability. In 80 studies comparing variance in achievement test scores, mastery students showed less variation in their scores 74% of the time, and less variability in retention scores 85% of the time. The reduction in variance was approximately 50% for both achievement and retention measures.

Kinds of Learning

Block and Burns also looked at learning in qualitative terms. They found some evidence to suggest that mastery strategies may be more likely to elicit higher order than lower order behaviours. Several researches found that mastery students were superior on essay-type questions that included comprehension, application, analysis, synthesis and evaluation skills, and that mastery students were more able to deal with these complex high order skills than the lower order multiple-choice knowledge questions.

There was some evidence, however, that for retention periods of 6 or 18 months it was the lower order behaviours associated with knowledge, rather than the complex application or analysis-synthesis, that were retained at a higher level in the mastery group.

The mastery approaches typically elicited more favourable affective responses from students, in particular there was a positive impact on students' interest in and attitudes to the subject matter, and also on self-concept,

academic self-confidence, attitudes to co-operative learning and towards instruction. Although there is some evidence to suggest that mastery set at the 100% level adversely affects such attitudes and that an optimum figure of 90% maintains both the positive affective and achievement responses.

Time Consequences

A group of studies investigated the amount of time spent in learning in mastery and non-mastery groups. Uniformly, they found that mastery strategies had a homogenizing effect on differences in study time. One study found that differences were reduced from 7 to 1, to 4 to 1 by mastery methods, and that there was progressively less variability in the mastery study time. In the final unit of eight there was 90% less variability

Block and Burns comment that this reduction in individual differences in study time is purchased at the price of additional study time for the slower students. Over the short term there is an increase in the average study time required for mastery of 70-80%, compared with non-mastery. In the long term this may be reduced to 20-50%

The extra time is available from two sources. The first is the teacher who is encouraged to prepare and organize the instructional plans, procedures and materials outside of class time and prior to the instruction. This allows more time for interaction with the students. The student can also be encouraged to exert more effort, especially in overcoming procrastination, and various incentive schemes have been used, such as frequent testing, varying the size of learning units, teacher/peer pacing of instruction. The emphasis is clearly on techniques that encourage teachers and students to spend more time learning as opposed to non-learning activities.

Achievements of the Behavioural Systems

In many ways the work of Bloom and Keller in their mastery learning systems represents the summit of achievement for the neo-behaviourist 'systems' approach to instruction. Both were built on firm, though differing, psychological principles and of all the applications of the behavioural science approach to the technology of education, LFM and PSI represent the only methods which consistently produce significant educational results. Clark (1983) attributes this to the emphasis placed on method, defined by Glaser (1976) as 'the conditions which can be implemented to foster acquisition of competence,' rather than on the media

involved in the delivery of instruction, as represented by the physical science approach.

It is fitting that Benjamin S. Bloom, whose work has been central to the development of instructional science for a generation, should have the last words:

> If you can be moved to try these ideas with a few teachers in your school for even as short a period as three months, you can determine the validity and limits of these ideas where they belong - in your classrooms and with your teachers and students.

He continues:

> However, neither further opportunity for education nor increased financial support for education will do much to improve the education of each of our students. The answer does not lie in additional funds, new fads, or major and sweeping changes in the organization of our educational system. As I see it, the solution lies in our views about students and their learning. These have grown out of our practices and they will not be changed until we alter these practices. When the changed practices succeed in promoting more effective learning, both teachers and students will change their views.
> (Bloom, 1978, p.563)

REFERENCES

Beneke, W.M. and Taylor, D.C. (1975) *Effects of Unit Deadlines on Performances in a PSI Course*. Paper presented at the annual meeting of the American Psychological Association, Chicago.

Bergan, J.R. (1980) 'The structural analysis of behaviour: an alternative to the learning-hierarchy model', *Review of Educational Research*, 50, 4, 625-646.

Blackburn, T., Semb, G. and Hopkins, B.L. (1975) 'The comparative effects of self-grading versus proctor grading on class efficiency and student performance'. In J.M. Johnson (Ed) *Behaviour Research and Technology in Higher Education*. Springfield Ill.: Charles C. Thomas.

Block, J.H. (1971) *Mastery Learning*. London: Holt Rinehart & Winston: .

Block, J.H. (1974) *Schools, Society and Mastery Learning*. London: Holt, Rinehart and Winston.

Block, J.H. and Burns, R.B. (1977) 'Mastery learning'. In L.S. Shulman (Ed) *Review of Research in Education*. Volume 4. Itasca, Ill.: F.E. Peacock, 3-49.

Bloom, B.S. (1964) *Stability and Change in Human Characteristics* . New York: John Wiley.

Bloom, B.S. (1968/81) 'Learning For Mastery', (UCLA-CSEIP) The Evaluation Comment, 1(2). In B.S. Bloom (Ed) All Our Children Learning. London: McGraw-Hill.

Bloom, B.S. (1976) Human Characteristics and School Learning. New York: McGraw-Hill.

Bloom, B.S. (1978) 'New views of the learner - implications for instruction and curriculum', Educational Leadership, 35, 562-76.

Bloom, B.S. et al. (1956) Taxonomy of Educational Objectives Handbook I: Cognitive Domain. New York: McKay.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. (1971) Handbook on Formative and Summative Evaluation. London: McGraw-Hill.

Born, D.G. (1971) Student Withdrawals in Personalized Instruction Courses. Paper presented at the meeting of the Rocky Mountain Psychological Association, Denver, May.

Born, D.G. and Herbert, E. (1971) 'A further study of personalized instruction in large university classes', Journal of Experimental Education, 40, 6-11.

Bruner, J.S. (1966) Toward a Theory of Instruction. Cambridge, Mass.: Harvard University Press.

Calhoun, J.F. (1976) 'The combination of elements in the personalized system of instruction', Teaching of Psychology, 3, 73-76.

Carroll, J. (1963) 'A model of school learning', Teachers College Record, 64, 723-733 .

Clark, R.E. (1983) 'Reconsidering the research on learning from media', Review of Educational Research, 53, 4, 445-459.

Dubin, R. and Taveggia, T.C. (1968) The Teaching Learning Paradox. University of Oregon Press: Eugene

Farmer, J., Lachter, G.D., Blaustein, J.L. and Brett, K. (1972) 'The role of proctoring in personalized instruction', Journal of Applied Behaviour Analysis, 5, 401-404.

Gagné, R.M. (1962a) 'The acquisition of knowledge', Psychological Review, 69(4), 355-365.

Gagné, R.M. (1965) The Conditions of Learning. (1st Edition) London: Holt, Rinehart and Winston.

Gagné, R.M., Mayor, J.R., Garstens, H.L. and Paradise, N.E. (1962) 'Factors in acquiring knowledge of a mathematical task', Psychological Monographs, 76(7), No.526.

Gagné, R.M. and Paradise, N.E. (1961) 'Abilities and learning sets in knowledge acquisition', Psychological Monographs, 75(14), Whole No. 518.

Gallup, H.F. (1970) Individual Instruction in an Introductory Psychology Course. Paper presented at the meeting of the American Psychological Association, Washington, D.C. (Sept.)

Gallup, H.F. (1974) Problems in the Implementation of a course in Personalized Instruction. In J.G. Sherman (Ed) PSI: 41 Germinal Papers. Menlo Pk., Cal.:Benjamin.

Glaser, R. (1976) 'Components of a psychology of instruction: towards a science of design', Review of Educational Research, 46(1), 1-24.

Green, B.A. (1971) 'Physics teaching by Keller Plan at MIT', American Journal of Physics, 39, 764-775.

Johnson, K.R. and Sulzer-Azaroff, B. (1974) The Effects of Different Proctoring Systems on Student Examination Performance and Preference. Paper presented at Second National Conference on Research and Technology in College and University Teaching. Georgia State University, Atlanta, GA. (Nov.)

Keller, F.S. (1968) 'Goodbye, Teacher', Journal of Applied Behaviour Analysis', 1, 79-87.

Keller, F.S. and Sherman, J.G. (1974) The Keller Plan Handbook. Menlo Park-California: Wa. Benjamin.

Kim, H. (1971) 'Mastery learning in the Korean middle schools', UNESCO Regional Office for Education in Asia, 6(1).

Knightly, J.J. and Sayre, J.L. (1972) 'Self-paced instruction for library science students', Journal of Education for Librarianship, 12, 193-197.

Kolb, J.R. (1967-68) 'Effects of relating mathematics to science instruction on the acquisition of quantitative science behaviours', Journal of Research in Science Teaching, 5, 174-182.

Kulik, J.A. and Jaksa, P. (1977) 'PSI and other educational technologies in college teaching', Educational Technology, 17, Sept. 12-19.

Kulik, J.A., Kulik, C.C. and Cohen, P.A. (1979) 'A meta-analysis of outcome studies of Keller's personalized system of instruction. American Psychologist, 34(4), 307-318.

Kulik, J.A., Kulik, C.C. and Smith, B. (1976) 'Research on the personalized system of instruction', Programmed Learning and Educational Technology, 13, 21-30.

Lee, Y.D. et al. (1971) Interaction Improvement Studies on the Mastery Learning Project - Final Report on Mastery Learning Programme. Educational Research Centre, Seoul National University. (Nov.)

Lewis, J.D. (1972) Manipulation of Two Variables in a Contingency-managed Statistics Class. Paper presented at the West Coast PSI Conference. California State University.

McKeachie, W.J. and Kulik, J.A. (1975) Effective college teaching. In F.N. Kerlinger (Ed) Review of Research in Education Vol.3 Itasca,Illinois: F.E. Peacock

McMichael, J. and Corey, J.R. (1969) 'Contingency management in an introductory psychology course produces better learning', Journal of Applied Behaviour Analysis, 2, 79-83.

Minke, K.A. and Carlson, J.G. (1973) A Comparison of Two Lecture Systems in Unit Mastery Instruction. Paper presented at the annual meeting of the American Psychological Association, Montreal. (Aug.)

Morrison, H.C. (1926) The Practice of Teaching in a Secondary School. University of Chicago Press: Chicago.

Myers, W.A. (1970) 'Operant learning principles applied to teaching introductory statistics', Journal of Applied Behaviour Analysis, 3, 213-220.

Nelson, T.F. and Bennet, M. (1973) 'Unit size and progress rates in self-paced instruction', Journal of College Science Teaching, 3(2).

Nelson, T.F. and Scott, D.W. (1972) 'Personalized instruction in educational psychology', Michigan Academician, 4, 293-302.

O'Neill, G.W., Johnston, J.M., Walters, W.M. and Rasheed, J.A. (1975) The Effects of Quantity of Assigned Material On College Student Academic Performance and Study Behaviour. In J.M. Johnston (Ed) Behaviour Research and Technology in Higher Education. Springfield, Illinois: Charles C. Thomas.

Phillips, T.W. and Semb, G. (1975) A Comparison of Pre-quiz and Post-quiz Lectures in a Contingency-managed University Course. Paper presented at the annual meeting of the American Association, Chicago. (Sept.)

Rowntree, D. (1977) Assessing Students: How Shall We Know Them? London: Harper and Row.

Ryan, D.W. and Schmidt, M. (1979) Mastery Learning: Theory, Research, and Implementation. Toronto, Ontario: The Ontario Institute for Studies in Education.

Semb, G. (1974) 'Personalized instruction: the effects of mastery criteria and assignment length on college student test performance', Journal of Applied Behaviour Analysis, 7, 61-69.

Sheppard, W.C. and MacDermot, H.G. (1970) 'Design and evaluation of a programmed course in introductory psychology', Journal of Applied Behavioural Analysis, 3, 5-11,

Skinner, B.F. (1954) 'The science of learning and the art of teaching', Harvard Educational Review, 24(2), 86-97.

Tyler, R.W. (1932) 'The construction of examinations in botany and zoology', Service Studies in Higher Education, Ohio State University, Bureau of Educational Research Monographs, 15, 49-50.

Washburne, C.W. (1922) 'Educational measurements as a key to individualizing instruction and promotions', Journal of Educational Research, 5, 195-206.

White, R.T. (1973) 'Research into learning hierarchies', Review of Educational Research, 43(3), 361-375.

CHAPTER 5

HUMAN INFORMATION PROCESSING AND THE AUDIO-VISUAL APPROACH TO EDUCATIONAL TECHNOLOGY

We receive information from a vast array of sources, some experienced directly and physically as we come into contact with our immediate environment, others experienced vicariously through media such as films and television, or through symbolic modes such as words or figures. Some of these are perceived through all our senses, while others are perceived through a more limited number of senses. Treichler (1967) presents a number of interesting conclusions, which may accord with many commonly held beliefs concerning the relationship between our senses and how we learn and remember (Figure 5-1), although they should be treated with caution because he gives no indication of their basis.

Figure 5-1. The Relative Effectiveness of the Primary Senses

WE LEARN:

    1.0% THROUGH TASTE
    1.5% THROUGH TOUCH
    3.5% THROUGH SMELL
    11.0% THROUGH HEARING
    83.0% THROUGH SIGHT

PEOPLE GENERALLY REMEMBER:

    10% OF WHAT THEY READ
    20% OF WHAT THEY HEAR
    30% OF WHAT THEY SEE
    50% OF WHAT THEY SEE AND HEAR
    70% OF WHAT THEY SAY AS THEY TALK
    90% OF WHAT THEY SAY AS THEY DO A THING!

Treichler's views have been reinforced by the theoretical orientations of a number of influential writers concerned with the audio-visual movement. For example, Dale (1954) listed the following proven contributions of such materials:

1. They supply a concrete basis for conceptual thinking

and hence reduce meaningless word-responses of students.
2. They have a high degree of interest for students.
3. They make learning more permanent.
4. They offer a reality of experience which stimulates self-activity on the part of pupils.
5. They develop a continuity of thought; this is especially true of motion pictures.
6. They contribute to growth of meaning and hence to vocabulary development.
7. They provide experiences not easily obtained through other materials and contribute to the efficiency, depth, and variety of learning.

Figure 5-2. Dale's 'Cone of Experience'
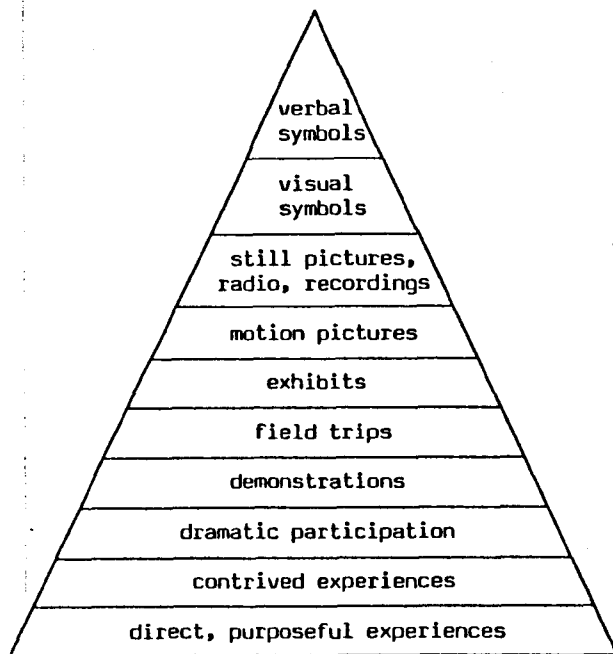
He concluded by asserting that these points do not constitute private beliefs or opinions, but are the 'distillation of a vast amount of research by many investigators.' Dale's 'Cone of Experience' (Figure 5-2) illustrates the narrowing of sensory experience from the concrete, purposeful experiences of direct, contingent experience to the higher reaches of abstract verbal and visual symbols. It reflects his original theoretical position that the predominant modes of instruction, found towards the pinnacle of the cone, provide restricted experiences for the pupil, who would greatly benefit from a more comprehensive range of educational experiences, encompassing those modes towards the base of the cone.

Early support for Dale's approach came from Carpenter (1953) in the form of the 'sign similarity hypothesis' which stated that:

> ...films whose signals, signs, and symbols have high degrees of similarity ('iconicity') to the objects and situation which they represent will be more effective for most instructional purposes than films whose signals, signs, and symbols have low degrees of 'iconicity'. (Carpenter, 1953, p. 41)

The term 'iconicity' was borrowed from Morris (1946), who suggested that signs vary in 'iconicity' to the degree to which they are similar to things or situations which they represent or signify. For example, a motion picture of a specific event may be described as highly iconic, whereas a pencil sketch may be described as being of low iconicity.

Such theories and other similar orientations (Knowlton, 1966; Gibson, 1954) have been given the collective term 'realism' theories by Dwyer (1978). In essence they suggest that the additional information flowing through several different sensory channels (ie. haptic, visual, auditory, etc.) will provide a multiplicity of stimuli which in turn will assist learners in organizing and structuring their perceptions, thereby ensuring more complete learning.

This approach has been preoccupied with the effects of the new media which were seen as being implicitly good because they were innovative devices that could supplant the verbal functions traditionally fulfilled by the teacher and text. Jonassen (1985) suggests that this materialistic conception of learning dominated the field in the early part of the century, yielding to the emergence of behaviourism towards the middle part of the century. In fact, it did not entirely give way and remained a powerful influence for several decades. Indeed, recent researches from cognitive psychology suggest that some, though by no means all, of the early claims may still be valid.

The major challenge to this physical science approach

to educational technology came from the emerging cognitive psychology of the 1950's and 1960's. This was epitomized by Travers (1964) who felt that the work of psychologists and researchers was at variance with the claims made by the proponents of the audio-visual movement. Travers concluded that:

> ...there does not seem to be a single contemporary scientist who takes the position that the human can receive more information if exposed to information transmitted through two sense modalities rather than one. The position of the research scientists and the designers of audio-visual materials are at such opposite poles that it hardly seems possible that both can be correct. (Travers, 1964, p.375)

Travers supported the idea that the human information processing system is best represented as a single channel of limited capacity and that increasing the available information to the human receiver, either by using several sensory modalities or using very realistic displays, will merely produce information overload, which in extreme cases will actually cause decreases in learning. His theoretical base was rooted in the work of many psychologists interested in the problems associated with human perception and communication, most notably that of Broadbent (1958).

This issue is a complex one. It has profound implications for education. On the one hand there are those who hold to the belief that direct, contingent experience, with multi-sensory inputs, is essential for intellectual development; and, set against these beliefs, are those which suggest that only a fraction of the total sensory information available at a given time is capable of being processed by the central nervous system and, therefore, that educators should aim to compress reality into a form most compatible with the central processing systems.

PSYCHOLOGICAL APPLICATIONS OF INFORMATION THEORY

Information theory was developed initially for communication systems such as the telephone or radio, the best known of the early models being that of Shannon and Weaver (1949), shown in Figure 5-3. According to Attneave (1959):

> The idea that information is measurable in precise terms was not widely appreciated until 1948, when Norbert Weiner's book "Cybernetics" appeared and Claude Shannon published a pair of articles titled "The Mathematical Theory of Communication"....The quick spread of

Shannon's ideas beyond the field of engineering may be attributed largely to a skillful and imaginative introductory article by Warren Weaver which appeared in the July 1949 issue of Scientific American.
(Forward to "Applications of Information Theory to Psychology").

In this system the message from the information source is converted into a signal by the transmitter, for example, a sentence spoken into a microphone which is broadcast by a radio transmitter. At some point this signal is contaminated by 'noise' (an electrical engineering term, denoting some form of interference), before being decoded by the receiver, eg. a radio receiver, and reconstituted as the message. As more 'noise' enters the system, so the intelligibility and

Figure 5-3. Shannon and Weaver's Linear Model of Communication (Shannon and Weaver, 1949)

INFORMATION SOURCE

↓

message

↓

TRANSMITTER

↓

signal

↓

NOISE SOURCE ────→

↓

received signal

↓

RECEIVER

↓

message

↓

DESTINATION
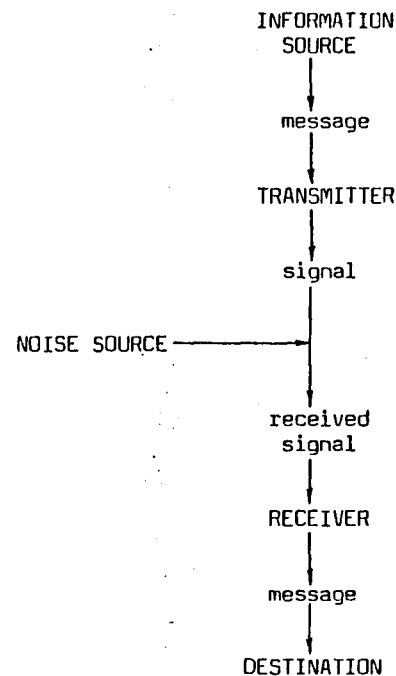
integrity of the original message changes, leading to new signals and information.

Shannon and Weaver viewed communication in a very broad sense to include all the procedures by which one mind may affect another: written and oral speech; music; pictorial arts; theatre; ballet; indeed, all human behaviour. Shannon's mathematical theory applied in the first instance only to the technical problem of accuracy of transference of various types of signals from sender to receiver. However, Weaver claimed that the theory was 'exceedingly general in its scope, fundamental in the problems it treats, and of classic simplicity and power in the results it reaches.'

The technical meaning of information is not radically different from every day usage, it is essentially 'choice', the narrowing down of alternatives, and the unit most often used for measurement is the 'bit', or binary digit. The word 'information' in communication theory relates not 'so much to what you do say, as to what you could say. The amount of information is defined in such a way that the nature of the information is irrelevant, just as weight is defined in such a way that the nature of the object is irrelevant: and the definition uses binary choice as its basis. In effect, the definition states that the information contained in a given situation, where a number of equally likely events could take place, is equal to the number of binary choices required to identify a specific event.

For example, imagine two boxes side by side, one of which contains an object (any object), with each box as likely to contain the object as the other. This situation represents one bit of information because it requires one binary choice to identify the box containing the object. No matter which box you choose to open you will know where the object is, and all uncertainty is removed. If you open the box and it is empty, you know unequivocally where the object is - in the other box! Of course, you also know where it is if you make the other choice. The point is, you only have to choose once to remove all uncertainty.

Thus, information is essentially choice, the narrowing down of alternatives. The mathematical theory of communication was developed in order to deal quantitatively with the transmission of messages over 'channels', to measure the amount of information transmitted, but Weaver indicated that the theory should not be restricted to engineering situations. It was to be seen as general in its scope.

The theory opened up a whole new avenue in psychology and some of Shannon's general concepts were rapidly introduced into the psychological literature by Miller and Frick in 1949. This led to a mushrooming of interest in information theory and its implications for the study of perception and communication. It was realized that, once

received by the sensory apparatus, signals are then passed through the nervous system, whose channels are just as susceptible to capacity limits, equivocation and the degrading effects of 'noise', as are those of the telephone and radio.

Information Capacities of the Eye, Ear and Central Nervous System

The initial application of information theory to radio-communications dealt with measures of information transmission rates and the capacities of channels of communication, and this was reflected in the field of physiological psychology. Jacobsen (1950, 1951) compared the relative capacities of the eye and ear, and concluded that there was a factor of about 30 in favour of the eye on the basis of nerve fibres leading from the organs. However, when consideration was given to the efficiency with which an individual nerve fibre transmits information, Jacobsen found that the 900,000 optic nerve fibres each produced 5 bits per second, compared with a maximum of 0.33 bits per second for each of the 30,000 auditory fibres, giving a factor of nearly 500 in favour of the eye.

These results indicated a far higher rate of information transmission for the eye when compared with the ear, at least as far as the optic and auditory nerves are concerned. However, further research indicated that there was a restriction on the amount of information that could be transmitted through the central nervous system, and that central information processing would be subjected to an upper limit. The existence of such a limit implies that any information beyond the limit is lost; in information theory this loss is called equivocation.

The earliest experiments indicated that this capacity of the central nervous system was rather low. Hake and Garner (1951) required subjects to make judgements concerning the position of a pointer between two markers and found that the channel capacity for judgements of visual position was approximately 3.25 bits. In other words, they could accurately and consistently assign up to about 9 alternative positions on the scale , but as the number of positions they were required to estimate increased beyond 9 the subjects became less efficient. Similar results were obtained by Pollack (1952, 1953) for absolute judgements of auditory pitch. Subjects were asked to listen to a tone and assign a numeral to it. When only two or three tones were used the listeners never confused them, with four tones confusions were quite rare, but with five or more tones confusions were frequent. The upper limit was estimated to be 2.5 bits (6 alternatives).

It was soon realized that for one-dimensional judgements we rarely confuse 7 or so alternatives, and this seems to be a general limit on our capacity for such judgements. With multi-dimensional judgements, with a number of independently variable attributes, the capacity would be expected to increase. The position of a dot inside a square is clearly a two dimensional proposition. If we have a capacity of 3.25 bits for estimating intervals, as the Hake and Garner experiment indicated, and if we do this for two dimensions we should get 6.5 bits as our capacity for locating points in a square. In fact, the limit appears to be 4.6 bits or 25 alternative positions (Klemmer and Frick, 1953), rather than the 90 predicted from the uni-dimensional judgements. Pollack (1953) anticipated that for listeners asked to judge both the loudness (2.3 bits) and pitch (2.5 bits) there would be an augmenting effect so that the multi-dimensional judgement would raise the capacity to 4.8 bits. The limit is raised, but only to 3.1 bits (9 alternatives rather than the 28 anticipated). In fact, the largest channel capacity in these early studies was found to be 7.2 bits (150 alternatives) when Pollack and Ficks (1954) used 6 different acoustic variables in their multi-dimensional display.

These results led Travers (1967) to state that:

....as more dimensions are added there is a decreasing increment in the channel capacity of the multi-dimensional system. People are less accurate in any one dimension if they must judge more than one attribute at a time.

The data suggest that the amount of information which can be taken in on a single presentation of a display is probably quite limited - perhaps of the order of 10 bits.

Further, the general conclusion was that the channel capacity is ordinarily less than that estimated by Jacobsen (1951), perhaps as little as a tenth of the figures given, and that the information used is probably less than one tenth of one percent of the channel capacity of the sensory input.

Human Information Processing: The Single-channel Model

At this time there was good evidence to suppose that the information available from the senses far exceeds the limits of the central processing system. In addition, evidence was emerging from a series of experiments which indicated that sensory information fed into a central system for processing information, which was effectively a single channel system, in which information could be processed sequentially from only one source at a time.

The question was, as there clearly is a limit to the capacity of the central system, how does the system select which information to pay attention to and which to ignore? The formulation of the first complete theory of human information processing and selective attention by Broadbent (1958) attempted to provide an answer to this. The theory had obvious implications for anyone involved in producing teaching materials, particularly if those materials use a combination of senses, as is the case with many audio-visual aids. Broadbent's model was later developed by Travers (1964) who placed the research scientists and audio-visual designers at such opposite poles in their ideas concerning human information processing capacities.

Broadbent's original theory evolved from a series of experiments and techniques developed in the early 1950's. As with information theory, the main impetus for interest in this area was the tremendous rate of development of communication systems during the second world war. Colin Cherry's work in this field demonstrates the international and inter-disciplinary nature of this branch of psychology. Although he was an Englishman his major contribution was published as a result of research conducted at the Massachusetts Institute of Technology, and it appeared in the 'Journal of the Acoustical Society of America.' It was called 'Some experiments on recognition of speech with one and two ears' (1953) and used the technique known as 'shadowing'. Shadowing consists of playing a message into a set of headphones and asking the subject to repeat the words of the message, keeping pace with the presentation as far as possible. It is, in fact, a lot easier than it may seem, though the spoken words are slightly delayed behind those to which the subject is attending, and the voice tends to have a pronounced monotony. If the material being shadowed is of a complex nature the subject will invariably have very little idea of what it is about, even though every word has been recognized and repeated.

Cherry was concerned with what he chose to call the 'cocktail party effect', which also has the more scientific title: selective attention. He wanted to know how we were able to select that one voice to which we are attending from the general babble of the party. He was also interested in the retention of information from conversations to which we do not pay attention. In the experiments one continuous spoken message was fed into a headphone on the subject's left ear and a different message was applied to the right ear. The subject was then asked to repeat one of the messages concurrently while listening, and to make no mistakes. The really fascinating thing about the experiment concerns the nature of the subject's perception of the unattended message at the unshadowed ear. The subject could say little about it, apart from acknowledging that sounds

occurred. In further experiments Cherry attempted to find out just what attributes of the rejected message were recognized.

In one experiment, for example, both messages started in English, and when the subject was comfortably shadowing the attended message, the unattended message was changed from English to German. When questioned later the subject was unaware of the change.

In another experiment the unattended message was composed of one of the following: normal English spoken by a male; English spoken by a female; English played in reverse; and, finally, a 400 cycle per second tone. With all the conditions of human speech, including the reversed speech, all subjects reported it as such, but could not identify specific words or even the language as English. Physical characteristics, such as a change to female voice or the introduction of the tone, were always noticed.

Cherry concluded from these experiments that detailed aspects such as language, individual words or semantic content are not noticed in the unattended channel. In effect, he felt that the utilization of information by the higher centres of the nervous system can be represented best by a single channel system. Broadbent extended this idea that the perceptual system generally functions as a single channel with information from only one source gaining access to it at a given time. But he cautioned against saying simply that 'a man cannot listen to two things at once'. On the contrary, he receives some information even from the unattended ear: but there is a limit to that amount of information. Broadbent (1954) had found that subjects could actually recall details of information presented to the unattended ear, under the specific conditions known as 'dichotic presentation'. Subjects were presented with three pairs of simultaneous digits, one of each pair to each ear (Figure 5-4), and were asked to recall as many digits as they could. Under normal circumstances subjects would have had little difficulty in remembering 7 or even 10 digits,

Figure 5-4. Dichotic listening experiment (Broadbent, 1954)

5	3	1	RIGHT EAR
6	4	2	LEFT EAR

Anticipated recall: 1,2,3,4,5,6 or 2,1,4,3,6,5

Actual recall   : 1,3,5,2,4,6 or 2,4,6,1,3,5

Figure 5-5. Simplified version of Broadbent's Model of Information Flow (1958).
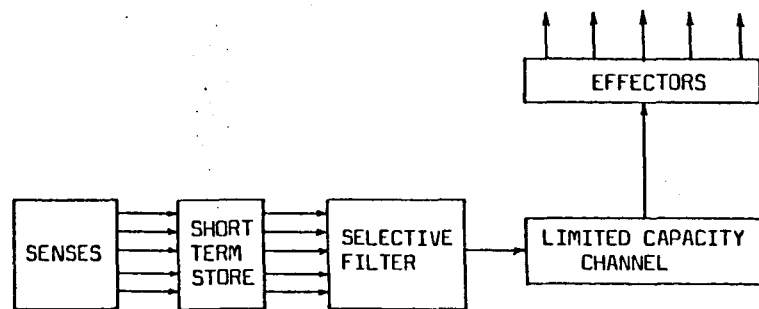
but under dichotic presentation they could recall only 4, 5 or 6.

The most interesting observation was that subjects preferred to report the digits by ear rather than in the order in which they were heard. In Figure 5-4, digits 1 and 2 are presented simultaneously, both being presented before the second set, digits 3 and 4. If subjects reported digits in temporal order both 1 and 2 would be recalled before either 3 or 4, giving two alternative temporal report patterns (anticipated). However, subjects chose to report the digits first by one ear and then by the other. When six pairs of digits were given, so that subjects had to recall 6 digits from one ear, before switching to the other, subjects were unable to recall the second set. Broadbent concluded that there must be a short-term memory store for information lying between the two stages of recognition of auditory messages. The experiment was repeated with half the digits presented to the ears and half to the eyes and the same result was obtained, showing that the buffer storage system was a general feature of human perception (Broadbent, 1956). This short-term memory store holds items which have not yet received attention and it may be thought of as keeping a record of raw sensory data. These results, and those of many other psychologists, physiologists and communications engineers, led Broadbent to the tentative formulation of his model of information flow in the human organism (Figure 5-5).

According to the Broadbent model, information passes from the senses and is held for a few seconds in the short term memory store where it will rapidly fade unless the selective filter mechanism permits access to the single limited channel, thence to the higher centres concerned with long term memory and control of behaviour.

The short-term store which feeds information into the processing channel, via the filter mechanism, was demonstrated to have certain limitations, in that only a relatively small number of items could be held in the memory at a given time, and that attention to new inputs causes old activations to fade away. Even short lists will completely fade in as little as 18 seconds if the subject is prevented from repeating the list to himself (a process known as 'rehearsal').

Beating the Channel Capacity: "Chunking" Information

The limit of the storage system is 6-10 units of information, and these units were called 'chunks' by Miller (1956) to differentiate them from 'bits', because the amount of information retained is not fixed. If the storage of items was like absolute judgement, then there should be a fixed amount of information that can be retained. We should remember less items when they contain a lot of information

Table 5-1. Ways of Recoding Sequences of Binary Digits.

Binary Digits (bits)   1 0 1 0 0 0 1 0 0 1 1 1 0 0 1 1 1 0

2:1 Chunks	10	10	00	10	01	11	00	11	10
Recoding	2	2	0	2	1	3	0	3	2

3:1 Chunks	101	000	100	111	001	110
Recoding	5	0	4	7	1	6

4:1 Chunks	1010	0010	0111	0011	10
Recoding	10	2	7	3	2

5:1 Chunks	10100	01001	11001	110
Recoding	20	9	25	6

Figure 5-6. Effects of Recoding Binary Digits on Short-Term Memory Performance.
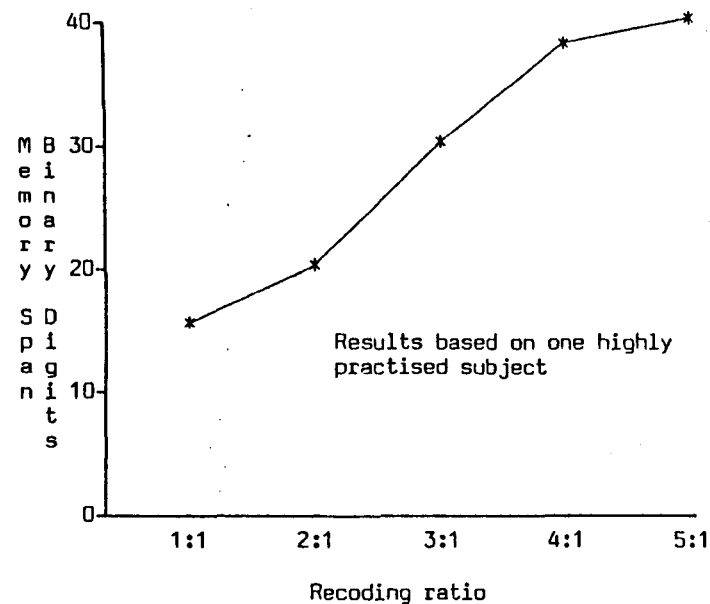
and remember more when they contain little information. As Miller pointed out:

> For example, decimal digits are worth 3.3 bits apiece. We can recall about 7 of them, for a total of 23 bits of information. Isolated English words are worth about 10 bits apiece. If the total amount of information is to remain constant at 23 bits, then we should be able to remember only 2 or 3 words chosen at random. (Miller, 1956, p.91)

Clearly, this is not the case, we can recall substantially the same number of items for a range of materials: digits, letters, words. In fact, by a process of coding and recoding information into larger chunks to increase the total amount of information, Miller demonstrated that the informational bottleneck, caused by the severe limitations placed on the amount of information

that we are able to receive, process and remember, could be stretched. Strings of binary digits were recoded by experimental subjects who were trained to code the strings into increasingly more complex chunks (Table 5-1). This meant that a string of 20 binary digits could be reduced to just four chunks of information using re-coding techniques. This, in turn, permitted the amount of information held in short-term storage to be increased dramatically (Figure 5-6). Miller concluded from this data that 'The kind of linguistic recoding that people do seems to me to be the very lifeblood of the thought processes.' In other words, given the inherent limitations of our information processing system, we extend our capacities by recoding information in a more economical form, from which we can reconstitute the original message when required, much as we do with dried vegetables when shelf storage space is limited.

Human Information Processing: Alternative Models

Miller had demonstrated one way in which the basic limitations of the single-channel information processing system could be overcome and within a short time further research began to challenge a central concept of the Broadbent model, that unattended information could only be analyzed at the level of gross physical characteristics. For example, Moray (1959) found that although his subjects could not recall nor recognize irrelevant messages presented as many as 35 times to the unattended ear, they frequently heard their own name when presented in this channel.

Figure 5-7. Channel Switching.

Attended Ear   : sitting at the mahogany / three
Unattended Ear: let us look at these / table

Further, Treisman (1960) demonstrated in an elegant fashion that subjects would follow the sense of a passage rather than the channel being shadowed, which indicated that processing was to some extent semantic. Figure 5-7 demonstrates how two sentences are switched in the shadowing experiment. Treisman found that subjects frequently shifted attention to the unattended ear to complete the sense of the sentence, so that it read 'sitting at the mahogany table', rather than 'sitting at the mahogany three'.

The conclusion from these experiments was that some, but not all, of the information from all the channels is

Figure 5-8. Treisman and Geffen's Attenuator Model of Attention.
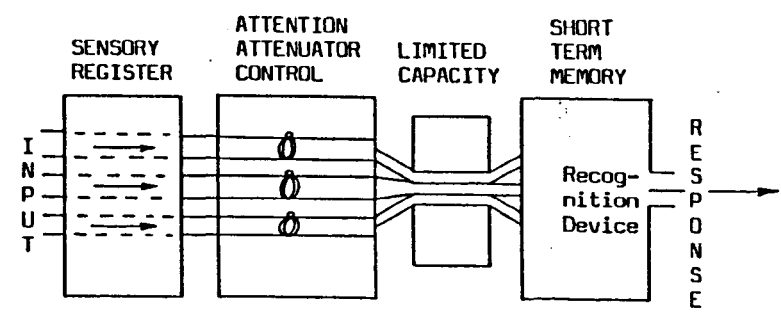
processed and this led to the development of Treisman and Geffen's (1967) 'attenuator' model of attention. Figure 5-8 shows how the model permits various sources to use the channel simultaneously. However, since there is a limited capacity, inputs from some channels must be attenuated.

This model has one primary channel that is being attended to, while all other channels are passing information in reduced amounts, but in such a way that whenever important information enters one of these attenuated channels it can be responded to. Treisman and Geffen (1967) obtained results strongly favouring this 'attenuation' view. Their subjects shadowed one of two dichotically presented prose passages, while simultaneously they had to monitor both channels for the occurrence of a target word, tapping each time it was detected. The results showed that whereas 90% of target words were identified in the primary channel, only about 9% were responded to in the unattended channel.

Norman (1968) extended this model of selection and attention to include a 'pertinence' factor (Figure 5-9). Pertinence, in the Norman model, is based on the expectations of future inputs and the properties of the currently attended channel of information. Certain classes of inputs, such as the sound of one's name, can be expected to have a permanently high level of pertinence, others will have levels which fluctuate with the expectations and analyses of ongoing events. The items in storage were termed 'logogens' by Morton (1969) in his complementary theoretical position. The logogens are seen as having thresholds, with the output for further processing depending on the level of activation from both sensory input processing and pertinence. Norman's model incorporates suggestions that for

Figure 5-9. Norman's Model of Selection and Attention.
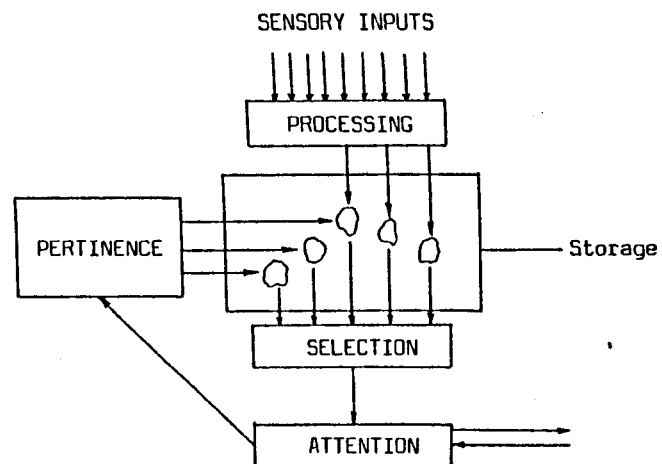
proper assessment all input information must have access to storage (Deutsch and Deutsch, 1963), and that an adjustable threshold pre-sensitizes the organism to sensory information of high relevance (Treisman and Geffen, 1967).

The item selected for further processing by the attention mechanism in the Norman model corresponds to the stored representation which received the greatest combination of pertinence and sensory activation. This was developed in a further paper by Norman and Bobrow (1975) who suggested that varying allocations of attention can be made up to the capacity limit of the processing system. An important distinction was made between resource limitations and data limitations. Resource limitations depend on the available internal resources of the human processor; data limitations relate to the sources of data forming the incoming signals. The suggestion was that some tasks require less than maximum processing resources and that if two tasks together required less than the maximum available they could both be performed simultaneously without any decrement in performance. However, if more than the maximum processing resources are required effort will be directed in favour of one or the other, resulting in a deterioration in the performance. The concept of data limitation introduces the constraints imposed by the external stimulus upon the processing system. With some tasks, applying more resources to processing may not result in an increase in performance.

The result is that the task has reached the point of pure data limitation. In such a case performance can only be improved by adjusting the stimulus data.

The Norman-Bobrow account resolved some anomalies in the data concerning information available in unattended channels and it is now generally accepted that information in these channels receives a level of processing higher than the crude physical characteristics suggested by the early Broadbent model. The processing is probably as high as semantic processing, and allows for channel switching as important information is registered in the attenuated channel. Cherry and Taylor (1954) demonstrated that this switching time was in the region of one sixth of a second or about 160-170 msecs, and Broadbent (1956) obtained a similar figure of 200 msecs in a later experiment.

The overall picture which emerges from these and similar studies of attention and information processing is one in which the major sense organs show great differences in their capacities for transmitting information, the eye being endowed with a greater number of nerve fibres, each of which is capable of transmitting more information than the corresponding auditory fibres. However, the transmitted information from the senses enters the central processing system which has a limited capacity. This limited capacity for information processing causes attention to be directed mainly towards one of several competing channels, the remaining channels being attenuated or receiving less than optimal processing but still receiving some high-level processing. Attention can be switched rapidly as important information is registered in the attenuated channels.

Depth of Processing and Mental Effort

Implicit in most of the early models of information processing was the division of memory into two parts: short-term memory (STM) and long-term memory (LTM), each with different characteristics but with both being essentially verbal storage systems. The distinction between STM and LTM has more recently given way to the view that memory should be viewed as an integral part of the whole information processing system, and that STM and LTM are merely artifacts of different 'depths of processing' (Craik and Lockhart, 1972). The different levels of processing range from shallow or physical analysis of a stimulus, to deep or semantic analysis. The depth of processing is defined in terms of the meaningfulness extracted from the stimulus rather than the number of analyses performed upon it. Information which is processed at a greater depth, which will be more meaningful than shallow processed information, will also be more memorable.

A typical experiment designed to demonstrate such an effect consists of having subjects perform different orienting tasks with a list of words before being given a recall test. Some tasks, such as detecting the occurrence of specific letters in the words or deciding the part of speech appropriate to each word, do not involve deep processing, whereas others, such as judging the pleasantness or frequency of usage, involve semantic processing at a deep level. Recall scores have generally tended to be higher for the deep processing list, although there is some circularity in the argument as to what constitutes a deep processing activity (Eysenck, 1984).

A recent adjunct to depth of processing is the elaboration of processing, with deep encodings being more elaborate or extensive and, therefore, providing more stored information than shallow processes. Retrieval, being affected by the amount of stored information, will be better for deep, elaborated encodings. Craik and Tulving (1975) ask 'Is spread of encoding a more satisfactory metaphor than depth?' and conclude that elaboration depends on the breadth of analysis carried out in each domain, and that a minimal semantic (deep) analysis will be more beneficial for memory than an elaborated structural (shallow) analysis.

Eysenck (1984) has suggested that a more complete theory than the original levels of processing proposed by Craik and Lockhart (1972) should take account of the fact that learning and memory depend upon at least four major factors: (1) the nature of the task (2) the nature of the stimulus materials (3) the characteristics of the subjects and (4) the nature of the retention test used.

Salomon (1984) has attempted to incorporate some of these factors into his application of the levels of processing theory in media research. He observed that while many U.S. experts claimed that television somehow inhibits 'deeper' processing of material and thus cannot serve well as an instructional medium, his cross-cultural study of Israeli and American children demonstrated that Israeli children retain more information and are more strongly affected by its formal features than their American peers. The Israeli children seemed to take the medium more seriously, processing the material with greater depth. He reasoned that above and beyond the cognitive activities a source of information activates or inhibits in learners (see Chapter 7), the depth with which its information is processed may depend on the way in which it is perceived. Pre-existing perceptions, particularly if socially reinforced, may preclude deep processing of available information.

In his experimental study he found that children thought of themselves as being more able to learn from tv than from books. When presented with similar material to learn from tv

or print, the tv group reported investment of significantly less mental effort than the book group. This self-reported amount of invested mental effort (AIME) was reasoned to be related to 'non-automatic, effortful mental elaborations' which are at the base of both depth of processing and Langer's (1984) idea of mindfulness.

With more tv group members reporting the tv programme to be 'easy stuff' and more book children reporting the text to be 'difficult', it should follow that the tv group should perform better than the text group on the test measuring inference making and factual recognition. This was not the case. It was the text group that achieved significantly higher scores. Salomon found that for the book group those who felt more 'efficacious' tended to invest more mental effort in reading the text, and those who invested more mental effort learned more. A different pattern emerged for the tv group. Those who felt more efficacious tended to invest less mental effort in the programme and thus learned less from it. Measures of the perceived realism of the different media showed tv to be rated higher and this correlated with the perception that the message was easier and therefore demanded less effort, a worrying result for advocates of realism theories! These results clearly indicate the complexity of the inter-relationships between those factors suggested by Eysenck (1984) as impinging on applications of depth of processing and elaboration paradigms.

The Dual-coding Hypothesis

Although the depth of processing approach has attempted to resolve the STM/LTM division, a new division has emerged based on two different but complementary systems of storage: a semantic/verbal logogen system and an image-based 'imagen' system. This division is necessary because of increasing evidence in support of a separate image-based system of information storage.

Many researchers have found that pictures are superior to words in a variety of tasks. In fact, people have an extraordinary recognition memory for pictures, and one of the most dramatic demonstrations was reported by Shepard (1967). Subjects viewed 600 pictures and were tested for recognition by pairing one of these pictures together with a new picture, not previously seen by the subject, who was required to identify the picture from the original set. Words were also presented in a similar manner: one set made up from fairly common words (100 times or more per million according to Thorndike and Lorge, 1944); the other from 'rare' words (less than once per million). Finally, a third group had sentences presented in the same way. The medians

of the percentage correct were 98.5, 90.0 and 88.2 for pictures, words and sentences. The results also indicated that subjects were better able to recognize words which were rare (92.5 percent correct) than when they were frequent (84.4 percent correct). Similar results have been found by Gorman (1961) and Olver (1965).   Further results showing the extent of visual recognition memory were obtained by Standing, Conezio and Haber (1970), who presented 2,560 photographs for 10 seconds each over a period of 2 to 4 days, with recognition being maintained at 90.5 percent for as long as 3 days. When recognition time was reduced to one second per picture and the orientation of the picture in the test was altered, performance remained largely unimpaired. It has been suggested that the superiority of pictures in recognition tests may be due to differential verbal responses rather than 'picturability'. However, much of the evidence indicates a higher recognition memory for pictures of familiar objects than their concrete-noun labels (Paivio, 1971, p.183), indeed, recognition memory appears to be a direct function of stimulus concreteness: recognition increases from abstract words, to concrete words, to pictures.

Paivio (1971) has concluded that these results suggest both verbal and non-verbal symbolic processes. The superiority of concrete over abstract words suggests that referent images, associatively evoked by the concrete nouns, also facilitate recognition memory. Paivio also found that, within the limitations imposed by task characteristics, there is evidence that non-verbal processes can also play a crucial mediational role in free recall. 'Picturability' has been shown to be consistently positive since the earliest experiments (Kirkpatrick, 1894; Calkins, 1898) when average

Figure 5-10. Availability of two coding systems as a function of stimulus concreteness; degree of availability is indicated by the number of plus (+) signs.

STIMULUS	CODING SYSTEM	
	Imagery	Verbal
Picture	+++	++
Concrete word	+	+++
Abstract word	-	+++

recall was highest for objects and pictures when compared with words. Substantially the same results have been obtained in more recent time (Paivio, Rogers and Smythe, 1968; Sampson, 1970).

These results are of great interest because the required response in recall tasks is verbal in both cases (for both pictures and their verbal labels) and it may be anticipated that the to-be-remembered name should be more readily available when the stimulus is a word rather than a picture. Since it is not, verbal processes alone cannot account for the finding. On the basis of such experimental results Paivio has suggested the dual-coding model for pictures and words (Figure 5-10), where the availability of an item is indicated by its total representation in the dual (imagery/verbal) coding system.

Figure 5-11. Schematization of Paivio's Dual-coding Model (Bleasdale, 1983).
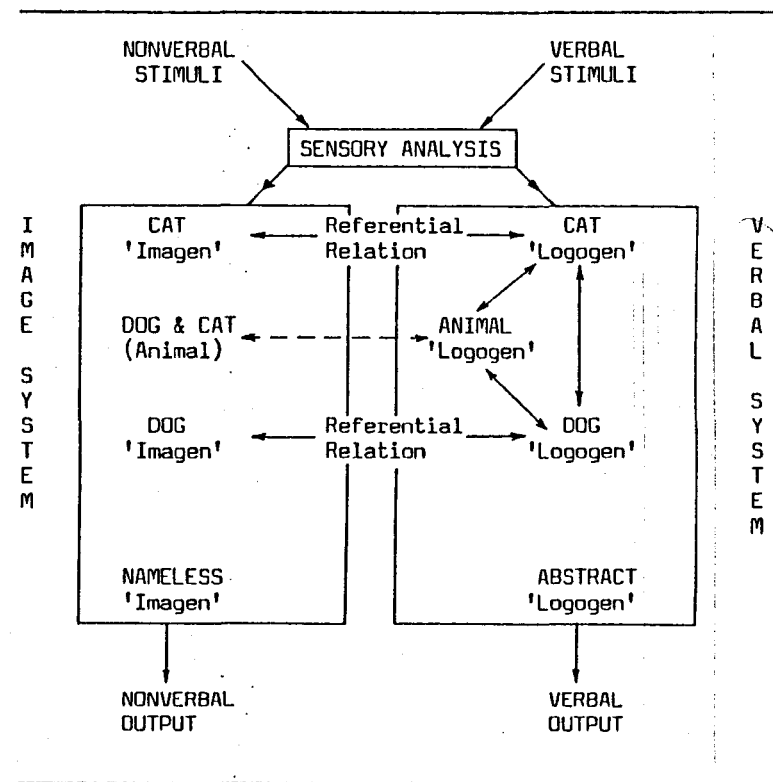
These results certainly indicate the presence of both a non-verbal (imagen) and a verbal (logogen) storage system. Indeed, Bleasdale (1983) claims that it would seem logical from a phylogenetic and ontogenetic perspective to assume that linguistic ability is preceded by the acquisition of perceptually-based information. Piaget (1953) suggested that knowledge structures (schemata) start to be formed almost from birth, and are basically formed from the co-ordinated remnants of sensory and motor signals that have been generated as the child has interacted with the world. This view has been incorporated by Paivio (1971), who also assumes that linguistic competence and performance are based on a substrate of imagery, although verbal behaviour and understanding ultimately becomes free of dependence not only upon a concrete situational context but to some extent from the imagery as well.

This concept of independent verbal and ·imaginal representational systems is at the heart of Paivio's dual-coding model (Figure 5-11). According to this model, words and pictures receive an initial, relatively peripheral, sensory analysis and are usually recognized by one of the stored representations (imagen/logogen), permitting access to further referential processing, during which time certain words may be imaged (especially concrete nouns) and pictures (or line drawings or objects) may be named. Finally, associative relations may be formed intraverbally among logogens and intra-imaginally among imagens.

Other models with similar dual-representational systems (Seymor, 1973; Morton, 1980), have also been developed to account for the accumulating evidence, coming from a wide variety of sources, for a separate image-related system.

WORDS, PICTURES AND INSTRUCTIONAL MEDIA

The pedagogical interest in the major channels of communication, their relative capacities, the limits placed on the processing of information, and the types of storage systems for processed information stems from a realization that the effectiveness of media generally, and instructional media in particular, will depend on the different properties of the channels used. Hsia (1968a) summarized these differences for the eye and ear and concluded that unless the fundamental information processing properties of the modalities are known then comparisons between different media using one or both senses will be meaningless (Table 5-1) and Travers (1967) agreed with this, clearly identifying those instructional tasks for which one or the other of the main channels is an obvious choice:

One cannot reasonably ask the general question whether

the eye or the ear is more efficient for the transmission of information, for clearly some information is more efficiently transmitted by one sensory channel than by another. If a person needed to know about the appearance or physical characteristics of an unfamiliar animal he would do much better to examine the animal visually than to have a description related by another observer. Such visual observation would be best until he asked about the vocal noises emitted by the animal, in which case he would do better to hear the actual sound of the animal than to obtain visual information about the vocal mechanism. Where direct information is needed about the world, the sensory modality should be used which can best transmit directly the particular physical Information involved. (Travers, 1967, p.88)

Table 5-1: Channel Effectiveness (Hsia,H.J., 1968a).

AUDITORY CHANNEL	VISUAL CHANNEL
Temporal in nature	Spatial in nature, excepting tv and motion pictures
Sequential presentation	Both sequential and simultaneous presentation
Poor referability	Good referability: information can be stored in the display.
Fewer dimensions in information coding	Numerous dimensions in information coding
Greater flexibility; variation in connotations, nuances and inflection	Restricted flexibility and advanced coding
Rate of transmission limited to speaking rate	Much faster rate of transmission
Less versatility	Greater versatility
More attention demanding	Less attention demanding
More resistant to fatigue	Less resistant to fatigue

Words: Print and Speech

However, under certain circumstances it is quite legitimate to compare the efficiency of the two main sensory modalities because both may seem appropriate. For example, identical material may be presented to both eye and ear in the form of print or speech. Based on the knowledge gained from cognitive psychology it can be postulated that the eye with its superior capacity will be more effective, although the limiting factor in the form of the central processing capacity may diminish the eye's overall superiority. Also, auditory information may have a greater tendency to become data limited because of the lower transmission rate of the peripheral nervous pathways. This would suggest that for certain types of information the application of additional processing resources would not effect improvements in performance, which could only be brought about by changes in the stimulus data. Further, it could be argued that print will gain access to both verbal and image systems, enhancing memory performance.

Day and Beach (1950) conducted one of the earliest and most thorough surveys of the research literature comparing visual and auditory presentation of information, and reached the following conclusions:

1. Meaningful, familiar material is more efficiently presented aurally, whereas meaningless and unfamiliar material is more efficiently presented visually.
2. The greater the reading ability of the receiver, the relatively more effective is a visual presentation.
3. The relative efficiency of a visual presentation increases with age from a definite inferiority at the age of six to a possible superiority at the age of sixteen.
4. Unusually difficult material is more effectively received with a visual presentation, whereas particularly easy material is better understood with an auditory presentation. The relative effectiveness of the visual presentation increases with increasing difficulty of the material.
5. The relative efficiency of a visual presentation diminishes as the interval of delayed recall increases.

There is a tendency here to support the predictions made on the basis of the new understanding of human information processing. Performance was found to be dependent on age, reading skill and the difficulty of the material and this was confirmed for studies using nonsense syllables or digits, meaningful words in serial order, and meaningful

prose reviewed by Hartman (1961a). He found that:

Relative channel effectiveness depends on the difficulty or complexity of the material for a given group of subjects provided the subjects can read. With material such as nonsense syllables and digits, adults find print an advantage. The reading skills of younger children, however, are so limited that the audio channel is probably the most effective for them regardless of task. (Hartman, 1961a, p.237)

In other words, the print channel becomes more effective relative to the audio channel as the difficulty of material increases, provided a fair degree of literacy is present. The meaningful prose materials consisted of educational talks, passages selected from fiction, essays and advertisements. Print was found to be superior in nine studies; audio was favoured in eleven; and three found no difference. With few exceptions, those studies favouring print dealt with adult subjects and complex stimulus materials and the studies supporting the relative effectiveness of the audio channel used young children as subjects, or adults and older children with simpler information.

Gulo and Baron (1965) presented prose material via print, lecture, television lecture and radio to college students and found reading to be superior to the other three presentations on a multiple-choice retention test. Thalberg (1964) found that for moderately difficult and very difficult material, reading was superior to listening for college students. These results were confirmed by Levie and Dickie (1973) in their later review.

One explanation for the superiority of reading relates to the greater referability of print, first recognized by Day and Beach (1950), combined with the fact that during equated time periods in which listening subjects were receiving only one exposure, reading subjects (who were receiving information at a faster rate, because reading is faster than speaking) could obtain more than one learning opportunity by re-reading. The results are also consonant with estimates of the higher rate of information transmission within the visual system and the probable lower data limitation threshold of the auditory system.

Words: Comparisons of Multi-channel with Single-channel Presentations

Results of studies comparing information presented to one sense modality with that presented to more than one modality are of particular importance when considering the Broadbent (1958) model of information processing as

elaborated by Travers (1967) for information transmission by means of audio-visual materials. In this model incoming information is compressed (redundancy is reduced) by the sense organs and peripheral nervous system so that some bits of information are 'chunked' together into fewer bits of information. This compressed information enters a central processing system which can handle sequential but not simultaneous information. If the central processing system is overloaded, some information may be stored briefly before it fades. The model predicts that if a presentation of the same information (spoken words and the same printed words), delivered simultaneously to both channels, is compared with either separately, no difference will be observed in the performance (recognition or recall) since all information must ultimately pass through a sequential processing system.

Later models, particularly those of Paivio (1971) and Morton (1980), support an alternative view of the results of such comparisons. A combination of spoken word and written word would activate both imagen and logogen systems, contributing to an increase in performance when compared with single presentations. Print alone would also have access to both systems, via the referential relation links, but not in such a direct manner.

The results of Day and Beach (1950) clearly support the later view. Another of their conclusions states unequivocally that a combined visual and auditory presentation leads to more efficient comprehension than either alone. Similarly, Hartman (1961a) reviewed nine studies comparing audio-print with print alone and found seven of them to favour the simultaneous presentation. Hartman (1961b) also found the audio-print presentation to be the most consistently effective of seven possible combinations of three channels tested. He concluded that a simultaneous audio-print presentation is more effective than either audio alone or print alone, when the simultaneously presented information is redundant (ie. when the same information is presented to both channels).

Van Mondfrans and Travers (1964), however, found evidence to support the Broadbent/Travers model. When recall of common words presented simultaneously by vision and audition was compared with words presented by vision alone and by audition alone no differences were found. Hsia (1968b) later found an audio-visual presentation of prose passages to be affected less than either audition or vision alone when 'noise' was introduced into the display (25% white noise in the audio channel; random black dots scattered over 25% of the visual display). Severin (1967) found that when recognition testing was provided in the channel or channels in which the information was presented, words presented simultaneously in audio and print were not significantly different from the print only condition, but

were superior to the audio only presentation. Interestingly, the most effective condition was audio plus related pictures, which was superior to the print condition, a result consonant with the dual-coding model.

Joy and John Menne (1972) confirmed Hartman's conclusion. They tested their subjects under aural, visual and audio-visual presentations. Three simple four-line verses, each containing 22 words were presented via television. The subjects each received an audio presentation, a visual presentation and a combined presentation of words, spoken and written. After each they were required to recite the verse. The results demonstrated that an audio-visual presentation was superior to either presentation alone.

The weight of evidence is clearly in favour of the later models of human information processing, with their emphasis on resource allocation, data limitation and dual coding of information in separate but related image and verbal systems. These studies do represent a rather special case, with the same material (words) being presented to the two channels.

There are some limited pedagogical applications that may be derived from this work, for example, in the teaching of languages, but comparisons of pictures and words are of much greater educational interest. After all, the principle aim of the early educational media theorists was to set education free from the constraints imposed by 'excessive verbalism'.

Words and Pictures: Comparisons of Single-channel Presentations

There is a complication in this area because it opens up intra-channel comparisons, as well as inter-channel comparisons. Thus, we can consider differences between words and their pictorial representations within one channel (print and pictorial: visual channel) and between channels (spoken words and pictures: auditory and visual channels). Much of the research has concentrated on the intra-channel comparison.

Hartman (1961a), in his review, concluded that the weight of studies comparing a pictorial presentation with an audio presentation (eg. pictures of objects compared with spoken words of the objects), favoured the pictorial mode, although he was cautious in his commitment to the generalization because of the small number of studies, and because one of the studies compared the effectiveness of the visuals in films with their sound tracks, a procedure which compares two elements which may have substantially different functions within the display. The results do, however,

confirm those obtained for spoken and printed words, which demonstrated superiority for the visual channel.

Comparisons between pictorial representations and written words have received much wider concern - the attraction of tighter experimental control is obviously one factor. Hartman (1961a) reported five such studies, all favouring the pictorial channel. Later studies (Gorman, 1961; Olver,1965; Rohwer, Lynch, Levin and Suzuki, 1967; Jenkins, Neale and Deno, 1967; Shepard, 1967) confirmed this in a variety of situations. Recall and recognition have been shown to increase from abstract words, to concrete words, to pictures, leading to Paivio's proposal for the dual-coding approach to perception and cognition.

Paivio (1971) describes one experiment in which subjects were shown words or pictures for a fraction of a second and were then asked to recall them. One group was tested after a delay of one week. Half of the subjects were not informed of the recall test until after the presentation ('incidental' group) . The stimulus items were composed of pictures, concrete words (ie. the words associated with the pictures) and abstract words. The immediate recall results demonstrated a clear advantage for pictures, with concrete words superior to abstract words, for both 'incidental' and 'intentional' groups. The delayed responses maintained this order, with the incidental group showing a higher delayed recall of pictures than was previously obtained for concrete words on immediate testing. Needless to say, this and many other studies led eventually to the development of the later models of human information processing.

Words and Pictures: Comparisons of Multi-channel with Single-channel Presentations

This strikes at the very heart of audio-visual theory. What evidence is there to support the idea that a combination of audio and visual will bring about more complete learning than either channel alone?

In most of the early studies information presented to the two channels was related (eg. a pictorial representation of an object and a verbal description of it) although other possibilities do exist: the information may be unrelated or even contradictory. Hartman's (1961a) review found the available thirteen studies to be inadequate for the formulation of generalizations, but did conclude that comparisons of pictorial-verbal presentations with single-channel presentations strongly indicated an advantage for the combination of channels.

Early researches dealing with motion film variables (Hoban and Van Ormer, 1950) supported the advantage for sound plus picture combinations, and this was confirmed by

Ketcham and Heath (1962) who demonstrated that sound plus relevant pictures was superior to sound alone, with sound plus irrelevant pictures proving to be least effective. When comparing the effectiveness of relevant pictures in multiple-channel communication Severin (1967) found audio plus related pictures to be the most successful of six combinations, including print alone and audio alone, with unrelated-picture/audio combinations least effective. More recently, Nugent (1982) demonstrated that when content is the same in visual, audio and print channels, younger students learn equally well from all modes, but combining pictures with print or audio generally maximized learning. This research suggests that by conveying information through both linguistic and iconic symbols students were provided with complementary processing systems, and they could alternate between the two to obtain information. This study offers support for the theoretical position that learners process pictorial and linguistic information through functionally independent, though interconnected, cognitive systems.

The most thorough exploration of multiple-channel communication has been undertaken by Dwyer (1978). In a series of experiments starting in 1967 Dwyer systematically studied the effects of different conditions of presentation of the same material. This consisted of a study unit 'The Human Heart', which was presented either as an oral/verbal presentation, with the recorded script being accompanied by keys words in the visual display, or as an illustrated presentation, with pictures accompanying the recorded script. Several different series of pictures were prepared, ranging from simple line drawings to full colour photographs. In addition to different types of illustration, there were four different types of test: a drawing test, in which the subject drew a labelled representative diagram of the heart; an identification test, in which subjects had to identify certain numbered parts of a heart; a terminology test which evaluated knowledge of specific terms; and a comprehension test which evaluated understanding and an ability to use information to explain phenomena. The individual scores were pooled to give a 'total criterial score', showing the subject's total knowledge and understanding.The results of the first experiment led Dwyer to conclude that:

The results of this study indicate that when students viewed their respective instructional presentations for equal amounts of time, the simple line drawing presentation was significantly more effective in facilitating achievement than was the oral presentation without visuals on the drawing, identification and total criterial tests. The oral presentation without visuals

of the heart was found to be as effective as each of the visually complemented treatments on both the terminology and comprehension tests. (Dwyer, 1972, p.22)

The results have not always been quite so clear-cut. With different populations of subjects the results have sometimes shown visuals to be beneficial for the terminology and comprehension tests. Visuals produce consistently higher levels of achievement for the drawing, terminology and total criterial test, when compared with the oral presentation, although such benefits are not sustained over a two week delayed-testing period. However, Dwyer's research consistently demonstrates the benefits associated with pictorial accompaniments to spoken commentaries, and provides additional support to the audio-visual theorists.

Texts and Illustrations

The effects of illustrations on the understanding of text is obviously of pedagogic interest. It is assumed that illustrations must add to recall, comprehension and understanding. Otherwise, why should publishers of school exts take so much trouble to prepare them? However, despite this assumption, much of the early research on text illustrations reported a slight advantage for text alone (Miller, 1938; Vernon, 1953, 1954) and this led Samuels in 1970 to conclude that pictures do not facilitate comprehension of text. He was particularly concerned that the assumed benefits of illustrations produced to supplement texts should be subjected to scrutiny, claiming that:

If fish were to become scientists, the last thing they might discover would be water. Similarly, researchers have too often failed to investigate important aspects of their environment because being immersed in it, they fail to notice certain components of it; or, having noticed a component, they simply assume it must be that way. One such example from reading is the ubiquitous use of illustrations.... (Samuels, 1970, p.397)

Samuels suggested that pictures are used in instructional materials under the assumption that they facilitate learning by appealing directly to the student's five senses. This was tested for by Bourisseau, Davis and Yamamoto (1965) but, as with other studies reviewed by Samuels, the results were negative and the authors concluded that although many people believed that one picture is worth a thousand words, so far as sensory response is concerned, the assumption is unwarranted.

In conclusion Samuels suggested that the answer to the question 'Should pictures be used as adjuncts to printed texts?' depended on the objectives.

If the objective is to promote acquisition of a sight vocabulary, the answer would seem to be 'no.' If the objective is to facilitate comprehension, the answer is less definite.... Although the research, in general, does not show that pictures aid comprehension, neither does it show that they hinder comprehension (Samuels, 1970, p.405)

Much subsequent research disagrees with this view, but the out-of-date reviews may still be sighted authoritatively as sources of wisdom concerning this topic, according to Levie and Lentz (1982). Having 'discovered water' later researchers began to analyze the constituent parts and determine how they interacted.

Levie and Lentz (1982) reviewed the later research, starting by looking at the effect of visual illustrations on the learning of verbal information presented in the text. This excluded oral or audio-visual presentations; word lists, single sentences, or deliberately ambiguous passages; charts, diagrams and graphs. Essentially they were looking at studies comparing learning from illustrated texts versus learning from text alone. They took what they termed an 'information-analysis approach' to the research, based on the possible relationships between the information in a text passage and in the illustrations. Information presented only in the text passage (ie. without illustration) they termed 'non-illustrated text information'; other information provided only in the pictures was termed 'picture-only information'; and when text material was accompanied by relevant illustrations it was called 'illustrated text information.'

The researcher's test of learning is a critical factor in evaluating results in this area. For example, does the test measure the learning of illustrated text information only, non-illustrated text information only, or some combination? Levie and Lentz (1982) claimed that when researchers control these relationships, greater clarity about the effects of text illustrations is possible. They identified a total of 46 comparisons of learning illustrated text information from illustrated text versus text alone. A majority of the results were statistically significant and in favour of the illustrated text, giving an average effect size of 0.55, with the illustrated text group score averaging 36% more than the text alone. This led Levie and Lentz to conclude that:

With a few exceptions, the results showed that illustrations had a significant positive effect on

learning illustrated text information and no effect on learning non-illustrated text information. (Levie and Lentz, 1982, p.198)

Vernon (1953) suggested that illustrations laid undue emphasis on certain points and therefore distracted attention from other aspects of the text, thus hindering overall performance. A more popular view is that illustrations have a generalized motivating effect and will improve learning of information which is not illustrated. Levie and Lentz (1982) lend support to the middle view, that illustrations simply have no effect on learning non-illustrated text information.

When learning a combination of illustrated and non-illustrated text information it can be predicted from the above results that the effect of including illustrations will depend on the proportion of supported text in the test employed. This is confirmed by Levie and Lentz, who indicate that 'the addition of pictures should not be expected to hinder learning; nor should pictures always be expected to facilitate learning.' However, they conclude that learning is better with pictures in most cases.

Dwyer's researches include many studies in which the materials are presented in the form of a text, and the results of these were also analyzed by Levie and Lentz. Similar differential effects were observed for the different types of test as were found in the studies using an audio-visual presentation, with the largest effect size in favour of illustrations being observed for the drawing test (0.58), lesser effects for the identification test (0.38) and terminology test (0.28), and a trivial effect for the comprehension test (0.09). These results were interpreted as showing that the degree to which an educational objective is aided by pictures depends on the emphasis given to knowledge about spatial information in the test of learning, demonstrating a consistency between Dwyer's work and other research in this area.

Levie and Lentz's results have been confirmed in a further meta-analysis conducted by Levin, Anglin and Carney (1987). Levin (1981) had previously speculated that illustrations could be categorized as having five different functions: decoration, representation, organization, interpretation and transformational. He suggested that each of these functions would produce different degrees of prose learning facilitation, ranging from no effect for the decorative function to substantial for transformational pictures designed specifically with a mnemonic objective in mind. Moderate to substantial facilitation was anticipated for representation, organization and interpretation functions. The meta-analysis confirmed these expectations with a near zero effect size for decorative illustrations

and an average 0.71 across other text-relevant studies. For children the average effect size ranged from 0.5 for representational pictures to 1.43 for transformational illustrations; for older subjects the range was from 0.7 to 1.04. These results further confirm the power of illustrations to enhance prose learning, particularly when pictorial information is further encoded as in the transformational mode.

Pictorial Complexity

Levie and Lentz (1982) comment that even within the limits imposed on studies considered in their review there is still considerable diversity, particularly in the types of illustration used, which varied from simple line drawings to colour photographs. Dwyer's work demonstrated advantages for including illustrations compared with no illustrations for specific types of test, offering support to the audio-visual approach, but as yet the differential effects of the various types of illustrations included in his researches have not been discussed. Four different types of

Figure 5-12. Dwyer's realism continuum for black and white and coloured illustrations.

1. Simple line representations (black and white)
2. Simple line representations (colour)
3. Drawings - detailed and shaded (black and white)
4. Drawings - detailed and shaded (colour)
5. Photographs of a model (black and white)
6. Photographs of a model (colour)
7. Photographs of a real heart (black and white)
8. Photographs of a real heart (colour)

illustration were used in most of Dwyer's experiments. They varied in the amount of detail and general pictorial complexity, and were produced in a black and white version and a colour version, giving a total of eight different types which could be arranged, according to Dwyer, along a 'realism continuum' (Figure 5-12).

Dwyer (1978) suggested that the two competing theories of human information processing, represented by the single-channel and multiple-channel perspectives, predicted different results for experiments varying the degree of realism in illustrations. The multiple-channel theorists, which Dwyer identified as stemming from the writings of Dale (1946) and Carpenter (1953), predicted an increase in the efficiency in facilitating learning as the realism increased

from the simple line representations to the photographs of a real heart. The single-channel theorists, based on the work of Broadbent (1958) and Travers (1964), predicted no effect from the increase in pictorial complexity, indeed when information in picture and text was redundant they predicted no effect from the addition of illustrations to the text. The results of the initial experiment (Dwyer, 1967) was equivocal. Illustrations did facilitate achievement for a majority of the test measures. However, increasing the degree of realism in the illustrations did not increase the performance on any of the tests. In other words, the addition of a simple line drawing enhanced performance on three of the five tests, but there was no difference in the effects of any of the different types of illustrations.

Further experiments reported by Dwyer (1972) tended to support the initial findings with a majority of the studies demonstrating superiority for the presentations illustrated with simple line drawings when assessed on the total criterial test and the drawing test. For the terminology and comprehension tests illustrations were generally not an advantage; and the identification test indicated an intermediate position, with many studies indicating benefits from the addition of simple line illustrations, and similar numbers indicating no such advantage. In only a minority of cases (16%) were illustrations which were more complex than the simple line drawings more effective, and there was no discernible pattern to this minor effect.

These results are not surprising when considered against the background of several decades of research in this area. Vandermeer (1954) had investigated the comparative effects of colour and black and white motion films, on the assumption that colour would add realism to the visual display, which would increase the efficiency of the films as teaching aids. This was not the case, and he concluded that:

> The use of colour in instructional films which may seem to 'call for colour' does not appear to be justified in terms of the greater learning on the part of those who view the films. If colour is to be used effectively in films there must be careful pre-production consideration of the probable psychological impact of specific uses of colour upon the learner. (Vandermeer, 1954, p.134)

These results were confirmed by later researchers (eg. Kanner and Rosenstein, 1960) and reviews (Kanner, 1968; Cox, 1976).

A fascinating report from the Yale film research programme (May and Lumsdaine, 1958) showed the lack of effects from increased pictorial complexity when they compared a crude black and white film version of a film, based on the story-board sketches, with the finished full colour motion picture. The results were totally unexpected: there was no difference in the achievement of the objectives between the two versions, although there was a large difference in cost. Twyford (1954) had obtained similar results when comparing a full colour motion picture with a version consisting of a series of still pictures in black and white. There were slight differences in favour of the expensive colour version, but the differences obtained were not of educational significance, and Twyford cautioned against the indiscriminate use of realistic presentations when they could not be justified in terms of instructional benefits.

Human Information Processing: Conclusions

There has been a healthy rivalry between theorists advocating either a single-channel of limited capacity or a multi-channel approach to human information processing. The present position calls for a resolution of the different positions. In part, both positions are correct.

The single-channel theory proposed by Broadbent (1958) and elaborated by Travers (1967) is no longer wholly tenable, and it is reasonable to assume that varying allocations of attention can be made to incoming information, with apparently unattended information receiving high levels of processing, as suggested by Norman and Bobrow (1975). Incoming information may receive variable amounts of processing, and the depth of processing may depend upon such factors as the 'realism' of the display, the receiver's perceived self-efficacy and the perceived demand characteristics of the particular medium.

There is good evidence to support two complementary representation systems, along the lines suggested by Paivio (1971), with extensive interconnections between verbal and image systems. This is clearly acceptable to the multi-channel advocates, and supports the early claims of the 'audio-visual' approach to educational technology.

However, there are limits to the audio-visual approach. Although multi-sensory presentations do seem to facilitate learning on specific tests, they do so only in circumstances where audio and visual components are mutually supportive. Visuals increase learning when they are relevant to the text or commentary, they do not produce a generalized effect on unrelated textual or verbal information. Visual information also appears to be compressed by the processing system, as suggested by Travers (1964). This is necessary because of the limited capacity of the system and precludes any facilitating effects anticipated by increasing pictorial complexity or realism. It suggests that for the more common activities associated with visual testing (eg. drawing,

identifying) students may benefit from visuals which are pre-compressed and simplified by the instructor.

REFERENCES

Attneave, F. (1959) Applications of Information Theory to Psychology. New York: Holt-Dryden.

Bleasdale, F. (1983) 'Paivio's dual-coding model of meaning revisited'. In J.C. Yuille (Ed) Imagery, Memory and Cognition. New Jersey: Lawrence Erlbaum.

Bourisseau, W., Davis, O.L. and Yamamoto, K. (1965) 'Sense impression responses to differing pictorial and verbal stimuli', Audio Visual Communication Review, 13, 249-254.

Broadbent, D.E. (1954) 'Role of auditory localization and attention in memory span', Journal of Experimental Psychology, 47, 191-196.

Broadbent, D.E. (1956) 'Successive responses to simultaneous stimuli', Quarterly Journal of Experimental Psychology, 8, 145-152.

Broadbent, D.E. (1958) Perception and Communication. Oxford: Pergamon Press.

Calkins, M.W. (1898) 'Short studies in memory and in association from the Wellesley College psychological laboratory', Psychological Review, 5, 451-462.

Carpenter, C.R. (1953) 'A theoretical orientation for instructional film research', Audio Visual Communication Review, 1, 38-52.

Cherry, E.C. (1953) 'Some experiments on the recognition of speech, with one and two ears', Journal of the Acoustical Society of America, 25, 975-979.

Cherry, E.C. and Taylor, W.K. (1954) 'Some further experiments on the reception of speech, with one and two ears', Journal of the Acoustical Society of America, 26, 554-559.

Cox, S. (1976) 'Colour in learning from film and tv: a survey of the research with some indications for future investigations', Educational Broadcasting International, (June.)

Craik, F.M. and Tulving, E. (1975) 'Depth of processing and retention of words in episodic memory', Journal of Experimental Psychology: General, 104, 208-294.

Craik, F.I.M. and Lockhart, R.S. (1972) 'Levels of processing: a framework for memory research', Journal of Verbal Learning and Verbal Behaviour, 11, 671-684.

Dale, E. (1946) Audiovisual Methods in Teaching. London: Holt, Rinehart and Winston.

Dale, E. (1954) Audiovisual Methods In Teaching (2nd Edition). London: Holt, Rinehart and Winston.

Day, W.F. and Beach, B.R. (1950) A Survey of the Research Literature Comparing the Visual and Auditory Presentation of Information. Air Force Technical Report 5921, Contract No. W33- 039- AC-21269. Charlottesville: University of Virginia.

Deutsch, J.A. and Deutsch, D. (1963) 'Attention: some theoretical considerations', Psychological Review, 70, 80-90.

Dwyer, F.M. (1967) 'Adapting visual illustrations for effective learning', Harvard Educational Review, 37, 250-263

Dwyer, F.M. (1972) A Guide for Improving Visualized Instruction. State College, Pennsylvania: Learning Services.

Dwyer, F.M. (1978) Strategies for Improving Visual Learning. State College, Pennsylvania: Learning Services.

Eysenck, M.W. (1984) A Handbook of Cognitive Psychology. London: Erlbaum

Gibson, J.J. (1954) 'A theory of pictorial perception', Audio Visual Communication Review, 2(1), 3-23.

Gorman, A.M. (1961) 'Recognition memory for nouns as a function of abstractness and frequency', Journal of Experimental Psychology, 61, 23-29.

Gulo, E.V. and Baron, A. (1965) 'Classroom learning of meaningful prose by college students as a function of sensory mode of stimulus presentation', Perceptual and Motor Skills, 21, 183-186.

Hake, H.W., and Garner, W.R. (1951) 'The effect of presenting various numbers of discrete steps on scale reading accuracy', Journal of Experimental Psychology, 42, 358-366.

Hartman, F.R. (1961a) 'Single and multiple channel communication: a review of research and a proposed model', Audio Visual Communication Review, 9, 235-262.

Hartman, F.R. (1961b) 'Recognition learning under multiple channel presentation and testing conditions', Audio Visual Communication Review, 9, 24-43.

Hoban, C.F. and Van Ormer, E.B. (1950) Instructional Film Research 1918-1950. Technical Report No. SDC-269-7-19 Port Washington, New York: Special Devices Centre, Office of Naval Research.

Hsia, H.J., (1968a) 'On channel effectiveness', Audio Visual Communication Review, 16, 245-267.

Hsia, H.J., (1968b) 'Effects of noise and difficulty level of input information in auditory, visual, and audiovisual information processing', Perceptual and Motor Skills, 26, 99-105.

Jacobsen, H. (1950) 'The information capacity of the human ear', Science, 112, 143-144.

Jacobsen, H. (1951) 'The informational capacity of the human eye', Science, 113, 292-293.

Jenkins, J.R., Neale, D.C., and Deno, S.L. (1967) 'Differential memory for picture and word stimuli', Journal of Educational Psychology, 58, 303-307.

Jonassen, D.H. (1985) 'Learning strategies: a new educational technology', Programmed Learning and Educational Technology, 22(1), 26-34.

Kanner, J.H. (1968) The instructional effectiveness of colour television: a review of the evidence. Stanford University: Stanford, California. Ed 015 675

Kanner, J.H. and Rosenstein, A.J. (1960) 'Television in army training: colour versus black and white', Audio Visual Communication Review, 8, 243-252.

Ketcham, C.H. and Heath, R.N. (1962) 'Teaching effectiveness of sound with pictures that do not embody the material being taught', Audio Visual Communication Review, 10, 89-93.

Kirkpatrick, E.A. (1894) 'An experimental study of memory', Psychological Review, 1, 602-609.

Klemmer, E.T., and Frick, F.C. (1953) 'Assimilation of information from dot and matrix patterns', Journal of Experimental Psychology, 45, 15-19.

Knowlton, J.Q. (1966) 'On the definition of a picture', Audio Visual Communication Review, 14, 157-183.

Langer, E.J. (1984) 'Playing the Middle Against Both Ends: The Influence of Adult Cognitive Activity as a Model for Cognitive Activity in Childhood and Old Age.' In S.R. Yussen (Ed) The Development of Reflection. New York: Academic Press.

Levie, W.H. and Dickie, K.E. (1973) 'The Analysis and Application of Media'. In Travers, R.M.W. (Ed), Second Handbook of Research on Teaching. Chicago: Rand McNally.

Levie, W.H. and Lentz, R. (1982) 'Effects of text illustrations: a review of research' Educational Communication and Technology Journal, 30, 195-232.

Levin, J.R. (1981) 'On the Function of Pictures in Prose'. In F.J. Pirozzolo and M.C. Wittrock (Eds) Neuropsychological and Cognitive Processes in Reading. New York: Academic Press.

Levin, J.R., Anglin, G.J. and Carney, R.N. (1987) 'On Empirically Validating Functions of Pictures in Prose'. In D.A. Willow and H.A. Houghton (Eds) The Psychology of Illustration (Vol.1). London: Springer-Verlag.

May, M.A. and Lumsdaine, A.A. (1958) Learning from Films. Yale University Press: New Haven

Menne, J.M., and Menne, J.W. (1972) 'The relative efficiency of bimodal presentation as an aid to learning', Audio Visual Communication Review, 20, 170-180.

Miller, G.A. (1956) 'The magical number seven, plus or minus two: some limits on our capacity for processing information', Psychological Review, 63, 81-97.

Miller, G.A. and Frick, F.C. (1949) 'Statistical behaviouristics and sequences of responses', Psychological Review, 56, 311-324.

Miller, W.A. (1938) 'Reading with and without pictures', Elementary School Journal, 38, 676-682.

Moray, N. (1959) 'Attention in dichotic listening', Quarterly Journal of Experimental Psychology, 11, 56-60.

Morris, C.W. (1946) Signs, Language and Behaviour. New York: Prentice-Hall

Morton, J. (1969) 'The interaction of information in word recognition', Psychological Review, 76, 175-178.

Morton, J. (1980) 'The logogen model and orthographic structure'. In U. Frith (Ed) Cognitive Processes in Spelling. London: Academic Press.

Norman, D.A. (1968) 'Toward a theory of memory and attention', Psychological Review, 75, 522-536.

Norman, D.A. and Bobrow, D.G. (1975) 'On data-limited and resource-limited processes', Cognitive Psychology, 7, 44-64.

Nugent, G. (1982) 'Pictures, audio, and print: symbolic representation and effect on learning', Educational Communication and Technology Journal, 30, 163-174.

Olver, M.A. (1965) 'Abstractness, Imagery, and Meaningfulness in Recognition and Free Recall.' Unpublished Master's Thesis, University of Western Ontario. Quoted in A. Paivio (1971) Imagery and Verbal Processes. New York: Holt, Rinehart and Winston.

Paivio, A. (1971) Imagery and Verbal Processes. New York: Holt, Rinehart and Winston.

Paivio, A., Rogers, T.B. and Smythe, P.C. (1968) 'Why are pictures easier to recall than words?', Psychonomic Science, 11, 137-138.

Piaget, J. (1953) The Origins of Intelligence in the Child. London : Routledge and Kegan Paul.

Pollack, I. (1952) 'The information of elementary auditory displays', Journal of the Acoustical Society of America, 24, 745-749.

Pollack, I. (1953) 'The information of elementary auditory displays II', Journal of the Acoustical Society of America, 25, 765-769.

Pollack, I., and Ficks, L. (1955) 'The information of elementary multi-dimensional auditory displays', Journal of the Acoustical Society of America', 26, 155-158.

Rohwer, W.D., Lynch, S., Levin, J.R. and Suzuki, N. (1967) 'Pictorial and verbal factors in the efficient learning of paired associates', Journal of Educational Psychology, 58, 278-284.

Salomon, G. (1984) 'Television is easy and print is tough: the differential investment of mental effort in learning as a function of perceptions and attributions', Journal of Educational Psychology, 76, 647-658.

Sampson, J.R. (1970) 'Free recall of verbal and non-verbal stimuli', Quarterly Journal of Experimental Psychology, 22, 215-221.

Samuels, S.J. (1970) 'Effects of pictures on learning to read, comprehension and attitudes', Review of Educational Research, 40, 397-407.

Severin, W.J. (1967) 'The effectiveness of relevant pictures in multiple-channel communications', Audio Visual Communication Review, 15, 386-401.

Seymour, P.H. (1973) 'A model for reading, naming, and comprehension', British Journal of Psychology, 64, 35-49.

Shannon, C.E. and Weaver,W. (1949) The Mathematical Theory of Communication. Urbana: University of Illinois Press.

Shepard, R.N. (1967) 'Recognition memory for words, sentences and pictures', Journal of Verbal Learning and Verbal Behaviour, 6, 156-163.

Standing, L., Conezio, J. and Haber, R.N. (1970) 'Perception and memory for pictures: single trial learning of 2,500 visual stimuli', Psychonomic Science, 19, 73-74.

Thalberg, P.S. (1964) An Experimental Investigation of the Relative Efficiency of the Auditory and Visual Modes of Presentation of Verbal Material. Doctoral dissertation, State University of Iowa. Ann Arbor, Michigan: University Microfilms, No. 64-7949.

Thorndike, E.L. and Lorge, I. (1944) The Teacher's Word Book of 30,000 Words. New York: Bureau of Publications. Teachers College.

Travers, R.M.W. (1964) 'The transmission of information to human receivers', Audio Visual Communication Review, 12, 373-385.

Travers, R.M.W. et al. (1964) Research and Theory Related to Audiovisual Information Transmission. Project No. 3-20-003. U.S. Department of Health, Education and Welfare, Office of Education.

Travers, R.M.W. et al. (1967) Research and Theory Related to Audiovisual Information Transmission, Revised Edition. Project No. 3-20-003. U.S. Department of Health, Education, and Welfare, Office of Education.

Treichler, D.G. (1967) 'Are you missing the boat in training aids?', Film and AV Communication, 1, 14-16.

Treisman, A.M. (1960) 'Contextual cues in selective listening', Quarterly Journal of Experimental Psychology, 12, 242-248.

Treisman, A.M. and Geffen, G. (1967) 'Selective attention: perception or response?', Quarterly Journal of Experimental Psychology, 19, 1-16.

Twyford, L.C., Davis, J.S. and Seitz, C.P. (1954) Evaluation of the Film: Military Police Support in Emergencies (Riot Control). Technical Report No. SDC 269-7-52. Instructional Film Program. Naval Training Devices Centre; Office of Naval Research: Port Washington-New York.

Van Mondfrans, A.P. and Travers, R.M.W. (1964) 'Learning of redundant material presented through two sensory modalities', Perceptual and Motor Skills, 19, 743-751.

Vandermeer, A.W. (1954) 'Colour vs black and white in instructional films', Audio Visual Communication Review, 2, 121-134.

Vernon, M.D. (1953) 'The value of pictorial illustration', British Journal of Educational Psychology, 23, 180-187.

Vernon, M.D. (1954) 'The instruction of children by pictorial illustration', British Journal of Educational Psychology, 24, 171-179.

CHAPTER 6

CONSTRUCTING REALITY

Eysenck (1984) has suggested that experimental psychology has increasingly become synonomous with cognitive psychology and identifies Broadbent's 'Perception and Communication' (1958) and Neisser's 'Cognitive Psychology' (1967) as being seminal influences on the 'booming, buzzing confusion that is contemporary cognitive psychology.'

Neisser (1976), in turn, identifies the developmental studies of Piaget as offering an alternative to earlier behaviouristic models of human behaviour, claiming that 'human nature is too important to be left to the behaviourists....' (p.8). Central to Neisser's position is the concept of the schema, which has been developed extensively by Piaget.

> Although perceiving does not change the world, it does change the perceiver. (So does action, of course.) The schema undergoes what Piaget calls 'accommodation,' and so does the perceiver. He has become what he is by virtue of what he has perceived (and done) in the past; he further creates and changes himself by what he perceives and does in the present. 'Existence precedes essence,' as the existentialists say. (Neisser, 1976, p.53)

Jonassen (1985) sees the present situation as a shift from a behaviouristic, mechanistic psychology of human behaviour to that of an active, constructive psychology, with learners viewed as organized entities actively constructing their own reality. He sees Piaget's theory of cognitive development as being of major importance to the new educational technologies, particularly Piaget's dictum that knowing is a process of constructing systems of transformations that model reality and that mature 'operational' thought is the product of:

> successive constructions, and the principal factor in this constructivism is an equilibration achieved by autoregulations that make it possible to remedy momentary incoherences, to resolve problems, and to surmount crises or periods of imbalance by a constant elaboration of fresh structures.... (Piaget, 1971, p.40)

PIAGET'S CONSTRUCTIVISM

Piaget rarely directed his attention specifically to education. Primarily he considered himself to be a biologist and a genetic epistemologist, indeed, he never really left biology even when he was so occupied with cognition.

For many years Piaget studied the water-snail, Limnaea stagnalis, which in tranquil water has an elongated shell. When moved to lakes where waves beat upon its shell, Piaget found that the young snail changes its shape, it contracts the muscles attaching it to rocks each time a wave breaks and eventually it becomes globular in shape. In time the snail's adaptation to the turbulent waters tends to become hereditary. Piaget moved some back to his aquarium where they retained their globular shape for many generations. For Piaget this is an example of how organisms may adapt genetically to novel surroundings. He does not fully subscribe to the Lamarkian heresy of the inheritance of acquired characteristics, but he disagrees with neo-Darwinian theory which sees organisms as essentially passive and dependent on chance mutation for adaptive change in the fight for survival. For Piaget, the adaptation is a dynamic, on-going process in which the hereditary structure of the organism interacts with the external environment in such a way as to reconstitute itself for better survival. Piaget (1971) identifies Waddington's (1957) theory of genetic assimilation as providing a basis for the mechanism by which such adaptive changes become genetically transmitted. This has led to Piaget's emphasis on the process that brings about progressive states of equilibrium: the dynamic, on-going self-regulating process which he calls equilibration.

His theory of cognitive development was presented in more than a hundred books and articles spanning more than fifty years. Many educationalists have been inspired by these writings, though Piaget has frequently pointed out that his ideas have not always been fully understood by such educationalists. Piaget's book 'The Science of Education and the Psychology of the Child' (1971) represents the fullest account of his thoughts on education. However, in order to understand these writings it is necessary to have a grasp of the basics of his theory.

Assimilation, Accommodation and Adaptation

Adaptation is a biological mode of functioning that characterizes all forms and levels of life. It consists of the dual processes of assimilation and accommodation, which go on continuously. Assimilation is the process of taking in from the environment all forms of stimulation and information, which are then organized and integrated into

the organism's existing forms or structures. Accommodation implies modification of the structures of the organism so that it can adapt to the environment.

The young infant investigating his immediate environment, grasps, sucks, explores, probes and absorbs it, assimilating the experiences. But the environment, whether of inert objects or living things, resists, moves, rewards and punishes the infant, compelling him to accommodate, to adjust to new and changing conditions, so that pre-existing patterns of behaviour are modified to cope with new information or feed-back from the external situations.

Piaget states that intellectual growth is propelled by disequilibrium, by a failure of accommodative and assimilative processes to keep pace with one another, by failure of the child's models of the environment to suffice in coping with newly learned complexities of that environment, giving rise to the jest that it is the 'trouble theory of development.'

Assimilation, Accommodation – an example: If I present you with the following numbers: 2 and 4, and ask you to decide what the next number is, you will probably form the theory that this is a progression of even numbers and therefore that the next number in the series is 6. When I tell you that you are correct you assimilate this information into your theory. The information fits. I now ask you to provide the next number in the series and you will probably surmise that it is 8. However, this time I have to tell you that your answer is wrong and that the correct answer is 10. You are now in a state of disequilibrium, to use Piaget's terminology, and until you accommodate to the new information you will remain in disequilibrium. In order to attain a steady state or equilibrium you must re-organize your theory, in order to reach a balance between assimilation and accommodation. You can see why it is called the trouble theory of development. The progression is achieved by adding the last number to the penultimate number in the series, so the next will be 10 plus 6. We can summarize this cognitive process by saying:

because of the intellect's organisation it exhibits adaptation, a two-way process comprising accommodation and assimilation.

Physical, Logico-mathematical and Social-arbitrary Knowledge

The concept of equilibration is a dynamic one, and it is therefore necessary to indicate what upsets the organism once it has achieved equilibrium. The first dynamic factor is maturation, the physiological growth of the genetic

organic structures. Maturation only opens up future possibilities in Piaget's scheme, experience is also necessary to actualize the potential, experience derived from commerce with the environment. These interactions with the environment lead the child to construct two different kinds of knowledge: physical and logico-mathematical.

Physical knowledge is associated with the results of actions on objects: touching, lifting, throwing, hitting, smelling etc. To construct this knowledge the child must have repeated active experiences in which he carries out manipulations of objects. When constructing physical knowledge the physical properties of the objects themselves correct or reinforce the child's learning. Implicit in Piaget's conception of the development of the child is the idea that children acquire knowledge about the physical properties of objects by manipulating the objects and that all other ways of coming to know objects are qualitatively inferior. Reading about or listening to someone talk about an object cannot provide the quality of physical knowledge that can be acquired by active manipulation, a fact that is acknowledged by Gagné:

The great value of concepts as means of thinking and communicating is the fact that they have concrete references.... but since concepts are learned by the human being via language, there is often the danger of losing sight of this concreteness. Learning can become over-verbalized which means that the concepts learned are highly inadequate in their reference to actual situations. The learner does not really know the meaning of the word even though he can use it correctly in a sentence. (Gagné, 1970, p.187)

Logico-mathematical knowledge is derived from acting upon objects, but does not come from the objects themselves, it is constructed from the child's actions on the objects. Logical relationships are constructed in the child's mind, such as higher, lower, faster. Mathematical relationships also are not implicit in objects but are constructed by the observing or counting child. Number and number-related concepts are logico-mathematical types of knowledge, but number is not a quality or characteristic of an object, it is an invention of each child.

Of course, physical knowledge and logico-mathematical knowledge develop in an inter-dependent manner, being tied together in their formation.

The third factor contributing to cognitive development is the transmission of social-arbitrary knowledge. Here information is learned from other children or transmitted by parents, teachers, or books in the process of education. Children learn language, values, rules, morality and symbol

systems, in fact social-arbitrary knowledge is everything that is handed on from generation to generation, that which constitutes the essential foundation of human society. Piaget believes that when a child hears contradictory statements concerning this realm of knowledge his equilibrium is disturbed, effecting a search for an answer which will reduce the cognitive conflict.

It is the equilibration process which co-ordinates and regulates the three factors of cognitive development (maturation, experience and social-arbitrary knowledge), bringing about progressive states of equilibrium. Equilibration is the organizing factor underlying all biological and intellectual development.

The Relation between Subject and Object

An implicit assumption in behaviourist, stimulus-response psychology is that knowledge is a copy of reality. It is also a frequent assumption of traditional and conventional educational practice.

Piagetian thought provokes the question: is knowledge a copy of reality or is it a unique, individual construction? The traditional view has been that knowledge is a directly acquired copy of reality, the child was thought to incorporate accurate images of the real world directly. The latest ideas in cognitive psychology reject this idea. As Neisser (1967) puts it in his 'Cognitive Psychology':

> Whatever we know about reality has been mediated, not only by organs of sense but by complex systems which interpret and re-interpret sensory information. (Neisser, 1967, p.3)

Piaget's major contention is that each individual constructs reality, and for each child this is a unique construction: and this is what he means when, in 'The Science of Education and the Psychology of the Child', he says:

> Knowledge is derived from action, not in the sense of simple associative responses, but in a much deeper sense of assimilation of reality into the necessary and general co-ordination of action. To know an object is to act upon it and transform it.... intelligence consists in executing and co-ordinating actions, though in an interiorised and reflective form. (Piaget, 1971, p.28-29)

For Piaget, children's conceptions develop from those initially based on limited information - unrefined and frequently inaccurate schemes - to increasingly more accurate and refined schemes. The essential functions of intelligence consist in understanding and inventing, in other words of building up structures by structuring reality.

In relation to the different types of knowledge that may be acquired Piaget (1971) describes two aspects of thinking that are different but functionally related: the figurative aspect and the operative aspect.

Figurative Knowledge

The person looking at a tree sees it figuratively, but in his mind he also conceives it as alive, breathing, having roots, branches etc. These latter aspects are operative conceptions that go beyond the limitations of the former figurative knowledge. Figurative thoughts are perceptions, imitations, mental images, those aspects of thought that deal with states as they appear in the senses. They are the thoughts that occur at the time of perception of external objects, or at the perception of mental images. When we look at an object and our thought produces identification of that object - that is figurative thought.

A child looking at a tree sees leaves, branches etc: what she sees is figurative knowledge about the tree.

Operative Knowledge

Operative thought is the product of perceptions (figurative) and intelligence. Operative knowing is the product of transformations of states resulting from actions. It is the organization, consolidation and interaction of figurative knowledge, and intellectual development can be characterized as the growth of operative knowledge.

Operative and figurative knowing are related. Figurative knowledge feeds the development of operative knowledge - it is its source. And, of course, changes in operative knowing result in changes in our subsequent perceptions. Thus, the child looking at the tree who realizes that it is alive (operative knowledge) looks at it in a different way and no longer gets the same figurative knowledge.

The Theory of Stages

According to Piaget, the organization of intelligence changes qualitatively in several stages throughout the individual's life. Similar conceptual processes are thought to account for strategies employed at each stage.

Figure 6-1. Piagetian Stages of Cognitive Development.

		MENTAL AGE
I	SENSORIMOTOR PHASE	0-1.5 years
II	PRE-OPERATIONAL PHASE	1.5-7 years
	(i) Pre-conceptual stage (ii) Stage of intuitive thought	(1.5-4 years) (4-7 years)
III	CONCRETE OPERATIONS PHASE	7-11 years
IV	FORMAL OPERATIONS PHASE	11 years and over

The order of the stages is fixed, each stage being regarded as essential preparation for the next one. In part, the stages are a result of maturation of the nervous system of the growing child. However, maturation is necessary but not sufficient for cognitive development: biological maturation does nothing more than open the way to possible constructions. It is interaction with the environment which actualizes them. Piaget acknowledges that some pedagogical interventions can accelerate and complete development (the American question according to Piaget) but they cannot change the order of constructions. There are also untimely pedagogical interventions such as teaching children to count before they have any concept of number.

Piaget (1970) identifies three principal periods of development: a) a sensorimotor period, which lasts until approximately 1.5 years of age; b) a period of representative intelligence which leads to concrete operations, running from 1.5-2 years to 11 years; c) finally, a period of propositional or formal operations, starting at about 11 years. There are, however, various sub-divisions, each of which is identified by particular cognitive characteristics (Figure 6-1).

Sensorimotor Phase

During the initial eighteen months of life the child begins by exercising its inborn reflexes and eventually co-ordinates its actions and perceptions to form acquired adaptations. In later life, in the school, the child will deal with his environment chiefly through his thoughts; the baby deals with it through its senses and motor activity. It is a period of pre-linguistic intelligence which allows, even without the use of language, certain rudimentary constructions, such as learning that objects are conserved and the development of notions of space and time. It is this period of child functioning that enables Piaget to place such emphasis on actions as the roots of logical operations.

> Turning to the question of the relations between language and logical operations, we have always maintained that the origin of logical operations is both deeper than and genetically prior to language; that is, it lies in the laws of the general co-ordinations of action, which control all activities including language itself. (Piaget, 1970, p.722)

Pre-operational Phase

This phase is divided into two stages. During the first, the pre-conceptual stage, the child represents one thing by another, plays games of pretence, plays some constructional games, draws and talks, learns letters and may read and write, but forms no true concepts and is considered to be egocentric. The child makes its first attempts at generalization at this stage, forming 'pre-concepts', which fall somewhere between mental images and true concepts (Pulaski, 1980). A pre-concept reveals the confusion between a class of objects and its members: one man is Daddy, so all men are daddies; one slug is much like the next, so the child will say 'There's the slug' on seeing one, and 'There's the slug again' on seeing another a few yards further on.

The second stage, the intuitive stage, starts at about four years. The child begins to form concepts but only sees one relationship in complex arrays and is dominated by perceptual judgements. The intuitive stage is so-called because it lacks the stable, reversible character of true operational thinking. It is typified by non-conservation in one of Piaget's most famous experiments. One of two glasses is partially filled with a liquid, and the second glass is then filled up to the same level. The child is asked if there is the same amount in both glasses, and adjustments are made until the child agrees that there is the same amount. The liquid from one of the glasses is now poured into another vessel, which is narrower and taller. To the question 'Is there more, less or the same amount?' in the tall, thin vessel, the child will reply that there is more, justifying the answer by focusing on a single dimension, even when the liquid is returned to the original glass.

During the pre-operational phase the child focuses his attention on the separate states or configurations of an array rather than on the transformations by which one state is transformed into another and such mental images as exist move slowly from one static configuration to the next, with no reversibility or flexibility, rather like slides in a projector. These are the figurative aspects of thought, in contrast to the operative aspects which are dynamic, related to transformations and external or internal actions.

Concrete Operations Phase

During the period from seven to eleven years physical actions are replaced by mental operations, the child forms classes and series but still works by trial and error, does not plan ahead and has difficulty with verbal definitions.

The child's mind at this stage is still tied to concrete experiences, even though hypotheses may be formed which go beyond the concrete evidence before them. Conservation, first of substance, followed by weight and then volume, is crucial to this stage. This requires an operational knowledge of reversal and involves the crucial process of decentring.

The child is capable of forming consistent classifications as concepts, rather than the pre-concepts formed by transductive reasoning in the previous stage. Related to classification is the ability to order similar objects according to size or position, a process known as seriation.

There are limitations in verbal reasoning, and when asked to use verbal propositions instead of objects, children at this stage tend to consider one statement at a time, just as children in the intuitive stage consider one relationship at a time when dealing with objects. The result is that concrete operational children have difficulty with such problems as: Edith is fairer than Susan; Edith is darker than Lily; who is the darkest of the three? They reason that Edith and Susan are fair, Edith and Lily are dark, therefore Lily is darkest, Susan is fairest, with Edith in between.

Related to the difficulties experienced with interiorized verbal reasoning is the difficulty in planning, especially in scientific experiments where the child tends to work by trial and error, rather than in a systematic manner.

Formal Operations Phase

The highest levels of thinking found in this phase are only possible when the child can follow the form of an argument without having the concrete materials at hand. This is the stage at which children begin to reason hypothetically, in the absence of material evidence. Hypotheses can then be tested to determine real solutions amongst possible solutions, a process known as hypothetico-deductive reasoning.

There is no higher phase than formal operations, and this represents the peak of human cognitive development, the functioning of the mature organism. Although it may represent the final phase, it does not follow that all adults actually achieve the full potential of this phase, indeed, many adults may remain in the previous stage and fail to show those characteristics associated with formal operations: reasoning increasingly in verbal propositions, making hypotheses which they test, planning ahead, seeking to explain events, considering many possibilities and looking for general laws.

PIAGET AND EDUCATION

Active Methods

In 'The Science of Education and the Psychology of the Child', Piaget acknowledged that direct instruction and imposed labour are a necessary part of education but he emphasized that educational programmes should 'give rise to active manipulation and discovery by the child itself.' Piaget is saying that basic operational structures of intelligence are not acquired through instruction but must be invented by the child. Furthermore, the child must reach an appropriate stage of maturational and intellectual readiness before she will be able to develop new structures of thought. The goal for education, according to Piaget, is to produce individuals who are critical, creative and inventive discoverers. So a major part of the child's learning depends on experimentation and discovery.

Piaget states that an active school is not necessarily a school of manual labour and, although a child's activity at certain levels necessarily entails the manipulation of objects, at other levels the most authentic research activity may take place in the spheres of reflection requiring the most advanced forms of abstraction or verbal manipulations.

A criticism of such methods has been that they inevitably lead to anarchic individualism, but Piaget counters this by suggesting that far from this, they actually help to combine individual work and team work, and encourage self-discipline and voluntary effort.

At the time of writing 'The Science of Education and

Psychology of the Child' (1971) Piaget felt that although active methods were gradually becoming accepted, no great progress had been made because of the difficulty of putting them into practice when compared with receptive methods. Active methods require much more varied and concentrated work from the teacher, whereas giving lessons is less demanding and tiring, and corresponds to a much more natural tendency in the adult generally, and in the adult pedagogue in particular.

Also, an active pedagogy presupposes a much more advanced kind of training and without an adequate knowledge of child psychology the teacher cannot properly understand the students' spontaneous procedures and therefore fails to take advantage of reactions that to him seem quite insignificant and a mere waste of time. Piaget claims that 'the heart-breaking difficulty in pedagogy... is, in fact, that the best methods are also the most difficult ones' (1971, p.69).

The active classroom has been associated with the term progressive teaching, which has pupils in an active role, learning predominantly by discovery techniques, with an accent on creative expression. Subject matter tends to be integrated, with the teacher acting as a guide to educational experiences and encouraging co-operative work. External rewards and punishments are seen as being unnecessary, and there is not so much concern with traditional academic standards and testing.

Receptive Methods

Piaget compares active methods with receptive methods, which are also referred to as traditional methods of teaching, in which the child is the passive receiver of information transmitted by the teacher. The danger with such methods, according to Piaget, is that the resulting knowledge will frequently be rote learned and not fully understood. In other words, receptive teaching methods lead to a predominance of figurative knowledge rather than the higher goal of operative knowledge.

Receptive methods place greater emphasis on the creative role of the adult, with a consequent emphasis on the transmission of knowledge by the teacher, rather than the constructive role of action which leads to placing an essential share of importance upon the activities of the student.

The accent is on memory, practice and rote learning, with a concern for regular testing and competitive, academic standards. The pupils are essentially passive and there is little emphasis on their creative expression.

Intuitive or Audio-visual Methods

Periodically emphasis has been placed on the role of audio-visual methods, according to Piaget. He uses the term 'intuitive' for such teaching methods, equating the function of audio-visual aids with the 'intuitive stage' in his developmental scheme, a period before operative thought is established, and during which figurative aspects of knowledge predominate. He accuses well-intentioned pedagogues of using such aids and believing that they have reached the summit of educational progress when, in fact, they are multiplying intuitive figurations in forms that no longer have anything active in them.

Piaget quotes an experiment which illustrates the dangers of these 'intuitive figurations.' He compared the memories that groups of children retain of a grouping of cubes, according to whether the cubes had been:

1. simply looked at
2. reconstructed by the child itself
3. constructed by an adult while the child watched.

Piaget found that the direct experience is superior (Piaget, 1971, p.35), and the demonstration by an adult is no better than the simple observation of the completed grouping. Thus, Piaget concludes, by carrying out experiments in the child's presence, either as demonstrations or on film, instead of making him carry them out, one loses the entire informational and formative value offered by direct action.

Piaget goes on to suggest that an additional reason why active methods have not been taken up is the confusion that sometimes occurs between active and intuitive methods: that they are either equivalent or at least the intuitive methods produce all the benefits that can be derived from the active methods. This arises from two confusions:

1. that any activity on the part of the student or child is a matter of physical actions. This is only true at the elementary levels, not in the later stages when a student may be totally active in the sense of making a personal rediscovery of the truths to be acquired, even though his activity is being directed toward interior and abstract reflection
2. a belief that an activity dealing with concrete objects is no more than a figurative process, in other words nothing but a way of producing a sort of precise copy in perceptions or mental images of the objects in question.

In other words, some audio-visual theorists believe that a mechanical copy of an object or situation, as in a film or

still picture, enables the child to gain knowledge of the object or situation. This is not so, as the experiment with the blocks revealed, knowledge is not at all the same thing as making a figurative copy of reality for oneself, it invariably consists of operative processes leading to a transformation of reality, ie. knowledge goes beyond the surface appearances.

Piaget is complaining about the rote-learning aspect of education, whereby the student makes a figurative mental copy of, for example, the stages of Piagetian theory, which are then duplicated in the examination or essay, without attempting to fit this information into previously acquired knowledge and experience in order to transform the old and new information in the process of thinking which leads ultimately to understanding.

Thus, intuitive methods merely provide students with speaking visual representation either of objects or events themselves or of the result of possible operations, but without leading to any effective realization of those operations. Piaget says that these methods are traditional methods which are continually reborn from their own ashes.

He concedes that they do certainly constitute an advance in relation to purely verbal or formal techniques, but they are totally inadequate in developing the child's operative activities and it is only as a result of a simple confusion between figurative and operative aspects of thought that it has been believed possible to pay tribute to the ideal of the active methods while at the same time giving concrete form to the subject matter of education in this purely figurative guise.

He relates this to the use of Cuisenaire rods, which are capable of being used in diametrically opposed ways:

1. a genuinely operative method, in which the child spontaneously manipulates the rods and forms configurations which lead to true operative understanding
2. intuitive or figurative methods, in which the teacher demonstrates and explains configurations laid out by the teacher rather than the child.

He also describes the use of a film to demonstrate geometrical concepts, including an illustration of Pythagoras's theorem in which the 'relations involved acquired visual clarity worthy of the highest praise.' But, Piaget asks 'Is this really training the child in geometrical reasoning and in operative construction in general?'

Pedagogy based on the image, even with the apparent dynamism of the film, remains inadequate for the training of operational constructivism, because intelligence cannot be reduced to a series of images on film. Piaget suggests that it might more correctly be compared to a series of cybernetic mechanisms ensuring a continuous flow of images by dint of auto-regulating and auto-correcting processes. This in turn leads to the conclusion that computer simulations with interaction between student and the machine may be more acceptable to Piaget.

Programmed Methods

Piaget's views of the figurative, rote-learning aspect of education are developed more fully in his criticisms of behaviouristic psychology, particularly as applied as an educational psychology. Piaget has this to say about behaviourism:

> This empty box conception of the organism, as it has been called, thus deliberately thumbs its nose at all kind of mental life, human or animal, and confines itself solely to behaviour in its most material aspects, ignoring any possible search for explanations in order to concentrate exclusively on the broad laws revealed by scrupulously detailed experimentation.
>
> This being so, Skinner, already in possession of the laws of learning he had personally verified or evolved, and freed of any theoretical preoccupations that might have hampered his testing of their general or practical application, observed in the first place that his experiments always worked much better when the interventions of the human experimenter were replaced by efficient mechanical apparatus. In other words, the pigeons produced much more regular reactions when dealing with teaching machines capable of applying the stimuli with greater precision and fewer minute variations. Skinner, a teacher by profession as well as a learning theorist, then had the brilliant idea that this observation of his would be equally valid when applied to humans, and that teaching machines, provided they were sufficiently well-programmed, would produce better results than an oral method of teaching susceptible of great variation in its application. And since the empty box conception of the organism renders so many preliminary considerations of the internal factors of human learning unnecessary, it was sufficient to be familiar with the general laws of learning and with the subject matter of the branches of knowledge to be taught, in order to construct programs at least equal in content to the body of knowledge commonly acquired.
>
> The experiment was tried and proved a total success. And it goes without saying, if we confine ourselves to

the usual methods of teaching by means of verbal transmission and receptive processes, that it could not fail to succeed. (Piaget, 1971, p.76-77)

Piaget does have some good words for Skinner's machines because they 'provide evidence of good psychology in that they make use exclusively of positive reinforcement and dispense totally with negative sanctions or punishments.' Further, he is not saddened by the fact that school teachers can be replaced by machines because this demonstrates the mechanical nature of traditional teaching methods, which merely require the repetition of what has been transmitted. The machines may even demonstrate the failure of traditional methods to accomplish their rather limited aims.

Piaget also appreciates the greater intensity of motivation found with machines than that found in many traditional lessons and questions whether the·teacher's affectivity always performs a desirable function. He is, in fact, much more positive about teaching machines than may be anticipated from some of his earlier remarks:

> In a time of great increase in student population and the scarcity of teachers they (teaching machines) are able to render undeniable services, and, in general, save a great deal of time in comparison with traditional methods of teaching. They are used not only in schools but also in commercial concerns where, for one reason or another, a necessity exists for rapid instruction of adults.
>
> Generally speaking, since every discipline must include a certain body of acquired facts as well as the possibility of giving rise to numerous research activities and activities of re-discovery, it is possible for a balance to be struck, varying from subject to subject, between the different parts to be played by memorizing and free activity. In which case, it is possible that the use of teaching machines will save time that would have been needlessly wasted by more traditional methods and therefore augment the number of hours available for active work. So that, particularly if the periods of active work include team work, with all that such work entails in the way of mutual incentives and checks, while the machine pre-supposes an essentially individualized kind of work, then this balance would at the same time be realizing yet another necessary kind of balance: that between the collective and individual aspects of intellectual effort, both so essential to harmonious school life. (Piaget, 1971, p.78-79)

Thus, according to Piaget, there is room for programmed

instruction, provided it is realized that it will be useful only in terms of acquiring facts (in a figurative manner). There must also be opportunities for activities and particularly manipulations which will lead to acts of discovery or re-discovery for the development of operational knowledge of a subject. But, such activities should not be misconstrued as being related to the use of audio-visual aids, which are perceptually dominated intuitive methods, producing figurative knowledge rather than true operational processes.

Research and Evaluation of Piagetian Theory

Piaget's contribution to cognitive psychology offers a considerable challenge to educational technology, because it personifies an alternative 'world-view' of educational matters. The emphasis is not on mechanical copies of reality as the basis of knowledge and understanding, but on active, or rather inter-active, processes of construction. Hence the choice of his 'cybernetic' analogy for intelligence in opposition to the mechanical film analogy. Piaget's view offers, in many ways, a richer framework for viewing cognition because it is based on a more complex model, which in turn has benefitted from richer conceptualizations and analogies than those available to the early behaviourists.

There have, however, been many criticisms of Piaget's theories and their educational applications, and it is necessary to take these into account before reaching a final evaluation of the potency of Piagetian ideas for educational technology.

It is necessary to point out that the age ranges given for the various stages are subject to substantial variation, predominant culture having a particularly strong effect. Shayer's (1978) investigation demonstrated the degree to which pupils in London comprehensives attained formal operations according to Piaget's time-scale. In fact, only 5.6% of 11 to 12-year-olds had reached the initial stages of formal operations, although many of the tasks demanded of students in the science classes were based on the assumption that a majority of pupils were in early formal operations, and could only be completed satisfactorily by the tiny minority of formal operational thinkers in this age range.

Clearly, the physical manipulation of objects and concrete experiences made available throughout the period of concrete operations, should also be available beyond the early years of secondary schooling and may even be required throughout the final years of schooling for many students.

There is some evidence to support the effects of manipulation rather than mere observation of manipulation by a teacher or instructor. For example, Macbeth (1974) found

that the direct first-hand manipulative experiences in the development of process skills in elementary school science supported the theoretical literature that pre-operational children must operate on concrete objects, though such manipulations may be more important for younger rather than older elementary school children.

There is also some evidence in secondary school science provided by Gabel and Sherwood (1980). In this case students who manipulated molecular models made greater progress than those merely observing teacher manipulation of the models.

However, although there is some support for Piaget's emphasis on the need for manipulation and active involvement in learning, his ideas concerning heuristics or discovery learning have received little empirical support.

Certainly, advocates of discovery approaches, such as Shulman and Keislar (1966), have claimed that discovery motivates more effectively than receptive 'learning approaches, and Suchman (1961) claimed that:

> The need for improvement is great. Current educational practice tends to make children less autonomous and less empirical in their search for understanding as they move up the elementary grades. The schools must have a new pedagogy with a new set of goals which subordinate retention to thinking. (Suchman, 1961, p.151)

Whereas others, such as Skinner (1968), have pointed out that discovery learning is highly inefficient because it involves a relatively high degree of trial and error responding. Wittrock's review of research (1966) offers no conclusive evidence for assuming any advantage for discovery approaches. He suggests that as a way to learn a few specific associations, discovery learning may be inferior to more highly directed procedures, but with the learning of concepts and hierarchically ordered subject matter, discovery may fare better, as Gagné and Brown (1961) had found.

Brainerd (1977) in a sustained critique of Piagetian theory reviewed alternative learning approaches to the acquisition of conservation. According to Piaget teaching children concepts that have not spontaneously developed is completely useless, and he asks, is it good to accelerate the acquisition of these concepts anyway? Brainerd accuses Piaget of being transparently Rousseauian and suggests that Piaget's proposal 'all but trivializes learning research.' Further, he feels that Piaget's collaborators merely aim to confirm existing Piagetian laws, rather than the discovery of new ones and this makes it difficult to become enthusiastic about experiments 'whose horizons are so limited.' What Brainerd calls 'tutorial methods', which includes all familiar procedures from classical learning

theory and social learning theory, were universally condemned by Sinclair (1973) and early evidence (Smedslund, 1961; Wohlwill and Lowe, 1962) certainly showed no evidence for learning effects on conservation tasks. However, other methods have been devised, including simple correction, rule learning, observational learning and conformity training, all of which yield positive results. Conformity training places a non-conserving child with a group who are conservers. The group is asked to observe a conservation experiment and to discuss each question and comment on the result, but they are restricted to formulating a consensual answer. Murray (1972) the originator of the technique found that 80% of pre-test non-conservers learned five conservation concepts and transfered these to two untrained conservation concepts. More recently Murray (1983) has commented on the fact that although shifts have generally been in the direction of conservation, there is no reason why the shift should not be in the other direction, according to social learning theory. However, although small numbers of conservers yield their position in the conformity situation, none revert to non-conservation on further tests, which means that there must be a powerful interactive effect between organismic maturation and the learning conditions.

There is no easy resolution of the two competing approaches because, as Murray (1983) concludes, 'each debater sees the child's deficiency or errors as due to a factor the other sees as tautological and non-falsifiable.'

On a different scale the Piagetian ideals of the active, discovery approach to teaching, have had a turbulent history when comparisons have been made with more traditional approaches. But, once again, there are different expectations associated with the two approaches. Lavatelli's 'Early Childhood Curriculum' (1970) concentrated on many concrete-operational concepts but did not show that a Piagetian programme produced a deeper understanding than any other system. Weikert (1973) found similar results, which were repeated in the purest of the Piagetian programs, the University of Wisconsin's 'Piagetian Preschool Education Programme' (Bingham-Newman, 1974).

Results from comparisons of approaches adopting many of the features associated with Piagetian methods have also been disappointing. Bennett (1976) ranged teacher styles along a twelve point continuum, from informal/progressive to formal/traditional, and identified a large group of teachers who closely matched the characteristics of the three major groupings: formal, mixed and informal. Data were collected for pupil performance at the start of the school year and then one year later on standard tests for reading, maths and English, together with measures for creative writing, punctuation and spelling, pupil behaviour and pupil personality.

The results indicated that pupils taught by formal and mixed styles show significantly superior progress as compared with those taught by informal styles, with the effect being more noticeable in average and above average achievers. In maths the formal style was superior to both mixed and informal, with the superiority existing at almost every level; and this result was confirmed in the English test.

The most difficult result for the progressive camp was not the fact that the more formal the teaching, the more time the students spent working on the subject matter at hand, hence the better the performance in reading, maths and English, but the fact that pupils in progressive settings did not perform at a higher level in their creative writing than their formally taught colleagues. Bruner (1976) who in much of his early writing had supported the discovery approach (Bruner, 1961) felt that the results indicated that 'common sense and technical inquiry are finally catching up with the romantic excess' and that 'new studies are now pointing to the critical role of the adult tutor, in social and intellectual development.' He did, however, urge caution in interpreting the results:

The progressives' fallacy was the assumption that you arrive at the terminus immediately - self-directed from the start and not just later. Now we know better. Self-direction is wasteful if one does not know where one is going or why. But to swing back to a Prussian model of authoritarian teaching will only assure that the pendulum will keep going through its dull, historically repetitive trajectory. (Bruner, 1976, p.225)

In fact, Bennett's book encouraged a great deal of debate about the issue of progressive versus traditional teaching, and also about the actual techniques of data collection and analysis in such studies. Bennett undertook a re-analysis of the original data using more appropriate procedures (Aitken, Bennett and Hesketh, 1981) as did others (Gray and Satterly, 1981) and the results led to the conclusion that:

First, the differences between teachers within styles were far greater than the differences between styles. Thus, one found 'effective' and 'ineffective' teachers, no matter which teaching style they adopted. Second, differences between teaching styles were so small as to be overwhelmed by differences between other systematic effects. And third, the direction of the differences between teaching styles did not consistently favour more 'formal' approaches over 'informal' ones. (Gray and Satterly, 1981, p.187)

These conclusions indicate that there is no easy answer to the progressive-traditional dilemma, certainly there seems to be no good evidence to support the superiority of traditional methods in reading, English and maths, and equally no more support for activities such as creative writing in the progressive classroom.

Aspects of this controversy are now emerging in the application of computers in education, in particular in the ideas of Seymour Papert and the ways in which the computer language LOGO should be used in the classroom.

PAPERT AND LOGO

Piagetian theory continues to exert an influence on teaching and the most recent form this has taken is the computer language LOGO developed by Seymour Papert, who describes in his book 'Mindstorms' (1980) the motivation which led to its creation:

I take from Jean Piaget a model of children as builders of their own intellectual structures. Children seem innately gifted learners, acquiring long before they go to school a vast quantity of knowledge by a process I call Piagetian learning or learning without being taught. (Papert, 1980, p.7)

However, he does differ in some ways from Piaget, and takes maths as an example of a case where Piaget would explain the slower development of a particular concept by its greater complexity or formality; Papert sees the relative cultural impoverishment in those materials that would make the concept simple and concrete, as the major element. In the case of mathematics this leads to mathophobia! Papert argues that since it is easier to learn French in France, then it should be easier to learn maths in mathsland. LOGO has been developed to provide a mathsland where children encounter and solve fundamental mathematical problems while achieving simple goals like drawing a house or snail.

Papert observes that children learn to speak naturally and easily, but they do not acquire equal competence in mathematical activities. He suggests that the difference lies not in innate differences nor even in the amount of learning, but in the nature of the learning process itself: it consists of learning the manipulation of symbols in a manner alien to the child's experience, using an unsuitable technology (paper and pencil). With the right dynamic technology children can become real mathematicians, and the right technology is a computer with LOGO programming language. The computer can provide an interactive

relationship, giving immediate feedback when errors occur, leading to meaningful learning of concepts in maths and other areas.

The method advocated by Papert is based on Piaget's learning by self-discovery, LOGO is a 'tool to think with' and the natural processes of learning should not be upset by teachers taking control. He argues (Papert, 1972) that the skills acquired during the design and debugging of programmes will be generalized, eventually being used to design and debug conceptual representations of problems. Papert feels that LOGO is capable of inducing general problem-solving strategies, rather than the 'domain-specific strategies induced from examples given by the teacher' (Howe, O'Shea and Plane, 1980).

There are many different ways in which any particular medium can be used in teaching, as Piaget (1971) noted, and LOGO is no exception. The Piagetian approaches adopted by Papert are 'horticulturalist' (Brainerd, 1978), with an emphasis on total intellectual growth derived from discovery learning, rather than traditional receptive methods which can only produce specific skills. Howe and his colleagues (Howe, O'Shea and Plane, 1980) adopt a more structured approach towards the use of LOGO. Howe (1979) has described his method as learning by model building, a process which enables students to learn about the underlying structure of an entity by building a model of it. Real learning is accomplished when the modeller can define a set of general principles which model a variety of entities. The structured approach developed by Howe is based on the following guidelines:

1. The teacher should always try to build on the child's existing knowledge.
2. The use of familiar materials (toys, pictures, apparatus) as manipulable models or metaphors makes teaching and learning more manageable.
3. Classroom teaching ought to be a partnership between teacher and child, with the child being given as much responsibility as possible for choosing, formulating and solving problems within the broad aims or objectives laid down by the teacher, in contrast to the very formal, rigid control used by many teachers.
4. Feedback of information to the learner is crucial, to check what has been done and to suggest what should be done in the event of failure.

There are many conflicting reports concerning the successes or failures of LOGO, and this arises from the different approaches to the nature of evaluation. Many studies have adopted an observational or case-study approach, following the pattern set by Piaget's method for obtaining data. Papert advocates taking a small group of

pupils and working with them exclusively for the time required to test the new method, even in the face of protests from headmasters, teachers and school authorities. He claims that such a method would yield results so qualitatively different that the question 'How do you measure that?' would be meaningless, or the results would be so poor that it would be unnecessary to even ask the question. The Edinburgh group (Howe, O'Shea and Plane, 1980) disagree and have adopted a more formal assessment procedure, for as they say:

> .... we cannot ignore cost considerations, nor objections by educational authorities. In our experience, the most important constraint is the examination diet. (Howe, O'Shea and Plane, 1980, p.5)

Most of the research has concentrated on problem solving and the learning of mathematical skills, but in general results have been disappointing. Michayluk (1986) concludes that, methodology aside, there is little evidence of LOGO influencing such skills. In 1969, Feurzig taught algebra and arithmetic to 13-year-olds using LOGO, but with no overall improvement in their mathematics and Howe, O'Shea and Plane (1980) reported similar findings with bottom stream 11 to 12 year-olds, although they did find that the children became more self-confident and keener to discuss mathematics with teachers.

For many groups the lack of traditional experimental research data does not appear to cause any concern, their descriptive data leads them to the conclusion that LOGO does improve pupils' reasoning and problem-solving, and that the powerful ideas do generalize and transfer (Noss, 1983 and 1984; Hoyles, Sutherland and Evans, 1985). Hoyles and Sutherland's research indicates that LOGO can act as a catalyst for change in the classroom, but do caution that LOGO alone cannot make a poor learning situation into a good one: good organization is crucial. They conclude that LOGO is a powerful new tool for provoking reflection in pupils and that pupils learn to make use of the feedback that the computer provides. The emphasis is on the pupils having freedom to test out their conjectures and to develop their own styles of learning.

However, these are still early days for LOGO research, and it will be many years before a full appraisal will be possible.

Constructivism and Educational Technology

Structured or unstructured, there is a qualitative difference in the approach adopted by both practitioners

when compared with behaviouristic methods. The purpose of education is the development of operative mental structures, rather than mere figurative, rote repetition, with children having much greater control over their learning activities. Educational technology must respond to the challenge of investigating these methods and the role such structures play in the development of the creative, innovative aspects of intelligence.

Behaviouristic and neo-behaviouristic methods may establish optimum learning of what is already known, but it is other methods, such as those advocated by Piaget and Papert, which may ultimately lead to the qualitatively different aspects of cognition associated with the invention of the presently unknown. Following the launching of the first man-made satellite by the U.S.S.R., on 4th October 1957, it was for just these reasons that Jerome Bruner's 'The Process of Education' (1960) became a best-seller in the early 1960's, in the U.S.

REFERENCES

Aitken, M.A., Bennett, S.N. and Hesketh, J. (1981) 'Teaching styles and pupil progress: a re-analysis', British Journal of Educational Psychology, 51, 170-186.

Bennett, S.N. (1976) Teaching Styles and Pupil Progress. London: Open Books.

Bingham-Newman, A.M. (1974) 'Development of Logical Abilities in Early Childhood. A Longitudinal Comparison of the Effects of Two Pre-school Settings.' Unpublished Doctoral Dissertation. University of Wisconsin, Madison. U.S.A.

Brainerd, C.J. (1977) 'Learning research and Piagetian theory'. In L. Seigal and C.J. Brainerd (Eds) Alternatives to Piaget: Critical Essays on the Theory.

Brainerd, C.J. (1978) Piaget's Theory of Intelligence. Englewood Cliffs, NJ.: Prentice-Hall.

Broadbent, D.E. (1958) Perception and Communication. Oxford: Pergamon Press.

Bruner, J.S. (1960) The Process of Education. Cambridge Mass.: Harvard University Press.

Bruner, J.S. (1961) 'The act of discovery', Harvard Educational Review, 31(1), 21-32.

Bruner, J.S. (1966) 'Some elements of discovery'. In L.S. Shulman & E.R. Keislar (Eds) Learning by Discovery. Chicago: Rand McNally.

Bruner, J.S. (1976) 'The styles of teaching', New Society, 29th April, 223-225.

Eysenck, M.W. (1984) A Handbook of Cognitive Psychology. London: Erlbaum.

Feurzig, W., Papert, S., Bloom, M., Grant, R. and Solomon, C. (1969) Programming Languages as a Conceptual Frame work for Teaching Mathematics. Mathematics, Report 1889, Cambridge, Mass.: Bolt Beranek and Newman.

Gabel, D. and Sherwood, R. (1980) 'The effect of student manipulation of molecular models on chemistry achievement according to Piagetian level', Journal of Research in Science Teaching, 17(1), 75-81.

Gagné, R.M. (1970) The Conditions of Learning (2nd. Edition). London: Holt, Rinehart and Winston.

Gagné, R.M. and Brown, L.T. (1961) 'Some factors in the programming of conceptual learning', Journal of Experimental Psychology, 62, 313-321.

Gray, J. and Satterly, D. (1981) 'Formal or informal? A re-assessment of the evidence', British Journal of Educational Psychology, 51, 187-196.

Howe, J.A.M. (1979) 'Learning through model-building.' In D. Mitchie (Ed) Expert Systems in the Micro-electronic Age. Edinburgh: Edinburgh University Press.

Howe, J.A.M., O'Shea, T. and Plane, F. (1980) Teaching Mathematics Through LOGO Programming: An Evaluation Study. DAI Research Paper No. 115, Edinburgh University.

Hoyles, C., Sutherland, R. and Evans, J. (1985) A Preliminary Investigation of the Pupil-centred Approach to the Learning of Logo in the Secondary School Mathematics Classroom. London: University of London Institute of Education.

Jonassen, D.H. (1985) 'Learning strategies: a new educational technology', Programmed Learning And Educational Technology, 22(1), 26-34.

Lavatelli, C. (1970) Early Childhood Curriculum - a Piaget Programme. Boston: American Science and Engineering.

Macbeth, D.R. (1974) 'The extent to which pupils manipulate materials and attainment of process skills in elementary school science', Journal of Research in Science Teaching, 11(1), 45-51.

Michayluk, J.O. (1986) 'LOGO: more than a decade later', British Journal of Educational Technology, 17(1), 35-41.

Murray, F.B. (1972) 'The acquisition of conservation through social interaction', Developmental Psychology, 6, 1-6.

Murray, F.B. (1983) 'Learning and development through social interaction and conflict: a challenge to social learning theory.' In L.S. Liben (Ed) Piaget and the Foundations of Knowledge. New Jersey: Lawrence Erlbaum.

Neisser, U. (1967) Cognitive Science. New York: Appleton-Century- Crofts.

Neisser, U. (1976) Cognition and Reality. San Francisco: Freeman.

Noss, R. (1983) Starting Logo: Interim Report of the Chiltern MEP Logo Project. Hatfield: MEP.

Noss, R. (1984) Children Learning Logo Programming: Interim

Report No.2. Hatfield: MEP.

Papert, S. (1972) 'Teaching children thinking', Programmed Learning and Educational Technology, 9, 245-255.

Papert, S. (1980) Mindstorms. Brighton: The Harvester Press.

Piaget, J. (1970) 'Piaget's theory'. In P.H. Mussen (Ed) Manual of Child Psychology, vol.1, 3rd Edition. London: John Wiley.

Piaget, J. (1971a) The Science of Education and the Psychology of the Child. London: Longmans.

Piaget, J. (1971) Biology and Knowledge. Chicago: Chicago University Press.

Pulaski, M. A. (1980) Understanding Piaget. London: Harper and Row

Shayer, M. (1978) 'The analysis of science curricula for Piagetian level of demand', Studies in Science Education, 5, 115-130.

Shulman, L.S. and Keislar, E.R. (1966) Learning by Discovery: A Critical Appraisal. Chicago: Rand-McNally

Sinclair, H. (1973) 'Recent Piagetian research in learning studies.' In M. Schwebel and J. Raph (Eds) Piaget in the Classroom. New York: Basic Books.

Skinner, B.F. (1968) The Technology of Teaching. London: Appleton-Century-Croft.

Smedslund, J. (1961) 'The acquisition of conservation of substance and weight in children. II. External reinforcement of conservation of weight and the operations of addition and subtraction', Scandanavian Journal of Psychology, 2, 71-84.

Suchman, J.R. (1961) 'Inquiry training: building skills for autonomous discovery', Merrill-Palmer Quarterly Behaviour and Development, 7, 147-169.

Waddington, C.H. (1957) The Strategy of the Genes. London: Allen and Unwin.

Weikert, D.P. (1973) Development of Effective Pre-school Programs: A Report on the High/Scope- Ypsilanti Pre-school Projects. Paper presented at the High/Scope Educational Research Foundation Conference, Ann Arbour, Michigan, USA.

Wittrock, M.C. (1966) 'The learning by discovery hypothesis.' In L.S. Shulman and E.R. Keislar (Eds) Learning by Discovery: A Critical Appraisal. Chicago: Rand-McNally.

Wohlwill, J.F. and Lowe, R.C. (1962) 'An experimental analysis of the conservation of number', Child Development, 33, 153-167.

CHAPTER 7

THOUGHT, CULTURE AND COGNITIVE MODELS

Piaget's theory has recently been criticized by Wartofsky (1983) for failing to adapt to new information and ideas in the field of cognitive development: there has been too much assimilation and not enough accommodation. This amounts to a criticism that the theory is no longer capable of growth. Wartofsky calls for a historical epistemology, rather than a genetic epistemology, because phylogenesis, in cognitive terms, can no longer be seen to be biologically fixed. It is historical, cultural and it continues. From this it follows that ontogenesis is not the socialization of child-thought to a fixed adult world, but to a changing world. Thus, Piaget cannot determine those universal and necessary features of the growth of knowledge, because this assumes a fixed species as the norm. What is needed, according to Wartofsky, is an evolutionary theory of the history of cognition. Implicit in this is the need for Piagetian theory to accommodate to new ideas and theories in order to develop and progress, and there is an implied criticism that this has not occurred.

Piagetian theory exerted a great influence on the early work of the American cognitive psychologist Jerome S. Bruner, following his visit to Geneva in 1956. However, Bruner has shown himself to be more receptive than his mentor to new ideas which have been in conflict with Piagetian theory. He has continued to admit his mistaken assumptions (Bruner, 1983) and press on with an evolving view of human development. Central to this is his use of the Lamarkian metaphor:

> Assume, for example, that man continues to adjust when he learns a language and certain ways of using tools. At that particular point evolution becomes Lamarkian in the sense of involving the passing on of acquired characteristics, not through the genes, but through the medium of culture. (Bruner, 1966a, p.101)

For Bruner culture is essential for humankind because it is a means of transmitting the accumulated knowledge of earlier generations. Each generation does not have to re-discover the past. It does, however, bring with it a certain vulnerability, in that it is reversible and may be forgotten as has happened in the Easter Islands and the Inca civilization of Peru.

Bruner's increasing emphasis on cultural aspects of

cognition, and in particular the amplification of cognition afforded by language, began with his contact with the works of the Soviet psychologist Vygotsky, who viewed intelligence as the capacity to benefit from instruction, with language having a powerful developmental role.

However, Bruner is still probably best known for his summary of the Woods Hole conference convened in 1959 by the National Academy of Sciences. Thirty-five 'scientists, scholars and educators' discussed ways in which education in science could be improved in primary and secondary schools, and the report of their deliberations was 'The Process of Education.'

The Process of Education

Paul Goodman announced in the N.Y. Herald Tribune that 'The Process of Education' (1960) was a classic comparable in its philosophical centrality and humane concreteness to Dewey's essays on education.

Bruner set forth a strong argument favouring the idea of models in the head based on general understanding, from which hypotheses about the particular could be generated and then tested against experience. According to Bruner the great disciplines (maths, physics, history) are not so much repositories of knowledge, they are methods for the use of mind. They provide the structure that gives meaning to the particulars, and the object of education is to get as swiftly as possible to that structure - to penetrate a subject, not to cover it. This is done by spiraling into it: a first pass to get the intuitive sense of it, later passes over the same domain to go more deeply into it.

Bruner complained that relatively little work by American psychologists had been done on the manner in which students could be trained to grasp the underlying structure of complex knowledge. He mentioned the latest researches which demonstrated that 'massive general transfer can be achieved by appropriate learning.' These studies were concerned with learning designed to produce general understanding of the structure of a subject, and as an example he considers mathematics:

> .... algebra is a way of arranging knowns and unknowns in equations so that the unknonwns are made knowable. The three fundamentals involved in working with these equations are commutation, distribution, and association. Once a student grasps the ideas embodied by these three fundamentals, he is in a position to recognize wherein 'new' equations to be solved are not

new at all, but variants on a familiar theme. (Bruner, 1960, p.7)

This reflected the emphasis from the cognitive revolution of the 1950's concerning the generativeness of knowledge: knowledge is not a store-house. Learning is most often figuring out how to use what is already known in order to go beyond what is currently being thought - and this involves knowing something structural about what is being contemplated - how it is put together. As Bruner says 'Knowing how something is put together is worth a 1,000 facts about it. It permits you to go beyond it.' (Bruner, 1983, p.183)

Perhaps the most controversial statement in the book was that any subject can be taught to anybody at any age in some form that is honest. Bruner has since complained that this is easy to take too literally; he was simply arguing that there are many intuitive notions in a range of subjects (calculus, the theory of evolution etc.) whose early grasp would help learning of the later, fully developed idea.

This hypothesis was based on three general ideas. From Piaget he took the idea that the child's understanding of any mathematical, scientific or moral idea would be framed by the level of intellectual operations that he had achieved. Lower levels of understanding are routes to higher-level ones. The lower level is not a degraded version of the higher one - each has a logic of its own.

In addition to the Piagetian perspective Bruner suggested that the act of learning appeared to consist of three simultaneous processes: acquisition of new information; transformation of information to make it fit new tasks; and evaluation or checking whether the manipulated information is adequate to the task.

Finally, there was the idea of the 'spiral curriculum'. A curriculum ought to be built around the great issues, principles, and values that a society sees as worthy of continual relevance for its members, according to Bruner, and he demonstrates how this may be applied to science:

> So too in science. If understanding of number, measure and probability is judged crucial in the pursuit of science, then instruction in these subjects should begin as intellectually honestly and as early as possible in a manner consistent with the child's forms of thought. Let the topics be developed and redeveloped in later grades. Thus, if most children are to take a tenth-grade unit in biology, need they approach the subject cold? Is it not possible, with a minimum of formal laboratory work if necessary, to introduce them to some of the major biological ideas earlier, in a spirit perhaps less exact and more intuitive. (Bruner, 1960, p.54)

These ideas reflected the intellectual ferment of the times, particularly the structuralist influences of Chomsky, Levi-Strauss and Piaget, and there were many different reactions to such ideas. In Russia the book tripped off a debate on the dogmatism of Russian education; in Italy it was used to attack the decaying formalism of the right and the utilitarian pragmatism of the left. In Japan it caught a wave of reform and became an emblem against traditional learning by rote.

Modes of Representation

Following on from the success of 'The Process of Education' Bruner became concerned with the techniques and technologies that aid growing human beings to represent in a manageable way the recurrent features of the complex world in which they live. He noted that the principal change in man during the past half-million years has been by linking himself with new external implementation systems rather than by any conspicuous change in morphology - 'evolution by prosthesis' as the anthropologist Weston La Barre put it. Bruner quoted Washburn and Howell (1960) as evidence for this:

It would appear that the large size of the brain of certain hominids was a relatively late development and that the brain evolved due to new selection pressures after bipedalism and consequent upon the use of tools.... (the) size of brain has increased some three-fold subsequent to the use and manufacture of implements. (Washburne and Howell, 1960, p.49)

Bruner suggested that the 'implement' systems have been of three main kinds:

1) Amplifiers of human motor capacities (levers, wheels, cutting tools).

2) Amplifiers of human sensory capacities (primitive devices such as smoke signals and modern ones such as radar).

3) Amplifiers of human ratiocinative (logical thought) capacities (language systems, explanatory theories).

He stated that these are conventionalized and transmitted by the culture, particularly the ratiocinative amplifiers, which involve symbol systems governed by rules that must be shared. He also suggested that there are three systems for processing information allowing human beings to construct what he calls 'models' of their worlds, which emerge in the developing child in the following order:

1) through action
2) through imagery
3) through symbols and language

These three modes of internal representation are termed:

1) through action.........................ENACTIVE
2) through imagery.........................ICONIC
3) through symbols and language...........SYMBOLIC

Initially, Bruner saw the enactive phase running from very early in life, with language being superimposed at around 18 months and imagery declining at the age of 6 or 7 years: first comes enactive, then iconic and finally symbolic forms of representation. Bruner acknowledges in his autobiography 'In Search Of Mind' (1983) that he fell into the Genevan trap of postulating a chronological straight jacket. He now feels that these modes are present throughout life and are partially translatable into one another.

Enactive Mode:

The enactive mode of representation is highly manipulative in character. It is knowing some aspect of reality without the use of imagery or words. Hence, it consists of representing past events through making appropriate motor responses. It consists mainly of knowing how to do something; it consists of a series of actions that are appropriate for achieving some result eg. sailing a boat, tying a knot, riding a bike.

Iconic Mode:

This is based on internal imagery. The knowledge is represented by a set of images that stand for the concept. Iconic representation depends upon visual or other sensory organisation and is principally defined by perceptual organisation and techniques for economically transforming perceptions. Although initially seen as fading from use in the child's cognitive apparatus at the age of 6 or 7 years, it is now recognized as an important element in the highest intellectual realms. Consider what Einstein had to say about his thought processes:

The words or language, as they are written or spoken, do not seem to play any role in my mechanisms of thought. The psychical entities which seem to serve as elements in thought are certain signs and more or less clear

images which can be voluntarily reproduced or combined.... The above mentioned elements are, in my case, of visual and some of muscular type. Conventional words or other signs have to be sought for laboriously only in some secondary stage... (J. Hadamard, 1945)

We have a similar account of Kekule's discovery of the benzene ring in a dream:

Again atoms were gamboling before my eyes. Smaller groups kept to the background. My mind's eye trained by repeated visions of a similar kind, now distinguished larger formations of various shapes.... everything in movement winding and turning like snakes. And look, what was that? One snake grabbed its own tail, and mockingly the shape whirled before my eyes. As if struck by lightning I awoke. (Anschutz, R., 1961, p.780)

And there are other examples, such as Watson and Crick's discovery of the DNA helix (J.D. Watson, 1968). Bruner is suggesting that images stand for perceptual events in the same way that a picture stands for the object pictured. In many circumstances such a representation of the world has many advantages, but, as we shall see, there are also disadvantages associated with predominantly iconic forms of representation.

Symbolic Mode:

Increasingly throughout life there is recourse to the symbolic mode of representation of thought. The emphasis here is that representation is based upon an abstract, arbitrary and flexible system of thought. It enables individuals to deal with what might be and what might not, and is a major tool in reflective thinking, though as we have seen not the only one. This mode is representative of a person's ability to consider propositions rather than objects (as we have seen in the stage of formal operations), to give concepts a hierarchical structure and to consider alternative possibilities in a combinatorial fashion.

Bruner gives the following example to distinguish between the three modes and to show how there is a natural progression from enactive through iconic to symbolic modes of thought:

Any domain of knowledge (or any problem within that domain of knowledge) can be represented in three ways: by a set of actions appropriate for achieving a certain result (enactive representation); by a set of summary images or graphics that stand for a concept without defining it fully (iconic representation); and

by a set of symbolic or logical propositions drawn from a symbolic system that is governed by rules or laws forming and transforming propositions (symbolic representation). The distinction can most conveniently be made concretely in terms of a balance beam.... A quite young child can plainly act on the 'principles' of a balance beam, and indicates that he can do so by being able to handle himself on a see-saw. He knows that to get his side down farther he has to move out farther from the centre. A somewhat older child can represent the balance beam to himself either by a model on which rings can be hung and balanced or by a drawing. The image of the balance beam can be varyingly refined, with fewer and fewer irrelevant details present, as in the typical diagrams in an introductory textbook in physics. Finally, a balance beam can be described in ordinary English, without diagrammatic aids, or it can be even better described mathematically by reference to Newton's Law of Moments in inertial physics. (Bruner, 1966b, p.45)

Dale (1969) found in Bruner's ideas a conceptual framework which related to the three main divisions of his 'Cone of Experience' in the fifth edition of 'Audio-visual Methods in Teaching.' He suggested that our experiences vary according to the degree in which they involve us physically or in thought. Some experiences call for a good deal of concrete, direct, immediate action in which we make full use of our senses and often our muscles as well. Observing something, however, requires less physical, or concrete, action than an experience of doing. And symbolic experiences have virtually all the manifest physical action removed. Dale agreed with Bruner that we begin our learning of a specific matter with a broad base of direct experience in action and gradually we omit these specific, first-hand, concrete occurrences and impressions as we come to rely on iconic substitutes or pictorial representations. At both stages, we develop a summarizing idea or symbol and when we understand a symbol we can use a word or formula to stand for everything out of which it developed.

This visual analogy (the Cone of Experience) is one of several devices that have been set up to show the progression of learning experiences from direct, first-hand participation to pictorial representation and on to purely abstract, symbolic expression.... the threefold arrangement of learning possibilities illustrates the three kinds of experience that we have found so important in the process of complete communication. In this respect it is similar to the analysis of the three major modes of learning made by Jerome Bruner. (Dale, 1969, p.108)

Thought, Language and the Zone of Proximal Development

Bruner went on to develop his curriculum package 'Man: A Course Of Study', which was concerned with the nature of man as a species, and the forces that shape his humanity. Five great humanizing forces, each closely associated with man's evolution as a species, were explored (Bruner, 1966b). These were tool making, language, social organization, the management of man's prolonged childhood and man's urge to explain his world and they reflected Bruner's increasing interest in Vygotsky's functionalist psychology.

Vygotsky was a Russian Jew, deeply interested in the arts and in language, and was a friend of the film-maker Eisenstein. His objective was to explore how human society provided instruments to aid the developing individual mind, and how the child clumsily takes over the forms and tools of the culture and then learns to use them appropriately. His functionalist psychology was interested in studying how the child uses hints and takes advantage of others helping him organize his thought processes until he can do so on his own. Vygotsky (1962) used the term 'Zone of Proximal Development' to describe the child's potential to use the help of others to gain consciousness and reach higher ground intellectually, transforming the meaning of the lower order concepts. Intelligence tests were inadequate measures of a child's ability, as far as Vygotsky was concerned, because they only show what she is capable of on her own. In the real world children have access to older children and adults to help them solve their problems, and so he argued that we should take into account the capacity of a child to profit from help that others can give when assessing a child's potential. The more they take advantage of an adult's support, the wider is their 'Zone of Proximal Development.' Two children may both be at the same stage when measured by conventional tests but may differ in the extent of their respective zones. With help one may manage to complete tasks usually completed independently by children four years her senior, whereas the other with similar help may only manage to extend his competence by two years, the limit of his zone. Of course, Vygotsky was writing at a time when advanced technological aids,in the form of calculators and computers, were not available to children, but there is no reason why the concept of a zone of proximal development should not be extended to include help derived from such powerful aids.

Vygotsky, working with Luria in the 1920's and early 1930's, offered a challenge to the prevailing Soviet view of psychology, which was predominantly a Pavlovian atomistic view, based on the reflex. They placed much greater emphasis on the transforming power of language and the internalized linguistic system represented in thought. Vygotsky's genius

was to exploit the distinction between classical conditioning prior to the growth of symbolic function, and thought after the intrusion of what Pavlov reluctantly termed the 'second signal system.'

He suggested that symbol systems restructure mental activity, and as symbol systems emerge in the course of history so different cognitions will be produced. Basic psychological processes (abstraction, generalization, inference) are universal and common to all humankind, but their functional organization will vary depending on the nature of the symbol systems available in different epochs. Language is a universal symbol system playing a crucial role in the development of higher psychological processes. Other symbol systems which are not universal introduce culture-specific differences. Luria (1976) led an expedition in the 1930's to test Vygotsky's theory and found that there were consitent differences between traditional non-literate farmers and other members of the same villages who had undergone brief literacy programmes. The most traditional and isolated populations with neither literacy nor schooling were context-bound, guided by perceptual and functional attributes and responded in a concrete manner to the tests. The most schooled group were able to respond in a more abstract manner and were more responsive to the conceptual and logical relationships between things. This certainly pointed to confirmation of Vygotsky's thesis that socio-cultural changes form the basis for the development of higher psychological processes, and literacy could be used as an indicator of the level of functioning. However, there were many comensurate changes at that time within Soviet society and differences in mental operations could not be attributed to literacy or schooling per se (Scribner and Cole, 1981).

This line of research has been pursued by Greenfield and Bruner (1966) who found similar results with Wolof children in Senegal. Greenfield has suggested that the differences are due to the school children's capacity for context-independent, abstract thought, and has provided a link between this and written language. Olson (1976) has also supported the theory that literacy biases cultures toward the development of formal reasoning systems. Scribner and Cole (1981) disagree having found no evidence to support the construct of a general 'literacy' phenomena. They suggest that literacies are highly differentiated and that the Arabic and Vai scripts studied did not trade off for each other in predicting cognitive performance, nor do they (singly or in combination) substitute for English literacy. Their results suggest a more general effect of the de-contextualization prevalent in the culture of Western schooling, as proposed by Donaldson (1978).

Cognition and Culture

Piaget always resisted the idea that there is a psychological reality to culture, that exists in the Popperian World 3 sense such that it can be internalized and serve as a prosthesis for the mind and development (Bruner, 1983), but for Bruner it became a crucial element in the child's development.

Bruner sees a person's maturing cognitive growth as being characterized by the increasing independence of his responses from the immediate nature of the stimuli and this growth depends upon the person internalizing events into a storage system that corresponds to aspects of the environment, by means of images, words, symbols. Through growth a person gains freedom from stimulus-control through mediating processes that transform the stimulus prior to the response.

Although Bruner sees mental growth as being like a staircase with rather sharp risers - a matter of spurts and rests, the spurts being touched off when capacities begin to develop, he differs from Piaget in that he believes that the steps or spurts are not very clearly linked to age; some environments can slow the sequence down or bring it to a halt, others move it along faster.

For example, Bruner anticipates improvement in performance in Piaget's conservation task when the environment is changed, forcing the child to activate language and take advantage of the 'remoteness of reference' that is a feature of language. In such a situation the child is forced to say his description of the world in the absence of the things to be described.

This was confirmed in Franks' perceptual shield experiment, reported in 'Studies in Cognitive Growth' (Bruner et al., 1966). Children who were not 'conservers' were shown two beakers of different dimensions, one containing liquid. A screen was placed in front of the beakers and the liquid poured from one to the other. The child was then asked 'Is there still the same amount of liquid?' In the traditional Piagetian conservation experiment children observe the pouring of the liquid from one beaker to the other and when asked if there is still the same amount of liquid they focus on the changed appearance and indicate that there is more or less depending on the dimensions of the beaker. Franks found that many children who failed on the traditional test were able to answer correctly with the shield in place. Correct responses jumped from 0% to 40% for the 4 year-old group, from 20% to 90% for 5 year-olds, and from 50% to 100% for 6 year-olds. The most common responses being 'its the same water' or 'you only poured it!' When the screen was removed all the 4-year-olds change their minds, the perceptual display overwhelmed them, but virtually all the 5-year-olds

held to their judgements, often invoking the difference between appearance and reality in their defence: 'It looks like more to drink, but it is only the same because it is the same water and it was only poured from there to there.' All 6 and 7-year-olds also held to their judgement.

When a post-test was administered with a different set of test materials to those used in the experiment it was found that in most cases the change in the child's reasoning had generalized and it was concluded that it was attributable to the powerful organizing effects of language.

The experiment reflects Bruner's view that language is a powerful ratiocinative technology which provides the child with a qualitatively different model of the world to that based purely on observation and this enables the child to predict the outcome of manipulations of the environment more accurately. Such technological advances, in a cognitive sense, promote the development of intellectual functioning, according to Bruner:

> I shall take the view in what follows that the development of human intellectual functioning from infancy to such perfection as it may reach is shaped by a series of technological advances in the use of mind. Growth depends upon the mastery of techniques and cannot be understood without reference to such mastery. These techniques.... are skills transmitted with varying efficiency and success by the culture, language being a prime example. (Bruner, 1964, p.1)

Direct and Mediated Experience

Bruner's continuing interest in the interaction between cognition and the technologies of culture led him to consider the nature of experiences which are transformed and contribute to a person's internal model of the world.

He dealt with the various forms of direct and mediated experience in the 73rd Year Book of the National Society for the Study of Education under the title 'Learning Through Experience and Learning Through Media' (Olson and Bruner, 1974). These major aspects of learning, including the three modes of internal representation, are considered in terms of their partial equivalence and substitutability, and also their differing potential roles in the intellectual development and acculturation of children.

When discussing direct experience he suggests that organisms have commerce with the environment on their own terms and actively select those features which are necessary for the construction of representations of the environment which enable survival of the individual or species. From this it follows that 'our conception of physical reality is

itself achieved by selective mediation' and it is the activities engaged in which determine the nature of such models of reality, as was suggested by Piaget (1971, Biology and Knowledge). Thus, we have a picture of reality that is biased or coded in terms of our actions upon it. Knowledge is always mediated or specified through some form of human activity and, according to Bruner and Olson, this activity has two facets:

1) Information about the world (knowledge)
2) Information about the activity used in gaining knowledge (skill, ability).

Further, it is suggested that there are two types of invariants that are specified through experience. Those features that are more or less invariant across different activities constitute our knowledge about those objects or events; the invariants across different objects or events constitute the basis of skills and abilities (Figure 7-1).

Figure 7-1. The Relationship between Knowledge and Skills.

Knowledge	Skill
Events	Locomotion
Objects	Linguistic
	Representation
TREE	DRAWING
BALL	THROWING
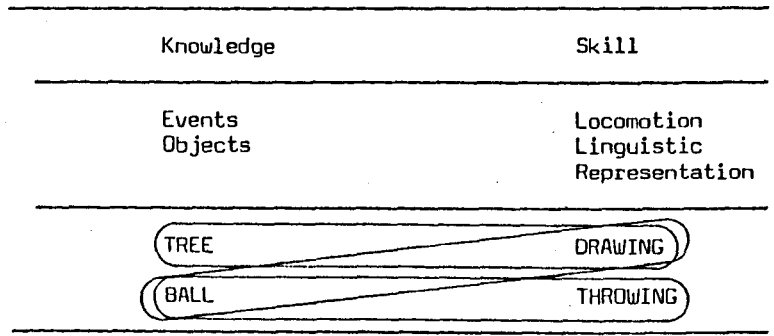
Within this context, Bruner claims that the goal or end-point is crucial in influencing the type of information to be sought and also the skills to be used in obtaining that information. He gives as an example the goal of 'learning about a country' and relates this to learning through direct experience:

The most obvious way to learn about a country (the goal, or end-point) is to walk its streets, read its poets, eat its foods, work in its fields and so on. In so doing, one will learn both about the country (that the country is poor or hilly, etc.) and how to proceed in the activities required to be of that country (how to mend a net or tell a story). This is surely what is

meant by learning through one's own direct contingent experience. (Olson and Bruner, 1974, p.130)

Vicarious Experience: Observational Learning

But there are other ways to acquire information eg. from seeing a person struggling with a load one can estimate its weight. In other words, one can experience vicariously and this facility opens up the pathways for learning from a variety of media: films and television, print and the spoken word. Such learning can occur when neither of the primary conditions for learning through contingent experience (self-initiated action or knowledge of its results) is fulfilled.

Bandura (1971) has summarized a wide range of data showing that behaviour can be changed by exposing the learner to modeling stimuli. In his classic 'Bobo doll' experiment (Bandura, 1965) children saw an inflatable plastic clown being abused in a television programme and were then allowed to play with a similar doll. The children were observed to copy the violent acts. Similar work has shown that altruistic behaviour can also be induced by observation of modelled behaviour.

Although early animal studies indicated that learning by observation was confined to higher mammals (Thorndike, 1898) later work has shown this not to be the case. Herbert and Harsh (1944), for example, had two groups of cats learn to pull strings and open doors by observing other cats. One group saw the final error-free performance, the other group saw early error-filled performance. Both groups of cats learned more quickly than the control group, which had not observed another cat's performance.

Bruner (1972) notes that while all animals learn from contingent experience, primates are distinctive in their capacity for learning by observation - there being an enormous amount of observation of adult behaviour by the young. He places great emphasis on the potential of observational learning and particularly the relaxed atmosphere of the social units of the great apes which allow the young to spend extensive periods of time observing techniques, particularly tool-using behaviour, and practising and experimenting with the component acts.

Zimmerman (1977) claims that Bandura's 'Social Learning Theory' (SLT) represents the most sophisticated formulation for human vicarious learning, in that it represents the 'wedding between behavioural and informational processing explanations for human functioning' (p.42). According to SLT observational learning is primarily a cognitive, representational process in which the representations are mentally transformed, stored either symbolically or iconically, and retrieved before being manifested as

imitation. Zimmerman argues that it is possible for teachers to use observational learning principles, as described by Bandura, to organize classroom demonstrations to teach concepts and abstract skills to children.

Bruner agrees and indicates that when observational learning is used as an instructional technique in the form of demonstrations the child should be shown critical alternatives and how to choose between them (Olson and Bruner, 1974). Good instruction through modeling depends upon the sensitivity of the instructor to the alternatives likely to be entertained by the child. And, just as providing clear demonstrations involves skill, so it seems probable that learning from demonstrations demands a skill. Modeling may well depend on the capacity not so much to imitate directly as to construct behaviour from already mastered constituent acts in order to match selected features of the model. As an instructional'technique modeling in the form of a demonstration is different from a skilled performance.

Vicarious Experience: Symbolic Learning

Bruner acknowledges that a further alternative to learning from direct experience is through symbolically coded information: words, diagrams, maps etc. It is learning through these symbolic systems that most readily substitutes for direct experience in formal schooling. Bruner and Vygotsky have both emphasised the extent to which language provides the means, par excellence, for teaching and learning out of context. Language provides an opportunity for acquiring knowledge in a form compatible with the rules of abstract thought. The human species is marked by its reliance on symbolically coded experience, so much so that language is taken as the distinctive human characteristic, and the development of literacy in various symbolic codes is the primary concern of formalized schooling. The major limitation of all cultural media such as language, numbers, graphs, diagrams etc, is that the information is conveyed through a symbolic system that places a high demand upon literacy in that medium. And the meaning extracted from those symbolic systems will be limited to the meaning acquired by the use of the symbol in the referential or experiential world. In other words, the major limitation of language is that no new information can be conveyed through language - if information falls outside the listener's competence the sentence will be interpreted in terms of knowledge already possessed. Thus, instruction through language is limited to rearranging, ordering and differentiating knowledge or information that the listener has available from other sources (modeling, direct

experience). Bruner makes the analogy here with the impossibility of verbally explaining perspective to the blind!

These different forms of experiencing a concept are seen by Bruner as being mapped on to a common underlying structure, or a coherent and generalized concept of reality and the very fact that the representations of reality are partially translatable, one to another, makes instruction possible. Thus, information relevant to action can be acquired by means other than direct action, so that we may learn to sail through watching films and reading books. But the three different forms of experience (enactive, iconic and symbolic) differ greatly in terms of the skills they assume and develop in the learner.

Forms of Instruction

Finally, Bruner suggests that there are three basic forms of instruction all of which provide information relevant to the acquisition of both knowledge and skills and these, in turn, are related to a variety of technological realizations:

1)   Systems of arranged contingent experience which are related to structured environments for learning, such as laboratories and experiments, simulations and even educational toys.

2)   Systems of observational learning which are associated with demonstrations and modeling, which may be live demonstrations or recordings on film or videotape.

3)   Symbolic systems which include such technological realizations as print, drawings, diagrams, graphs and maps.

Thus, Bruner has related our interactions with the world, in terms of the forms of mediation, to the forms of representation which allow our construction of models of the world. These in turn have been related to technological realizations and the development of skills associated with the decoding of communications.

Olson (1976) has taken this further with his theory of instructional means. He makes a distinction between 'utterance' and 'text', in which oral language is characterized as 'a flexible, unspecialized, all-purpose instrument with a low degree of conventionalization' in which meanings are dependent on social relations, context and prior world knowledge of the participants. Written

language, with its demands for explicitness of meaning, its permanence and its logical function, serves the cultivation of analytic, scientific and philosophical knowledge. Olsen claims that there is a literacy bias in schools because 'school intelligence' is skill in the medium of text. Clark and Salomon (1986) indicate that such considerations lead to hypotheses concerning further cognitive effects of which may lead to new 'literacy biases' succeeding older ones, for example, as may be the case with computer programming which is a highly structured and analytic activity in a rigidly constrained symbol system.

— ——

Cognition, Learning and Media

Salomon (1979) has extended these ideas in his exploration of how symbolic forms cultivate 'mental skills and affect knowledge acquisition. This acquisition is seen as being mediated by skills of information reception and processing, and because these skills are affected by the nature of the symbol system associated with a given medium, he suggests that each medium may have its own specific effects on how knowledge is extracted.

Salomon takes Olson and Bruner's (1974) 'sought after end' and expands the idea as the 'task to be performed.' He argues that the perception of the task to be performed determines the kind of information one wishes to extract from a coded message. Learning can thus be facilitated to the extent that the activated skills are relevant to the demands of the learning task. He gives as an example a task calling for analytic comparison which activates imagery instead, leading to debilitated learning. Clearly, for effective communication there must be a match between the cognitive demands of the task, the skills required by the codes of the message and the learner's mastery of the codes. There is some evidence which indicates that when certain skills are deficient switching instruction to a medium which places reliance on more highly developed skills or which compensates for the deficiency will enable information to be more readily assimilated.

To demonstrate this Salomon prepared several different versions of a film, which he reasoned would make different demands on the skills available to his fifth-grade subjects. Two versions were produced that were identical except for the fact that the shifts from long-shots to close-ups were accomplished by zoom-ins and outs in the Z version but were left out in the CU version. The Z version was designed to supplant the mental skills of connecting parts and wholes, which the CU version was expected to call upon. Prior to viewing the films the subjects were given a battery of mental tests, including a Detail and Whole Test in which a detail of a drawing was presented alongside the whole object to measure the ability to relate details to perceptual wholes, especially important in the CU version. The results indicated that:

> This difference between the Close-Up/Long-Shot (CU) and ZOOM (Z) correlational patterns strongly suports the expectation that element-specific skill-mastery is required for the acquisition of specific knowledge to the extent that the coding element deviates from the viewer's anticipatory schemata. Indeed, the Close-Up/Long-Shot version seems to call on specific skills, whereas the Zoom version overtly supplants them. Hence, intitial mastery of the relevant skills is far less necessary for knowledge acquistion from the Zoom version. (Salomon, 1979, p.102)

This relates directly to Olson and Bruner's (1974) assertion that knowledge is always mediated through some form of human activity, and the knowledge acquired through such activity has two facets: knowledge about the world and knowledge about the skills involved in gaining the knowledge.

Salomon also found that, at the time that television was being introduced to Israel, heavy viewers of the only children's tv programme (Sesame Street) did better on the Changing Points of View Test. He reasoned from this that a technique which is all pervasive in film and television is the showing of the same scene from different camera angles and repeated exposure to these changes in physical points of view develops skill in shifting points of view in the mind.

A similar effect was observed with Salomon's Space Construction Test. The task is to put four pieces of a picture together so that they form a room. Children who did well on this test were also found to be better able to understand edited films. Salomon reasoned that this is because of visual techniques that are intrinsic to film and television. Three-dimensional space is divided up by pans or cuts in films rather than revealing the whole in one shot. To have a sense of the whole the viewer must mentally integrate the various elements, and commerce with this form of coding information develops the necessary cognitve skill, which enables information to be extracted from the communication.

Implicit in this reasoning is the idea that the symbol system associated with a given medium may produce cognitive changes which result in the cultivation of certain mental skills. Salomon suggests that this is the case with language, which appears to prime cognitive operations, even to trigger them. This receives support from Vygotsky (1962), and Luria (1979) and is in line with Bruner, Olver and

Greenfield's (1966) claim that amplifiers of human capacities must produce an appropriate internal counterpart to be effective. Language is not, however, the only symbol system which participates in cognition and Salomon indicates that it should be possible to learn to think in graphic codes and he quotes the work of Hatano, Miyake and Binks (1977) which shows that expert abacus users internalize the operations of the abacus. Intermediate users still use visible finger movements to accompany mental calculations and if prevented from doing so their performance is impeded. Here the finger movements would appear to have a similar function to Vygotsky's egocentric speech, which is also on route to becoming totally internalized.

If symbol systems of media are to cultivate mental skills, as Salomon suggests, what are the psychological mechanisms involved? He indicates three ways in which cognition can be affected: (1) activation of skills that transform external codes into internal ones (2) short-circuiting skills by overtly providing the end result of mental tranformations that the learner should have employed and (3) overt supplantation, or modeling, of transformational procedures. Each of these have been shown to have a different effect on skill facilitation (Salomon, 1979). Coding elements that activate skills facilitate skill-mastery in already skillful learners; coding elements that short-circuit skills will have little cultivating effects; and coding elements that overtly model skills facilitate skill-mastery in intially unskillful learners.

Although Salomon's work has investigated symbolic codes in film and television, other media have also been shown to develop skills which are transferable. Gagnon (1985) found that giving Harvard College students 5 hours of arcade-game play improved their performance on a standardized paper-and-pencil test of visual-spatial ability, the games serving something like a remedial function for people with relatively undeveloped spatial skills. Greenfield (1987) has also demonstrated a carry-over from video game practice to tests which involved scientific-technical thinking, although novices, with only 2.5 hours practice, performed on the transfer task as well as expert games players with hundreds of hours practice. It has also been suggested that in learning to programme computers pupils will acquire powerfully general higher cognitive skills such as planning abilities, problem-solving heuristics, and 'reflectiveness on the revisionary character of the problem solving process itself.' However, Pea and Kurland (1984) caution against too much optimism and indicate that this is an old idea in a new form: similar arguments have been offered in centuries past for the beneficial cognitive consequences of other powerful symbolic systems, including mathematics, logic, writing systems and even Latin!

A New Direction for Educational Technology

Fosnot (1984) suggested that for most of its history, educational technology has attempted to justify and verify its own basic assumption that both the processes of technology and the products of technology can help improve instructional effectiveness. This has led to a systems approach to instructional design which is grounded in empiricism. However, educational technology need not remain in this position and Fosnot calls for educational technologists to come to terms with the new ideas in cognitive psychology, particularly those of constructivist psychology. Winn (1982) expresses a similar concern with instructional technology and concludes that research in cognitive psychology is further ahead than the application of research findings in instructional design. For many years the work of Piaget and Bruner, the very foundations of much exciting thinking in cognitive psychology, has been largely ignored in the field of educational technology. Now, thanks to the pioneering drive of such researchers as Papert and Salomon the new educational technology is inextricably linked to the 'booming, buzzing' world of cognitive psychology.

REFERENCES

Anschutz, R. (1961) 'August Kekule'. In E. Farber (Ed) Great Chemists. New York: Interscience Publishers.
Bandura, A. (1965) 'Influences of models' reinforcement contingencies on the acquisition of imitative response', Journal of Personality and Social Psychology, 1, 589-595.
Bandura, A. (1971) Social Learning Theory. New York: General Learning Press.
Bruner, J.S. (1960) The Process of Education. Cambridge Mass.: Harvard University Press.
Bruner, J.S. (1964) 'The course of cognitive growth', American Psychologist, 19, 1-15.
Bruner, J.S. (1966a) 'Some Elements of Discovery'. In L.S. Shulman & E.R. Keislar (Eds) Learning by Discovery. Chicago: Rand McNally.
Bruner, J.S. (1966b) Toward a Theory of Instruction. Cambridge, Mass.: Harvard University Press.
Bruner, J.S. (1972) 'Nature and uses of immaturity', American Psychologist, Aug., 687-708.
Bruner, J.S. (1983) In Search of Mind. London: Harper and Row.
Bruner, J.S., Olver, R.R. and Greenfield, P.M. (1966) Studies in Cognitive Growth. Wiley: New York.

Clark, R.E. and Salomon, G. (1986) 'Media in teaching'. In M.C. Wittrock (Ed) Handbook of Research on Teaching (3rd Edition). New York: MacMillan.

Dale, E. (1969) Audiovisual Methods in Teaching (3rd Edition). London: Holt, Rinehart and Winston.

Donaldson, M. (1978) Children's Minds. Glasgow: Fontana /Collins.

Fosnot, C.T. (1984) 'Media and technology in education: a constructivist view', Educational Communication and Technology Journal, 32, 195-205.

Gagnon, D. (1985) 'Videogames and spatial skills: an exploratory study', Educational Communication and Technology Journal, 33(4), 263-275.

Greenfield, P.M. (1987) 'Electronic Technologies, Education and Cognitive Development'. In D.E. Berger, K. Pezdek and W.P. Banks (Eds) Applications of Cognitive Psychology: Problem Solving, Education and Computing. London: Erlbaum.

Greenfield, P.M. and Bruner, J.S. (1966) 'Culture and cognitive growth', International Journal of Psychology, 1(2), 89-107.

Hadamard, J. (1945) An Essay on the Psychology of Invention in the Mathematical Field. New Jersey: Princeton University Press.

Hatano, G., Miyake, Y. and Binks, M.G. (1977) 'Performance of expert abacus operators', Cognition, 5, 57-71.

Herbert, J.J. and Harsh, C.M. (1944) 'Observational learning by cat', Journal of Comparative Psychology, 37, 81-95.

Luria, A.R. (1976) Cognitive Development: Its Cultural and Social Foundations. Cambridge, Massachusetts: Harvard University Press.

Luria, A.R. (1978) The Selected Writings of A.R. Luria. Edited by M. Cole. New York: M.E. Sharpe.

Olson, D.R. (1976) 'Towards a theory of instructional means', Educational Psychologist, 12, 14-35.

Olson, D.R. (1977) 'From utterance to text: the bias of language in speech and writing', Harvard Educational Review, 47(3), 257-281.

Olson, D.R. and Bruner, J.S. (1974) 'Learning Through Experience and Learning Through Media'. In D. R. Olson (Ed) Media and Symbols: The Form Of Expression, Communication, and Education. The Seventy-third Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Pea, R.D., and Kurland, D.M. (1984) 'On the cognitive effects of learning computer programming', New Ideas in Psychology, 2(2), 137-168.

Piaget, J. (1971) Biology and Knowledge. Chicago: Chicago University Press.

Salomon, G. (1979) Interaction of Media, Cognition and Learning. London: Jossey-Bass.

Scribner, S. and Cole, M. (1981) The Psychology of Literacy. Cambridge, Massachusetts: Harvard University Press.

Thorndike, E.L. (1898) 'Animal Intelligence', Psychological Review, Monograph Supplement, 2, No. 8.

Vygotsky, L.S. (1962) Thought and Language. London: John Wiley.

Wartofsky, M.W. (1983) 'From Genetic Epistemology to Historical Epistemology: Kant, Marx and Piaget'. In L.S. Liben (Ed) Piaget and the Foundations of Knowledge. London: Lawrence Erlbaum.

Washburn, S.L. and Howell, F.C (1960) 'Human Evolution and Culture'. In S. Tax (Ed) The Evolution of Man, Volume 2. Chicago: University of Chicago Press.

Watson, J.D. (1968) The Double Helix; A Personal Account of the Discovery of the Structure of DNA. London: Weidenfeld and Nicolson.

Winn, W. (1982) 'Visualization in learning and instruction: A cognitive effect', Educational Communication and Technology Journal, 30(1), 3-25 .

Zimmerman, B. J. (1977) 'Modeling'. In H. L. Hom and P. A. Robinson (Eds) Psychological Processes in Early Education. London: Academic Press.

ability grouping 30, 73
ability to understand
  instruction 68, 71
abstract thought 177, 182
abstract words 122, 130
academic games 72
academic standards 154
accommodation 144-146, 169
accountability 59
active manipulation 147, 153
active methods 154-156, 161
active role 7, 154
adaptation 6, 145-146, 150
affective domain 43-44, 47
amount of invested mental
  effort 121
amplifiers of human
  capacities 172, 186
analysis 43, 45, 48, 97
animal intelligence 4, 12
animal learning 4, 6, 38,
  181
application 43, 45, 97
aptitude 38, 68-71, 73-77,
  92
aptitude tests 36, 71
arithmetic 26-27, 85, 165
assessment 49, 56-57, 74-75,
  81, 118, 165
assimilation 145-146, 148,
  169
Attneave, F. 106
audio-visual methods/aids 1,
  10, 72, 103, 106, 111,
  128-130, 132-135, 137, 155,
  159, 175
auditory system/channel 105,
  109, 119, 125-129
aversive control 25-26

ballet 108
Bandura, A. 181-182
Beach, B.R. 126-128
behavioural objectives
  42-43, 49-50, 52, 57-63
behavioural sciences 1-2, 98

behaviourism, behaviouristic
  psychology 1-3, 9, 15-16,
  19, 21, 25, 44, 53, 87, 94,
  98, 105, 144, 148, 157,
  159, 166
Bennett, S.N. 161-162
Bernard, C. 11
binary digit 108, 114-115,
  116
biology 85, 145, 171
Block, J.H. 42, 67, 83,
  85-86, 97-98
Bloom, B.S. 42-50, 57-58,
  67-86, 97-99
Bobbitt, F. 47-48
bobo doll 181
Bobrow, D.G. 118-119,137
Brainerd, C.J. 160, 164
branching programmes 33-34
Briggs, L.J. 46, 52, 55-57
Broadbent, D.E. 106, 111-
  114, 116, 119, 127-128,
  136-137, 144
Bruner, J.S. 67, 162-166,
  169-187
Bugelski, B.R. 35
Burns, R.B. 86, 97-98

calculus 171
Carpenter, C.R. 18, 30, 105,
  135
Carroll, J.B. 67-75, 78, 86
Cartesian doctrine 11
central nervous system 106,
  109
chaining 54
channel capacity 110, 114,
  117
channel switching 116, 119
Charters, W.W. 48
charts 92, 119
Cherry, E.C. 111-112, 119
Chomsky, N. 172
chunking information 114
Cicero 47
Clark, R.E. 98, 184
classical conditioning 21,
  32, 54, 177

7. Spencer, K.A. (1990a) HyperCard: teaching technology for successful learning, Journal of Audiovisual Media in Medicine, 13, 25-30.

HyperCard: teaching technology for successful learning

K. A. SPENCER

A brief review of the effectiveness of instructional media is presented, with reference to mastery learning and the effects of illustrations. This is related to the need for instructional media to incorporate both of these essential elements if they are to succeed in enhancing the level of student performance. The role of computer technology, and in particular the authoring system HyperCard, in achieving this is discussed. An experimental program, HyperHeart, is used to illustrate the potential of such a system in biology.

I have been interested in the use of audiovisual teaching aids since the late 1960s. At that time film, video and the new media seemed to be exciting options to traditional chalk and talk teaching. Perhaps it was the influence of the new techniques which were being employed in the media, particularly the excitement generated by such films at Woodstock, with the emphasis on split screen effects and stereo sound. It seemed to the young media specialists as though we had access to technology which could revolutionize the process of education. Computers could be used to put men on the moon, or when harnessed to concepts borrowed from the psychological sciences, to subtly guide the process of learning for students so that none should fail. Technology was to play the central role in individualizing learning, leading to success for all.

My initial research investigations as a research student looked at the use of film as an instructional medium. By the mid-1970s, when I had completed my initial research, the results were not very encouraging. Having not conducted a full review of the research literature I had made the assumption that such methods would be superior to 'traditional' verbal methods of teaching', as many audiovisual theorists claimed. My initial research results

did not confirm my expectations and I was forced to conduct a more thorough literature search in order to gain some insights into what was happening.

A plethora of comparisons between traditional and new methods showed few differences in student performance. This led to the simplified conclusion that any medium could be used to teach any subject and that the major consideration should be the cost of instruction. My earliest research confirmed this, showing that full colour motion films were no more effective in teaching the mentally handicapped than a traditional classroom approach with the teacher using readily available materials, which were always vastly cheaper and easier to prepare than the more sophisticated movie technology.

More recently, however, I have found that there are some fairly convincing studies which show that certain approaches to teaching really do offer enormous benefits — substantial enough to be considered educationally significant.

To illustrate this I will present some comparative results, between innovatory methods and media, and a traditional method or medium. The differences in performance will be in the form of an Effect Size (ES). This is generally estimated as the average test score difference between treatment and control groups divided by the standard deviation of the control group (Glass *et al.*, 1981). This indicates when differences are educationally significant, when they will produce differences of such significance that teachers will notice the difference in the performance of their class. An ES of 1.0 means that the innovatory method or medium increases the performance of the group by an amount equal to one standard deviation unit of the control or traditional method or medium. This would take the average student from her position in the middle of the traditional group to the position occupied by the top 20 per cent of that group. This is a large effect and is educationally significant. An ES of 0.2 is a small or trivial effect, whereas an ES of 0.5 shows a medium effect, produced by a new technique which certainly merits further investigation.

There are two features of this research which are particularly striking. The first is the requirement for mastery and the second concerns the use of appropriate visual illustrations in instructional materials.

Mastery learning

Mastery approaches emphasize the achievement of high levels of performance within early teaching units before students are permitted to move on to later units. Mastery learning has been derived from two different but related strategies. The Personalized System of Instruction (PSI) was pioneered by Fred Keller, a psychologist who was impressed by Skinner's work with his teaching machines and programmed

K. A. Spencer, B.Sc., M.Sc., is a Lecturer in Educational Technology and Media, Department of Educational Studies, University of Hull, Cottingham Road, Hull HU6 7RX.

texts and realized that positive consequences (good grades, feelings of achievement etc.) are more effective facilitators of learning than negative consequences (boredom, failure etc.). Bloom's Learning For Mastery (LFM) is derived from Carroll's model of school learning (1963) which relates mastery to two major variables in learning:

1. Time needed in learning, which depends on student aptitude, the quality of instruction and the student's ability to understand instruction
2. Time spent in learning, which depends on the time allowed for learning and the student's perseverance in learning.

Although not all the variables in Carroll's scheme can be readily altered, Bloom maintains that some can, particularly the quality of instruction and the time allowed for learning. Using such techniques American researchers have demonstrated that the average student in a class can have her performance increased to that of top grade students (i.e. from 50th place in a class of 100 students to a position in the top 10–15).

From the earliest research reviews there has been a tendency for PSI to out-perform traditional methods. An early review demonstrated that 38 of the 39 studies conducted showed a better final exam performance for PSI. In 34 cases the difference was statistically significant, giving an average ES of 0.66 (Kulik, Kulik and Smith, 1976). A later meta-analysis looked at 61 studies and found an average ES of 0.5, with 35 per cent less variation in the performance of the PSI group (Kulik, Kulik and Cohen, 1979). Block and Burns (1977) found a high average ES of 0.8 for LFM studies and a combined PSI and LFM retention analysis revealed an average ES of 0.7. Bloom (1984) sees such approaches to group instruction as moving in the direction of the very large effects (ES = 2.00) produced by one-to-one tutoring.

Illustrations

Research using illustrations which aid text or verbal explanations also shows improvements in students' achievement. Such illustrations do not have to be costly, realistic representations. The extensive work conducted by

Frank Dwyer, at the University of Pennsylvania, shows that simple illustrations are often all that is required to convey concepts in the biological sciences.

Although it is now taken for granted that illustrations aid recall, comprehension and understanding, the early researches on text illustrations did not support this assumption. Samuels (1970), summarizing the research, concluded that pictures do not facilitate comprehension of text as much as had been supposed. However, later reviews disagree with Samuels. Levie and Lentz (1982), adopted an information-analysis approach and found an ES of 0.55 in favour of illustrations. In this case, the research concentrates on the learning of illustrated text information from illustrated text versus text alone. Their further analysis of Dwyer's (1978) research studies demonstrates that illustrations provide moderate benefits (ES = 0.58) when a drawing test is used, but are less beneficial for more verbal tests (terminology ES = 0.28; comprehension ES = 0.09). Levin et al. (1987) reviewed the research on illustrations according to their functions and found that the effects for decorational illustrations were near zero, but that representational, organizational and interpretational illustrations produced substantially the same ES of 0.75, with transformational illustrations (those having a direct effect on memory by targeting the critical information to be learned) having a very large effect (ES = 1.5).

Tools for the instructional designer

Given the results of research which extends over half a century, I have looked for a medium which can present information in text and graphical form and which can assess student performance and provide remedial information to enable the average student to reach levels of performance more usually associated with A-grade students. Conventional media, such as film, slides or video, fail to do this because they are essentially passive, passing information to the student, but not requiring responses from the student, although they may be under user control for such actions as stopping and starting a program. They cannot provide the means for determining the student's rate and level of

learning, and they cannot provide the necessary remedial help to enable the student to gain mastery of a given subject.

Early teaching machines could do this, in a simplified way, but they were frequently limited by the storage system used for their programs. This storage problem was resolved when computers became generally available to education in the 1960s. The computer combined the psychological principle of providing remedial help with storage for large programs, limited only by the type of computer. Early work, such as the PLATO project, used mainframe machines and could combine text, synthesized voice, graphics and animated sequences. Such projects were successful in educational terms, but tended to be expensive on a cost per student hour basis. These large scale projects paved the way for the smaller desktop machines which eventually made it possible to have computers available in most educational establishments, offering hardware and software at relatively low cost. However, the programing languages, such as BASIC, used to develop the software for these new machines were unsatisfactory for most educationalists and programs tended to be developed by computer programers rather than instructional designers or teachers. The result has been that software has generally failed to meet the requirements of teachers and their students.

There are languages which have been developed to enable educators to play a more central role in the creation of teaching programs and they can provide a useful environment for the production of software. However, the major limitation of these authoring languages, such as MicroText, is the reliance on textual presentation formats, with only very basic provision for graphics. This has been a great obstacle to the development of worthwhile software by my students in recent years. Simplified diagrams can be produced within the programming language, and drawn or scanned illustrations produced with graphics packages can be incorporated, with some difficulty, into programs running on machines which are PC compatible. Although I have made some use of the available languages, I have really been waiting for that ultimate language which will enable me to easily combine graphics and text with a means of determining rate and level of performance. T

ance of students, providing feedback and remediation when necessary, all at a reasonable cost, in cash and computer memory!

Fortunately, there is an emerging micro-based technology which can be put into the hands of teachers and instructional designers, who will find it to be an easy-to-use but powerful means of improving teaching. As with most of the useful teaching aids, it was not designed specifically for education. It is the HyperCard program which runs on the Apple Macintosh range of computers. The Macintosh computer is the result of the *enabling technology* concept which fired the imagination of its developer Steve Jobs. This concept suggests that very powerful technology should be placed at the disposal of operators to make difficult jobs easier. It was not a new concept in most walks of life, but it did represent a new approach in the world of desktop computers.

HyperCard

The HyperCard software complements the Macintosh hardware to provide the kind of tool that I have suggested should be placed in the hands of teachers and instructional designers. The program was designed by Bill Atkinson who had produced the MacPaint art program for the Macintosh and, because of its origins, graphics are an essential feature. Atkinson describes it as an authoring tool and an information organizer. Programmers usually need to invest years in acquiring the necessary programing skills associated with Mac programs (because of the complexity, which gives the appearance of simplicity). Atkinson's original purpose was to enable all users to produce sophisticated Mac programs without the need to invest so much time in acquiring the necessary programming skills. I believe he has succeeded in this. When I started, within a couple of hours I was happily drawing on cards and linking them together to form simple teaching stacks.

Cards, stacks and layers

The card is the basic unit in Hyper-Card. Several cards are put together to form a stack of cards, and it is possible to move from one stack to another as required. In Figure 1, cards 1–4 form one stack and cards 5–8 are in a

separate stack. The card has several components. There are two layers for each card, a background and a foreground. The background contains information which is common to several cards, for example the illustration of the heart in Figure 1 is placed in the background layer and appears in the following three cards. However, a stack of cards can have several different background layers.

Text and graphics

Text fields and graphics can be placed in either back- or foreground. Text can be of a variety of different fonts and sizes and can be entered in the field either by the author, in which case the field can be locked to prevent users corrupting the information, or by users. This facility allows students to add comments or insert data which can then be acted upon. Scroll bars are provided to permit extended items of text to be pulled into view as required. The Macintosh computer was designed with a graphic interface and this has been exploited within HyperCard to provide a full array of graphic tools. These include freehand drawing with the pencil tool, which is moved around the screen by the computer's mouse, and various line drawing tools, with facilities for different line widths. A variety of boxes can be produced and filled with patterns selected from a menu. Brushes of different shapes are available, as is a spray facility, both having access to the patterns menu. There is, of course, an eraser. Text can also be added within the drawing tools menu. Newly drawn items can be highlighted and moved to new locations or deleted before they are assimilated onto the card layer, giving plenty of scope for experimentation. Once assimilated, all areas of the drawn screen can be highlighted and moved using the lasso tools. And, for final touches, the screen can be put into the strangely worded 'FatBits' mode, which permits individual screen pixel editing.

Links and buttons

Having drawn the illustrations for a teaching program and having written the accompanying text, it is necessary to provide links between the various cards. This is done with the button tool. Buttons are areas of the screen which are activated when the cursor, in the shape of a hand, is placed in

them and the user 'clicks' the mouse button, i.e. when the button on the external mouse device is pressed and released. Buttons can be placed in both card layers and can be of several different appearances. Usually an icon is attached to the button to indicate its function and presence. Establishing a link between one card button and another card is done using a simple menu, which asks for the name of the relevant card.

HyperTalk

The HyperCard User's Guide, which is provided with the program, gives some indication of what can be achieved by linking cards together, but it does not provide sufficient information about the HyperTalk scripting language, which really turns the program into a useful educational tool. A more complete understanding of the authoring language is given in Goodman (1987) and Brown (1988).

I have mentioned the various features of HyperCard: the stack, the card, the background layer, the card or foreground layer, text fields, graphic areas and buttons. Each of these elements can play a dynamic role in the way a user moves through the set of cards when a script, written in the HyperTalk scripting language, is associated with it. The HyperTalk language has a syntax which is very similar to that of English, for example, on mouseUp/go to the next card/end mouseUp, is the script for a button which moves on to the next card when the mouse is clicked on the button. The script can be quite simple, as in the previous example, or it may be complex, as in the example in Figure 2. Several sections of the screen are designated as areas associated with answers to the question posed on the card. This script detects the location of the click on the screen. Having determined the location of the user's response the script compares the answer given with the correct answer and then either sends the user on to the next test item or, depending on the error made, to a remedial card.

HyperHeart: anatomy of an instructional HyperCard program

The HyperHeart program is an attempt to provide instruction in the structure and functioning of the heart

Panel 1:
The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Panel 2:
VENA CAVAE

The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Panel 3:
The blood flows from the lungs into the left atrium through the four PULMONARY veins.

The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Panel 4:
Click to cancel

The CHORDAE TENDINAE or tendinous cords (the heart strings) stop the atrio-ventricular valves from turning inside out when the ventricle contracts.

The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Click to see motion sequence

Panel 5:
The CHORDAE TENDINAE or tendinous cords (the heart strings) stop the atrio-ventricular valves from turning inside out when the ventricle contracts.

The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Click to stop sequence

Panel 6:
The CHORDAE TENDINAE or tendinous cords (the heart strings) stop the atrio-ventricular valves from turning inside out when the ventricle contracts.

The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Click to stop sequence

Panel 7:
The CHORDAE TENDINAE or tendinous cords (the heart strings) stop the atrio-ventricular valves from turning inside out when the ventricle contracts.

The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Click to stop sequence

Panel 8:
The CHORDAE TENDINAE or tendinous cords (the heart strings) stop the atrio-ventricular valves from turning inside out when the ventricle contracts.

The heart muscle is in a relaxed state and is filling with blood. The blood is being forced into the relaxed heart because of the pressure in the VENAE CAVAE and the PULMONARY VEINS. The two ATRIO-VENTRICULAR valves are open allowing blood to flow through from the ATRIA to the VENTRICLES. The CHORDAE TENDINAE are not under tension.
The SEMI-LUNAR VALVES are closed to prevent blood flowing back into the ventricles from the AORTA and PULMONARY ARTERY.

Click to stop sequence

Figure 1. HyperHeart teaching stacks.

r 14–16-year olds. There are several rsions, including the browse version own in the eight cards in Figure 1. he background layer of the stack ntains the heart diagram, the text ld giving information about the art, and two arrow buttons which ow the user to move to the next card r the previous card in the stack. The rd layer in Figure 1.1 contains a ther eight transparent buttons ich are placed around the words ich are in capital letters. Each these buttons presents more infor-ation about the word which is

covered by the button. The small hand icon is the cursor, which is moved about the screen by the user. In Figure 1.2 the cursor has been placed over the words 'venae cavae' and the external mouse button has been clicked, which causes the label to appear on the screen. This actually moves the user to a separate card which has the label in place, but the user is unaware of the move. A second click returns the display to that shown in Figure 1.1. Similarly, clicking over the words 'pulmonary veins' calls up the card with information about the

pulmonary veins, as in Figure 1.3. More detailed information is available for some of the items. When the mouse is clicked over 'chordae tendinae', the card showing a simplified illustration and associated text is displayed. This card, Figure 1.4, has two new buttons. The 'click to cancel button' returns the user to Figure 1.1. The 'click to see motion sequence' is linked to a new stack (Figure 1.5–1.8). This stack is an animation sequence in which the valves are shown opening and closing; it runs cyclically, (5,6,7,8 and back to 5) until the button 'Click to stop sequence' is activated and the user is returned to Figure 1.4.

Figure 1 demonstrates the essential features of the HyperHeart stacks, although some have been developed along more traditional lines, in which the path through the information is set out by the author in linear fashion, leaving no choice for the student other than moving forward or backward as in a traditional text.

Mastery is assessed with a series of test items similar to that shown in Figure 3. Here five possible answers are given and the student selects the answer by clicking in the appropriate box. Each box has a separate transparent button associated with it. The script for the card is shown in Figure 2. This script (and similar scripts associated with other test cards) determines which button has been clicked on and compares this with the correct answer, stored in the card name. This card is used both as a pre-test and a post-test item and the next action within the script depends on the mode (pre- or post-test). If the answer is correct a point is added to the total score and the next item is displayed, in both modes. In pre-test

```
on mouseUp
  set the cursor to 4
  set the lockscreen to true

  global TestScore, Response, RemedialCue, EntryTest

  get name of card
  put char 5 of second word of it into cardAnswer
  get short name of target
  if it is not "a" and it is not "b" and it is not "c"—
  and it is not "d" and it is not "e" then
    beep
    answer "You must click INSIDE the box" with "OK"
    exit mouseUp
  end if
  put it into RemedialCue
  put RemedialCue after Response

  if EntryTest is true then
    if RemedialCue is cardAnswer then
      add 1 to TestScore
      go next card
      exit mouseUp
    else
      go next card
      exit mouseUp
    and if
  end if

  if RemedialCue is cardAnswer then
    add 1 to TestScore
    Global QuestionSequence
    get char 1 to 4 of QuestionSequence
    put empty into char 1 to 4 of QuestionSequence
    if it is empty then go card EndTest
    else go card it
    exit mouseUp
  end if

  if RemedialCue is not cardAnswer then
    if RemedialCue is "c" then go card R01dSI
    else go card R01dNT
  end if

end mouseUp
```

Figure 2. Example HyperTalk script for HyperHeart test card.
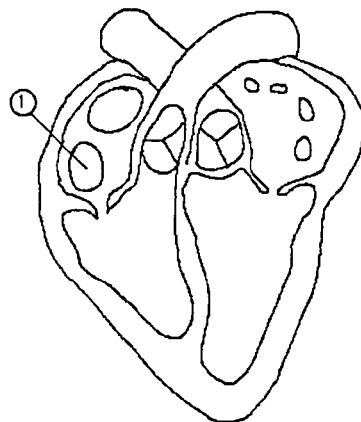
Select the answer you feel best identifies the part of the heart indicated by the numbered line.
Click in the box next to the answer you have selected

Number 1 points to the:

☐ a. aorta

☐ b. mitral valve

☐ c. superior vena cava opening

☐ d. inferior vena cava opening

☐ e. septum

Figure 3. Example HyperHeart test card.

mode, if the answer is incorrect, no point is added to the total score and the next test card is presented. In post-test mode, if the answer is incorrect, no point is added to the total score and the next test card is presented. In post-test mode, if the answer is incorrect remedial action is taken, depending on the answer given. In this case, two remedial branches are provided, a general corrective and a more specific corrective if a superior/inferior confusion is detected. In post-test mode, when all test items have been answered, the total score is determined by the script written to a 'mastery' card. Failure to reach a score of 95 per cent correct leads to a re-test and further remedial help, as required.

Initial trials with this type of program have been very encouraging. Teachers felt that a criterion of 95 per cent would deter most students and that they would simply abandon the program. This has not been the case. Even students of low ability have persevered and mastered the material, gaining a positive self-image in the process, to the surprise of their teachers. The only students who have abandoned their work were those assigned to a control condition which was text-only. The power of illustrations to aid in understanding and their importance in maintaining perseverance has been amply demonstrated in the preliminary experimental studies.

Conclusions

The results of these initial studies are most encouraging and lead to the inevitable conclusion that HyperCard and similar programmes which are now appearing, such as the Link program for PC compatibles, represent a major advance in the process of developing technological tools for teachers and instructional designers.

References

Block J. H. and Burns R. B. (1977) Mastery learning. In L. S. Shulman (Ed) *Review Of Research In Education*. Volume 4. Itasca, Ill, F. E. Peacock, 3–49.

Bloom B. S. (1984) The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership,* May, 4–17.

Brown P. J. (1988) *HyperDictionary*. London. Kogan Page.

Carroll J. (1963) A model of school learning. *Teachers College Record,* **64,** 723–733.

Dwyer F. M. (1978) *Strategies for Improving Visual Learning*. Pennsylvania, Pennsylvania State University.

Glass G. V. *et al.* (1981) *Meta-Analysis in Social Research*. Beverley Hills, Ca.: Sage Publications.

Goodman D. (1987) *The Complete HyperCard Handbook*. London, Bantam Books.

Kulik J. A., Kulik C. C. and Cohen P. A. (1979) A meta-analysis of outcome studies of Keller's personalized system of instruction. *Am. Psychol.* **34,** 307–318.

Kulik J. A., Kulik C. C. and Smith B. (1976) Research on the personalized system of instruction. *Programmed Learning and Educational Technology* **13,** 21–30.

Levie W. H. and Lentz R. (1982) Effects of text illustrations: a review of research. *Educ. Communic. Technol. J.* **30,** 195–232.

Levin J. R., Anglin G. J. and Carney R. N. (1987) On empirically validating functions of pictures in prose. In D. A. Willow and H. A. Houghton (Eds) *The Psychology of Illustration*. London: Springer-Verlag.

Samuels S. J. (1970) Effects of pictures on learning to read, comprehension and attitudes. *Rev. Educ. Res.* **40,** 397–407.

8. Spencer, K.A. (1990b) HyperHeart: does animated illustration contribute to mastery learning?, British Journal of Educational Technology, 21(3), 227-8.

HyperHeart: does animated illustration contribute to mastery learning?

Ken Spencer, School of Education, University of Hull, HU6 7RX, UK

Having been disappointed with the *no significant difference* performance of educational media innovations for more than 20 years, I am fortunate to have found that, looking from a slightly different perspective, there are now some fairly convincing meta-analysis studies which show that certain approaches to teaching really do offer benefits substantial enough to be considered educationally significant. Two areas showing consistently superior effects are mastery learning methods and the use of pictorial illustrations in instructional media.

Effect Size (ES) is generally estimated as the average test score difference between treatment and control groups divided by the standard deviation of the control group; it gives an indication of when differences are educationally significant. An ES of 1.0 means that the innovatory method or medium increases the performance of the group by an amount equal to one standard deviation unit of traditional method; this is a large effect and is educationally significant. An ES of 0.2 is a small or trivial effect.

Research reviews show a tendency for mastery methods (eg Learning For Mastery or Personalised System of Instruction) to out-perform traditional approaches, with average ESs ranging from 0.5 to 0.8, with 35% less variation in the performance of the mastery group, and a retention ES of 0.7. Illustrations which aid text or verbal explanations also shows improvements in students' achievement with average ESs ranging from 0.6 to 0.8 in favour of illustrations whose function was representational, organisational or interpretational. They are less beneficial when purely verbal tests are used (terminology ES = 0.28; comprehension ES = 0·09).

Given the above results, I have looked for a medium which can present information in text and graphical form and which can assess student performance and provide remedial information to enable the average student to reach levels of performance more usually associated with the top 20%. Conventional media fail to do this, being essentially passive and non-interactive, but I think I have found it in the form of HyperCard, which runs on the Apple Macintosh range of computers. It brings to fruition the *enabling* technology concept, which suggests that very powerful technology should be placed at the disposal of users to make difficult jobs easier.

The card is the basic unit in HyperCard and a collection of related cards is called a stack. The card has several components: background and foreground layers, text fields, graphic tools and buttons. Cards within a stack are controlled by the HyperTalk scripting language, which really turns the program into a useful educational tool.

HyperHeart

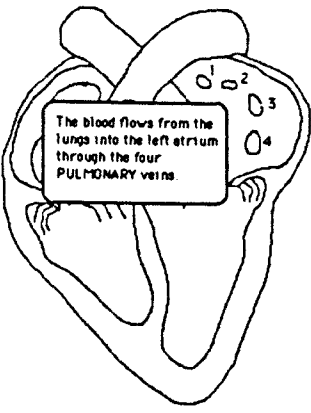The HyperHeart program is an attempt to provide instruction in the structure and functioning of the heart for 14–16 year olds based on the program which has been used by Frank Dwyer for many years. There are several versions: text-only; text plus simple line drawings; text plus simple line drawings with animation sequences. The text-only version has been shown by Dwyer to be inferior to the simple line illustrations. simple line drawings; text plus simple line drawings, with animation sequences. The

text-only version has been shown by Dwyer to be inferior to the simple line illustrations, particularly when the test incorporates illustrations. The version with animation sequences was produced to assess the value of moving pictures to explain complex, dynamic situations. The previous research in this area has tended to show that only small gains are likely to accrue from the addition of motion sequences.

Each of these versions can be of two types: mastery and non-mastery. With the non-mastery type of program, subjects pass through the structure of the heart material at their own pace, taking an illustrated test before continuing with the functioning of the heart and ending with two verbal tests and, finally, a drawing test. The mastery programs test the subject after the structure of the heart section and have a 95% criterion level. Each test item diagnoses the probable reason for an incorrect answer and provides correctives. For example, left-right, vein-artery and auricle-ventricle confusions are identified and corrected; several confusions may be identified and each corrected within one response eg left-right/auricle-ventricle. Test results and completion times are conveniently stored in a log file for each student.

Research has shown that lowering of the criterion level for mastery to 80% (from 95–100%) results in a halving of the observed ES. Initial trials with the high criterion level in this program (95%) have been very encouraging. When the program was introduced in schools, the teachers felt that such a criterion would deter most students and that they would simply abandon the program. This has not been the case. Even students of low ability have persevered and mastered the material, gaining a positive self-image in the process, much to the surprise of their teachers. The only mastery condition students who have abandoned their work were a minority of those assigned to the text-only condition. These students found too great a discrepancy between the teaching program and the demands of the illustrated test items. Certainly, the power of illustrations to aid in understanding and their importance in maintaining perseverance has been amply demonstrated in the preliminary experimental studies. The additional time taken to achieve mastery has ranged from 25% to 100%, but in all cases students have felt that the extra effort has been worthwhile.

The addition of animation sequences, which can be repeated as required by the student, obviously increases the time taken to complete the program but does not appear to influence the performance as measured by the tests. This confirms the prediction based on previous research in this area, and once again raises the question of the value of motion sequences for teaching such topics. Indeed, it may be said to challenge the whole concept of interactive video: why link the computer to an expensive motion videodisc system when simpler, less expensive systems are just as efficient?

The results of this study, which is in its early stages, strongly support the view that the latest generation of computers, combining text and graphic displays at an affordable price, can provide instruction that leads to excellence, provided that educational technologists designing the educational software apply the knowledge gained through the years from research and avoid the seduction of superfluous, but superficially more attractive methods of display.

9. Spencer, K.A. (1991) Modes, media and methods: the search for educational effectiveness, British Journal of Educational Technology, 22(1), 12-22.

Modes, media and methods:
the search for educational effectiveness

Ken Spencer

Ken Spencer is a lecturer in educational technology and instructional media in the School of Education, University of Hull, UK.

Abstract
The educational effectiveness of illustrations, visual-based instructional media, programmed learning, computer assisted learning, audio-tutorial instruction, organisation of groups, and mastery learning strategies, are discussed. Attention is drawn to those aspects which produce educationally significant improvements in student performance. It is suggested that this research, which clearly indicates a variety of ways in which levels of attainment can be greatly enhanced, is being ignored in new curriculum developments. The author concludes that the improvements sought by educationists and politicians will not be achieved until the knowledge gained from many years of research is acted upon.

For decades educationists have been searching for improved instructional techniques and investigating many different approaches. However, the general result has been that most innovations have resulted in about the same level of performance as that obtained using traditional approaches. This is frequently reported, in statistical terms, as 'no significant difference' or NSD. Some investigations have looked at the effects of specific media, or even specific modes of instruction, while others have looked at broader approaches which may incoporate several different media and/or particular techniques, such as mastery learning.

Here I use the term *mode*, in the sense that Rowntree (1982) defines it, as the kind of stimulus presented to the student. According to this scheme, five modes are available, including human interaction, realia (real things), pictorial representations, written symbols and recorded sounds. However, I prefer the use of a general *interactive* mode term, rather than one specifically designated as human interaction. This broader term includes both human and machine interactions, and comes closer to Rowntree's intended meaning.

In Rowntree's scheme, the *vehicle which carries* the stimulus mode is referred to as the

medium. Thus, several media can present a given stimulus mode eg written symbols may be found in such media as television, filmstrips and books. Similarly, a given medium, may carry several modes, eg film may convey representation, written symbols and recorded sounds.

In addition to modes and media there is the broader category which may be termed *methods* or approaches. Here a variety of different media may be incorporated, together with a particular emphasis on the style of teaching. For example, the audio-tutorial approach to biological instruction places great emphasis on individualised instruction in a multi-media environment including audiotape, slides, movies or video, a variety of printed materials and experimental apparatus, together with opportunities for human interaction. Most of these media are also used in traditional approaches to biology teaching. The distinguishing features are individualised learning and formative evaluation.

Comparative studies
There are few comparisons dealing specifically with instructional modes because most media and methods tend to include several different modes. However, one area, pictorial representation, has received considerable attention in the form of comparisons between illustrated and non-illustrated teaching materials.

Much of the available research has dealt with comparisons between a particular medium (or method/approach) and the traditional way of teaching the subject. For many media comparisons this consists of the new medium being used to deliver most of the course content, for example, through the medium of television or video. This is then compared with the usual way of teaching the material, which may be chalk and talk or may include alternative media eg teacher plus overhead projector.

Similarly, new methods are compared with those used prior to the new development. This could include comparisons of essentially the same instructional media used in conjunction with different teaching styles or methods of organisation, as is the case when the Audio-tutorial approach is compared with a traditional biological sciences course.

Box scores or effect sizes?
Early research studies tended to report comparative statistics which gave no indication of the size of differences, if any, between the types of instruction. Conclusions tended to be based on the presence or absence of a statistically significant result. When groups of researches were reviewed there was a tendency to use a 'box score' tally approach, frequently resulting in a small number of studies favouring the innovation, a similar number favouring the traditional approach, and the vast majority showing NSD.

Criticisms of this approach to summarising research have been accommodated by the new meta-analytic methods of formulating generalisations (Glass *et al* 1981). Meta-analysis is simply the statistical analysis of a large collection of results from individual

studies for the purpose of integrating findings. It seeks to determine the effect of a specific treatment, influence or intervention from a large variety of individual studies on a particular subject. Of course, the quality of the individual studies is of crucial importance and selection criteria must be stringent. The meta-analytic procedures yield effect size estimates which are converted to 'percentage of standard deviation gains' due to more powerful treatment, if any.

The effect size (ES) is a measure of the educational importance of any performance changes produced. It is generally estimated as the average test score difference between treatment and control groups divided by the standard deviation of the control group. In education we are not simply interested in differences in a statistical sense, we wish to know when such differences are educationally significant. An ES of 1.0 means that the innovation increases the performance of the group by an amount equal to one standard deviation unit of the control treatment. This would take the average student from a position in the middle of the traditional group to the position occupied by the top 20% of that group. This is a large effect and is educationally significant. An ES of 0.2 is a small or trivial effect, whereas an ES of 0.5 shows a moderate effect, produced by a treatment which certainly merits further investigation.

Modes, media and methods

Whatever the method of reporting the results, general media comparisons and studies pertaining to their overall instructional impact have yielded little that warrants optimism and Clark provides a typical commentary on this state of affairs:

> The best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition. Basically, the choice of vehicle might influence the cost or extent of distributing instruction, but only the content of the vehicle can influence achievement (Clark 1983).

Comparisons of new methods give more room for optimism. Some, but not all, demonstrate consistently superior, educationally significant levels of performance when compared with traditional methods. Similarly, the main line of what may be termed *pictorial mode* research looks encouraging. These results can, I believe, offer guidelines for the development of successful teaching strategies, but they may be in conflict with *a priori* perceptions of the essential ingredients for effective instruction.

Pictorial mode

Illustrations may be said to form a very significant part of the pictorial representation mode. The are the very essence of visual-based instructional media and methods. The pictorial mode of instruction applies to many different media and it is often assumed that it must aid recall, comprehension and understanding. However, the early researches did not support this assumption (Miller 1938, Vernon 1953), leading Samuels (1970) to conclude that pictures do not facilitate comprehension as much as had been supposed.

Later research, reviewed by Levie and Lentz (1982), disagreed with Samuels. When an *information-analysis* approach was adopted, Levie and Lentz found an ES of 0.55 in favour of illustrations. In this case, the research concentrated on the learning of illustrated text information from illustrated text versus text alone. Their further analysis of Dwyer's (1978) research studies demonstrated that illustrations provide moderate benefits (ES = 0.58) when a drawing test is used, but are less beneficial for more verbal tests (terminology ES = 0.28; comprehension ES = 0.09).

Levin *et al* (1987) reviewed the research on illustrations (including tests which were presented orally to children) according to their functions and found that the effects for decorational illustrations were near zero, but that representational, organisational and interpretational illustrations produced substantially the same ES of 0.75, with transformational illustrations having a very large effect (ES = 1.5). Transformational illustrations have a direct effect on memory by targeting the critical information to be learned, the representation frequently embodying disparate elements in a coherent (though possibly bizarre) whole. Educational media make extensive use of representational, organizational and interpretational illustrations. However, the general application of transformational illustrations is probably rather limited.

These experimental results clearly demonstrate the value of the pictorial representation mode. Since illustrations are an essential part of all visual-based media, the research is a vindication of visual-based instructional theory.

Visual-based instructional media
Visual-based instruction is a term which can be applied to a variety of different media (filmstrips, films, television and so on). Such educational media can be seen to make learning more concrete and relevant by bringing the real world into the classroom through the use of a variety of still and moving pictorial displays. They are seen as the antithesis of traditional verbal methods.

Films and filmstrips
In studies in different age groups and for different subjects, films and filmstrips generally have shown no significant differences when compared with traditional chalk and talk, or talk plus relevant illustrations: 'Most studies do not reveal reliable differences between instruction by films as compared to conventional methods, over a wide range of subject matter, age ranges, abilities, and conditions of use' (Campeau 1966). Looking closer, studies also showed that expensive full-colour movie presentations were, in fact, no more effective than crude black and white still pictures (Twyford 1954, May and Lumsdaine 1956).

Television
Television was introduced in the early 1950s and it was hoped that, by putting the best teachers on the screen and using television as a window on the world, TV would prove to be more effective than traditional methods. By the mid-1960s there was sufficient evidence to show that, again, there was no real benefit, in terms of effectiveness, from

the new medium in a majority of more than 200 research studies. Chu and Schramm (1968) found that in 73% of studies there was no significant difference between traditional and television methods.

Cohen *et al* (1981) have reviewed the visual-based instruction research, which included still projection, film, multi-media, closed-circuit television, educational television and video, using meta-analysis techniques. The outcome is predictable from previous results with an average ES of 0.15. A moderate ES of 0.41 was observed for the use of video as a source of feedback for teacher training or skill acquisition.

Programmed learning methods
At about the same time that TV was entering the educational arena, a new method also appeared. Programmed learning was rooted in behaviouristic psychology and soon became the most evaluated method in the history of education. A wide variety of media was employed to implement this method, ranging from simple linear teaching machines to group programmed learning via television. Programmed texts were also popular and the method has also been applied to the computer medium. Unfortunately, it didn't live up to early expectations and by 1964 the cumulative evidence showed that it could be used for many age groups covering almost any aspect of education and that students did indeed learn as much as by traditional methods, but not much more, although there could be savings in learning time (Schramm 1964).

The latest evidence suggests that small but positive ESs will generally be found. For higher education this is in the region of 0.24 (Kulik Cohen and Ebeling 1980); but for secondary schools it is only 0.08 (Kulik *et al* 1982). Some subjects benefit more than others: surprisingly, social sciences show a moderate ES of 0.6 whereas maths produce a small negative ES (Hartley 1977).

Computer aided instructional methods
The computer is the main medium in such methods, potentially delivering all modes of instruction, except realia. The medium can be used in a variety of different ways, ranging from basic drill and practice routines to more sophisticated interactive modes such as tutoring and simulations. Computer aided instruction, the great hope of the 1960s, failed to deliver consistent educational benefits, probably because it was only seen as a convenient medium for the implementation of programmed learning techniques. Mid-1970s reviews concluded that only small differences were to be expected from CAI: 'At secondary school and college levels CAI is about as effective as traditional instruction' (Jamison *et al* 1974).

The picture remains the same for more recent reviews, with small ESs for most of the educational uses of computers. In higher education these included tutoring, simulation, managing and use of computers for programming, with substantially the same small ES of 0.25 (Kulik Kulik and Cohen 1980). For secondary schools an average ES of 0.32 was obtained (Kulik *et al* 1983), although interactive simulations gave a promising ES of 0.49 and tutoring an ES of 0.36.

The audio-tutorial method (AT)

Postlethwait's system evolved from the use of audiotaped lectures for remedial instruction. In its developed form students work independently, listening to the audiotape and following instructions which may involve carrying out experiments, reading articles or watching multi-media displays. There is regular formative evaluation of student performance and although this may include a diagnostic element, there is not a mastery requirement. It has proved to be a very popular method of instruction in a wide range of subjects. The first major review of research was favourable, showing more than 60% of studies favouring AT when compared with more traditional teaching (Fisher and MacWhinney 1976). These results are indicative of the false impression that some comparative statistical approaches convey. Later reviews pointed out the small performance difference anticipated from AT implementation (Kulik and Jaksa 1977) and when meta-analytic procedures were applied to the research data this was confirmed, with a small ES of 0.25 (Kulik *et al* 1979).

Methods of group organization

Accumulated evidence also shows that various methods of grouping and teaching, ranging from tutorials to lectures and two-way telephone discussions, fail to produce significant instructional benefits. Dubin and Taveggia (1968) summarised a total of 91 studies from 1924 to 1965 which compared such approaches to learning with more traditional methods. The results led them to title two of their chapters 'The Sound and Fury Signify Nothing' and 'There Are No Differences!' Later meta-analyses have shown a moderate ES of 0.4 for the effects of individual tutoring (Cohen *et al* 1982) but have shown only a small average ES of 0.1 for the effects of ability groups.

> Effects were near zero in the four programs designed especially for academically deficient students: such students learned as much in mixed-ability classrooms as they did in homogeneous classrooms. Finally, effects were also near zero in the 33 studies that compared effects of multi-track versus mixed-ability classrooms on unrestricted populations (Kulik and Kulik 1982).

The same trivial average ES was found for studies comparing individualised systems of instruction (Bangert *et al* 1983) and for class-size effects (Walberg 1984).

Mastery learning methods

The Personalised System of Instruction (PSI)

The Personalised System of Instruction was pioneered by Fred Keller, a psychologist who was impressed by Skinner's work with his teaching machines and programmed texts and realised that positive consequences (good grades, feelings of achievement and so on) are more effective facilitators of learning than negative consequences (boredom, failure etc). Keller Plan specifies objectives and provides reinforcement for their successful achievement, as well as giving more opportunities for personal interaction than traditional systems. This is achieved by the provision of proctors (student tutors) who monitor student progress and ensure mastery of each teaching unit.

Learning for Mastery (LFM)
Bloom's Learning for Mastery is derived from Carroll's (1963) model of school learning which relates mastery to two major variables in learning:

- time needed in learning, which depends on student aptitude, the quality of instruction and the student's ability to understand instruction
- time spent in learning, which depends on the time allowed for learning and the student's perseverance in learning.

Although not all the variables in Carroll's scheme can be readily altered, Bloom maintains that some can, particularly the quality of instruction and the time allowed for learning. Diagnostic, formative evaluation is a crucial element in this method. Students who fail to master study units are provided with additional tuition which may consist of a variety of different instructional media. Bloom emphasises that this method can be applied within the usual weekly timescale of schools and does not require large-scale change.

Mastery research
From the earliest research reviews there has been a tendency for PSI to out-perform traditional methods. An early review demonstrated that 38 of the 39 studies conducted showed a better final exam performance for PSI. In 34 cases the difference was statistically significant, giving an average ES of 0.66 (Kulik *et al* 1976). A later meta-analysis looked at 61 studies and found an average ES of 0.5, with 35% less variation in the performance of the PSI group (Kulik *et al* 1979). A crucial feature is the high mastery criterion. Relaxation of this criterion leads to a decline in final test performance. With mastery at 70–85% the ES is 0.4, but with mastery set at 100% the ES is 0.82 (Kulik *et al* 1986). Block and Burns (1977) found a high average ES of 0.8 for LFM studies and a combined PSI and LFM retention analysis revealed an average ES of 0.7.

These methods are certainly producing significant educational effects and this has led Walberg (1984) to comment that:

> . . . the psychological components of mastery learning rank first and fourth in their effects on educational outcomes: Skinnerian reinforcement or reward for correct performance has the largest overall average effect—1.17 standard deviations: instructional cues, engagements, and corrective feedback have effects equal to approximately one standard deviation. Separate syntheses of mastery programs in science show an average effect of 0.8.

The emphasis on feedback and correctives, so essential to the mastery programmes, may also account for the discrepancy between assigned and graded homework. Walberg also indicated that homework that is commented upon and graded has nearly three times the effect of homework that is merely assigned, giving an ES of 0.79.

Conclusions concerning future developments
I have summarised the ESs associated with the different researches in Table 1. In order

Table 1: Modes, media and methods: summary of effect sizes

	Effect size	
Small effect sizes (under 0.35)		
Programmed learning (secondary)	0.08	*
Individualized instruction	0.1	*
Ability groups	0.1	*
Visual-based instruction	0.15	**
Programmed learning (higher)	0.24	**
Audio-Tutorial	0.25	***
Computer Assisted Instruction (higher)	0.25	***
Assigned homework	0.28	***
Computer Assisted Instruction (secondary)	0.32	***
Moderate effect sizes (0.35–0.69)		
Tutoring	0.4	****
Video feedback	0.41	****
Computer Simulations	0.49	*****
PSI	0.58†	******
Large effect sizes (0.7 and over)		
LFM and PSI (retention test)	0.7	*******
Illustrations	0.75	********
Graded homework	0.79	********
LFM	0.8	********

* represents the size of effect in tenths of a standard deviation
† average

to make the data more intelligible I have divided the studies into groups according to whether the reported effect size was small (under 0.35), moderate (0.35 to 0.69) or large (0.7 and over). Using these figures, let us consider and resolve some of the conflicting sets of data before attempting to formulate a general position for future developments in instructional techniques.

Modes and media

Visual illustrations are of central importance when considering the effects of visual-based teaching media. Why then are significant educational effects observed for the pictorial mode studies but not for visual-based media? The answer is that pictorial mode effects are more powerful than media effects because the mode studies compare two fundamentally different ways of presenting information: uni-modal and bi-modal. For example, Dwyer compares a verbal only presentation (uni-modal) of the structure and functioning of the heart with a combined verbal and pictorial presentation (bi-modal). When the illustrations are removed, student performance will, not surprisingly, be affected adversely, particularly when tests incorporate a visual dimension.

Cognitive psychology is beginning to supply an explanation for this. There is now strong evidence that information is stored in two separate but inter-connected systems

within the human organism: a verbal system and an image system. Media which involve bi-modal presentations, addressing both storage systems, will be more effective than uni-modal media, particularly when the tests are bi-modal.

There is evidence that the human information system compresses incoming information, reducing redundancy. Illustrations which are compressed forms of reality and therefore less demanding of the processing system, appear to be stored and retrieved more efficiently. Hence, simple line illustrations are found to be as effective as more complex, realistic representations.

Media research has frequently made the mistake of comparing like with like in terms of modes—for example, a lecture illustrated with blackboard illustrations might be compared with a televised lecture using more elaborate illustrations. The outcome is the same in both cases because, although the pictorial mode is a necessary condition for the activation of the iconic memory storage system, transparencies, television graphics, printed illustrations or blackboard drawings are *all* sufficient for this activation to take place.

Of course, not all concepts are readily illustrated with pictorial representations; in such cases, a word is worth a thousand pictures, and accompanying illustrations may even have a negative effect.

Modes and methods
The results for the different methods also present an interpretive challenge. A mode analysis may also be appropriate here. I would argue that the interactive mode, which includes machine simulations of human interactions, provides the key to the success of those methods which are more effective. The essence of this mode is that it provides opportunities for feedback. The nature of that feedback is crucial. It can vary from the simplest indication, without explanation, that an answer is incorrect, to the full diagnostic interactions with a proctor that are found in PSI. The primary purpose of the diagnosis in mastery learning methods is to ensure complete understanding of the taught material. The most effective methods of instruction seem to include these diagnostic interactions combined with mastery conditions.

On this basis, programmed learning methods which simply report success or failure, providing minimal interactive features, no diagnosis and no mastery conditions, are not very effective. Similarly, computer based methods which show all the hallmarks of early programmed learning methods, are likely to be no more effective. It is interesting to note that the computer is most successful when interactively simulating real world events or tutoring.

Although nominally a tutorial approach, Postlethwait's audio-tutorial method uses a non-interactive medium (audiotape) to provide the tutoring. This proves to be inadequate in terms of improving student performance because the level of learning is not checked at this stage. Weekly testing and grading sessions do not appear to be

effective in raising student performance, probably because the emphasis is misplaced, with grading of students taking precedence over mastery of course units. LFM and PSI combine diagnostic interactions with high mastery criteria and consequently show impressive, educationally significant benefits. However, the large effect produced by graded homework can only be partially explained on the basis of diagnostic interactions.

Raising standards of performance

These results indicate that the technology for substantially enhancing learning is available and can be delivered using a variety of different techniques. Keller Plan uses traditional media vehicles, human tutors and print materials, to implement an interactive mastery approach; it represents a simple and effective method for raising student performance. Bloom's Learning for Mastery offers a similar approach within a classroom context. Effective computer programs, which simulate human activities and are efficient at handling mastery requirements, are also available. What seems to be crucial is the application of the technology *of* education rather than the provision of technology *in* education. Unfortunately, the latter is frequently given priority when non-educational agencies are involved, ill-informed decision-makers being seduced by the superficial appeal of new hardware approaches.

Finally, recent curriculum development proposals in this country, ostensibly aimed at raising standards of performance, seem to place considerable emphasis on increased testing. However, they fail to acknowledge that unless testing is both diagnostic and related to remedial strategies aimed at ensuring high levels of mastery performance, it will merely reflect the same cumulative deficits and failure, the same low standard of performance, as before. If we wish to raise standards, this is not the way. As Bloom has pointed out:

> the solution lies in our views about students and their learning. These have grown out of our practices and they will not be changed until we alter these practices. When the changed practices suceed in promoting more effective learning, both teachers and students will change their views (Bloom 1978, 563).

References

Bangert, R L, Kulik, J A and Kulik, C C (1983) Individualized systems of instruction in secondary schools *Reviews of Eucational Research* 53 2 143–158

Block, J H and Burns, R B (1977) Mastery learning *in* Shulman LS (ed) *Review of Research in Education.* Volume 4, FE Peacock, Itasca, Illinois, 3–49

Bloom, B S (1978) New views of the learner—implications for instruction and curriculum *Educational Leadership* 35 562–76

Campeau, P L (1966) Selective review of literature on audiovisual media of instruction *in* Briggs, L J, Campeau, P L, Gagne, R M and May, M A (eds) *Instructional Media: A Procedure for the Design of Multi-media Instruction and Suggestions for Future Research* Monograph 2, American Institutes for Research, Pittsburgh

Carroll, J (1963) A model of school learning *Teachers College Record*, 64 723–733

Chu, G C and Schramm, W (1968) *Learning from Television: What the Research Says*, Institute for Communication Research, Stanford University

Clark, R E (1983) Reconsidering the research on learning from media *Review of Educational Research* 53 4 445–459

Cohen, P A, Ebeling, B and Kulik, J A (1981) A meta-analysis of the outcome studies of visual-based instruction *Educational Communication & Technology Journal* 29, 1 26–36

Cohen, P A, Kulik, J A and Kulik, CC (1982) Educational outcomes of tutoring: a meta-analysis of findings *American Educational Research Journal* 19 237–248

Dubin, R and Taveggia, T C (1968) *The Teaching-Learning Paradox: A Comparative Analysis of College Teaching Methods*, Centre for the Advanced Study of Educational Administration, University of Oregon

Dwyer, F M (1978) *Strategies for Improving Visual Learning*, Pennsylvania State University, Pennsylvania

Fisher, K M and MacWhinney, B (1976) AV autotutorial instruction: a review of evaluative research *Audio Visual Communication Review* 24 3 229–261

Glass, G V et al (1981) *Meta-Analysis in Social Research*, Sage Publications, Beverley Hills

Hartley, S S (1977) *Meta-analysis of the Effects of Individually-paced Instruction in Mathematics*, Doctoral Dissertation, University of Colorado

Jamison, D T, Suppes, P and Wells, (1974) The effectiveness of alternative instructional media: a survey *Review of Educational Research* 44 1 1–67

Kulik, C C and Kulik, J A Effects of ability grouping on secondary school students: a meta-analysis of evaluation findings *American Educational Research Journal* 19 415–428

Kulik, C C, Kulik, J A and Bangert-Drowns, R L (1986) 'Effects of testing for mastery on student learning' paper presented at the annual meeting of the American Educational Research Association, San Francisco

Kulik, C C, Schwalb, B J and Kulik, J A (1982) Programmed instruction in secondary education: a meta-analysis of evaluation findings *Journal of Educational Research* 15 3 133–138

Kulik, J A and Jaksa, P (1977) PSI and other educational technologies in college teaching *Educational Technology* 17 12–19

Kulik, J A, Bangert, R L and Williams, G W (1983) Effects of computer based teaching on secondary school students: a meta-analysis *Journal of Educational Psychology* 75 19–26

Kulik, J A, Cohen, P A and Ebeling, B J (1980) Effectiveness of programmed instruction in higher education: a meta-analysis of the findings *Educational Evaluation & Policy Analysis* 2 6 51–64

Kulik, J A, Kulik, C C and Cohen, P A (1974) A meta-analysis of outcome studies of Keller's personalised system of instruction *American Psychologist* 34, 4 307–318

Kulik, J A, Kulik, C C and Cohen, P A Effectiveness of computer-based college teaching: a meta-analysis of findings *Review of Educational Research* 50, 4 525–545

Kulik, J A, Kulik, C C and Cohen, P A (1979) Research on audio-tutorial instruction: a meta-analysis of comparative studies *Research in Higher Education* 11 4 321–341

Kulik, J A, Kulik, C C and Smith, B (1976) Research on the personalised system of instruction *Programmed Learning and Educational Technology* 13 21–30

Levie, W H and Lentz, R (1982) Effects of text illustrations: a review of research *Educational Communication and Technology Journal* 30 195–232

Levin, J R, Anglin, G J and Carney, R N (1987) On empirically validating functions of pictures in prose *in* Willows D A and Houghton H A (eds) *The Psychology of Illustration* (Vol. 1) Springer-Verlag, London

May, M A and Lumsdaine, A A (1958) *Learning from Films*, Yale University Press, New Haven

Miller, W A (1938) Reading with and without pictures *Elementary School Journal*, 38 676–682

Rowntree, D (1982) *Educational Technology in Curriculum Development*, Harper and Row, London

Samuels, S J (1970) Effects of pictures on learning to read, comprehension and attitudes *Review of Educational Research* 40, 397–407

Schramm, W L, *The Research on Programmed Instruction*, US Government Printing Office, Washington DC, 1964

Twyford, L C, Davis, J S and Seitz, C P (1954) *Evaluation of the Film: Military Police Support in Emergencies (Riot Control)*, Technical Report No. SDC 269-7-52, Instructional Film Program, Naval Training Devices Centre, Office of Naval Research, Port Washington, New York

Vernon, M D (1953) The value of pictorial illustration *British Journal of Educational Psychology* 23 180–187

Walberg, H J (1984) Improving the productivity of America's schools *Educational Leadership* May 19–27

10. Spencer, K.A. (1992) Information technology and human information processing: a critical perspective, Proceedings of the European Conference about Information Technology in Education, Barcelona, November, 513-522.

Information Technology and Human Information Processing: a Critical Perspective

Ken Spencer,
School of Education,
Hull University,
Hull HU6 7RX, UK.

Abstract

Information technology is the latest in a long line of media and methods to be applied to education. It would be unfortunate if research into the effectiveness of earlier technological innovations was not brought to bear on current IT practice: a critical insight into the value of information technology must capitalize on the results of past research.

Attention is drawn to those aspects of information technologies which produce educationally significant improvements in student performance and have relevance for current IT practice. The author concludes that these improvements, sought by educationalists, students, parents and politicians, will not be achieved until the knowledge gained from many years of research is acted upon.

Key words: human information processing media mastery

Introduction

For decades educationalists have been searching for improved instructional techniques and have investigated many different approaches. Some investigations have looked at the effects of specific media (eg. interactive video), or even specific modes of instruction (eg. illustrations), while others have looked at broader approaches, which may incorporate several different media and/or particular techniques, such as mastery learning.

Whatever the method of reporting the results, general media comparisons and studies looking at their overall instructional impact have yielded little that warrants optimism, and Clark provides a typical commentary on this state of affairs:

> The best current evidence is that media are mere vehicles that deliver
> instruction but do not influence student achievement any more than the
> truck that delivers our groceries causes changes in our nutrition.
> (Clark, 1983)

513

However, comparisons of some methods give more room for optimism because they demonstrate consistently superior, educationally-significant levels of performance when compared with traditional methods. Such research studies can, I believe, offer guide-lines for the development of successful teaching strategies involving information technology.

Graphic representations

Illustrations are the very essence of visual-based instructional media and methods. The pictorial mode of instruction applies to many different media and is central to the idea of modern information technologies, with their emphasis on graphic environments. It is usually assumed that graphics must aid recall, comprehension and understanding. However, early researches into the educational effects of illustrations did not support this assumption (Miller, 1938; Vernon, 1953), leading Samuels (1970) to conclude that pictures do not facilitate comprehension as much as had been supposed.

Later research disagreed with Samuels interpretation. The review by Levie and Lentz (1982) indicated an effect size (ES) of 0.55 in favour of illustrations, and their analysis of Dwyer's (1978) research studies demonstrated that illustrations provide moderate benefits (ES=0.58) when a drawing test is used, but are less beneficial for verbal tests (terminology ES=0.28; comprehension ES=0.09).

Levin et al. (1987) reviewed the research on illustrations according to five functions. They found that the effects for decorational illustrations were near zero, but that representational, organisational and interpretational illustrations, those used most frequently in educational media, produced substantially the same ES of 0.75, with transformational illustrations having a very large effect (ES=1.5).

These experimental results clearly demonstrate the value of the pictorial representation mode and, as illustrations are an essential part of all visual-based media, the research seems to support a visual-based theory of instruction. The results from research on visual-based media are, however, contradictory!

Visual-based media

Visual-based instruction is a term which can be applied to a variety of different media (filmstrips, television, interactive video etc.). Such educational media are thought to make learning more concrete and relevant by bring the real world into the classroom through the use of a variety of still and moving pictorial displays and are seen as the antithesis of traditional verbal methods. However, research

514

conducted with different age groups and for different subjects generally shows no significant difference between visual-based media and traditional methods.

Television provides a good example. By the mid-1960s there was sufficient evidence from 200 studies to show that in 73% of comparisons there was no significant difference between traditional and television methods (Chu and Schramm, 1968).

Cohen et al. (1981) have reviewed the visual-based instruction research, which included still projection, film, multi-media, closed-circuit television, educational television and video, using meta-analysis techniques. The outcome is predictable from previous results with an average ES of 0.15. Interestingly, a moderate ES of 0.41 was observed for the use of video as a source of feedback for teacher training or skill acquisition.

Looking closer, studies also showed that expensive full-colour movie presentations were, in fact, no more effective than crude black and white still pictures (Twyford, 1954; May and Lumsdaine, 1956).

Graphic representations are of central importance when considering the effects of the new information technologies and we must understand why significant educational effects are observed for the pictorial studies but not for visual-based media.

Human information processing
Cognitive psychology is beginning to supply an explanation for this. There is now strong evidence that information is stored in two separate but inter-connected systems within the brain: a verbal system, which accepts speech or print, and an image system. Research shows that media which involve bi-modal presentations, such as the new generation of graphics-based computers, address both storage systems and are more effective than uni-modal media, particularly when the tests are also bi-modal (Spencer, 1991).

Dwyer's (1978) research also suggests that the human information system compresses incoming information, reducing redundancy, and that illustrations such as simple line drawings, which are themselves compressed forms of reality and therefore less demanding of the processing system, are stored and retrieved more efficiently. Dwyer found that these line illustrations are as effective as more

complex, realistic representations such as colour photographs, contrary to what have been termed 'realism' theories.

The conflicting research results have come about because visual-based media research has made the mistake of comparing like with like, in terms of modes. Here the pictorial dimension is a necessary condition for the activation of the iconic memory storage system, but the form (eg. transparencies, computer graphics, printed illustrations or blackboard drawings) is irrelevant: all are sufficient for this activation to take place.

There is likely to be a conflict here between researchers and the information technology hardware and software producers. The researchers have produced evidence which supports the absolute necessity of visual illustrative instructional material for effective teaching and learning. But the evidence also indicates that simple illustrations are just as effective as more elaborate presentations.

HyperHeart research

I have recently investigated several different approaches to teaching with information technology. One comparison featured two versions of an illustrated presentation on the structure and function of the human heart (Spencer, 1990). The first used simple line illustrations, similar to those in Dwyer's experiment. The second was the same but also had animation sequences to explain particular aspects of the functioning of the heart. These animation sequences were under the student's control and could be viewed at will.

The results confirmed my suspicions: the human information processing system is able to fill in the gaps between the various stages of such a presentation and animated graphics do not appear to be very beneficial in terms of increasing student achievement. Research from several earlier studies with film, television and transparencies have also reached the same conclusion (Spencer, 1981).

These results are not going to please many hardware and software producers who are committed to giving us systems which offer increasingly elaborate displays in colour with full motion pictures. Such displays may be very attractive, but are they going to be more effective in educational terms?

Please do not misunderstand me: I am not against technology as a means of improving student performance; I merely wish to see the technology used for what it can do best and my second line of research has shown what this may be!

Much of the research on media and methods in education has been disappointing, but one area has shown more promise than others: the provision of corrective feedback and a demand for mastery performance.

Table 1.   Modes, Media and Methods: Summary of Effect Sizes (Spencer, 1991).

Effect Size

Small Effect Sizes (Range 0.1-0.3)

Programmed learning (Secondary)	0.08	*
Individualized instruction	0.1	*
Ability groups	0.1	*
Visual-based instruction	0.15	**
Programmed learning (Higher)	0.24	**
Audio-tutorial	0.25	***
Computer assisted instruction (Higher)	0.25	***
Assigned homework	0.28	***
Computer assisted instruction (Secondary)	0.32	***

Moderate Effect Sizes (Range 0.4-0.6)

Tutoring	0.4	****
Video feedback	0.41	****
Computer simulations	0.49	*****
PSI	0.58 †	******

Large Effect Sizes (Range 0.7 and above)

LFM and PSI (retention test)	0.7	*******
Illustrations	0.75	********
Graded homework	0.79	********
LFM	0.8	********

†	average
*	represents the size of effect in tenths of a standard deviation

This approach has shown a tendency to consistently out-perform traditional methods. An early review demonstrated that 38 of the 39 studies conducted showed a better final exam performance for the Personalized System of Instruction (PSI), a mastery method pioneered by Fred Keller, giving an average ES of 0.66 (Kulik, Kulik and Smith, 1976). A later meta-analysis looked at 61 studies and found an average ES of 0.5, with 35% less variation in the performance of the PSI group

517

(Kulik, Kulik and Cohen, 1979). Block and Burns (1977) found a high average ES of 0.8 for Bloom's Learning for Mastery (LFM) strategy and a combined PSI and LFM retention analysis revealed an average ES of 0.7.

These methods are certainly producing significant educational effects and this has led Walberg (1984) to comment that:

> .... the psychological components of mastery learning rank first and fourth in their effects on educational outcomes:... instructional cues, engagement, and corrective feedback have effects equal to approximately one standard deviation. Separate syntheses of mastery programs in science show an average effect of .8.

My work in this area has featured a variation of the human heart unit previously described, with programme modules which analyzed student responses and diagnosed specific learning errors leading to corrective feedback. The mastery criterion is set at the high level of 95%.

Figure 1. HyperHeart Mastery Learning and Transfer.
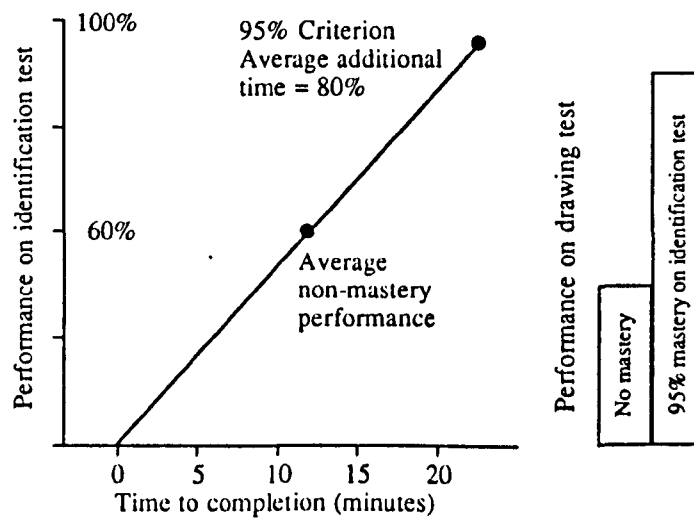
Working with students with learning difficulties I found that, although it took on average 82% longer for students to achieve mastery, they reached levels of performance which were well above those predicted by teaching staff. The high level of performance on the initial test which required students to identify parts of the heart also transferred to a test requiring students to draw diagrammatic representations of the heart.

The interactive mode of the new information technologies, which includes machine simulations of human interactions, provides the key to the success of those approaches which are more effective. The essence of this mode is that it provides opportunities for feedback. The nature of that feedback is crucial. It can vary from the simplest indication, without explanation, that an answer is incorrect, to the full diagnostic interactions which artificial intelligence systems can offer. The primary purpose of the diagnosis in mastery learning is to ensure complete understanding of the taught material. The most effective methods of instruction seem to include these diagnostic interactions combined with mastery conditions.

On this basis, it is not surprising that earlier approaches such as programmed learning methods, which simply report success or failure with minimal interactive features and no diagnosis or mastery conditions, are not very effective. Similarly, computer-based methods which show all the hallmarks of early programmed learning methods, are likely to be no more effective.

It is interesting to note that research confirms that the computer is most successful when interactively simulating real world events or tutoring, both of which capitalise on the provision of feedback (Kulik et al, 1980).

Raising Standards of Performance

These results indicate that the technology for substantially enhancing learning is available and can be delivered using current graphics hardware and software, which is capable of providing corrective feedback and is efficient at handling mastery requirements. What seems to be crucial is the application of our knowledge of educational research rather than the blanket provision of new information technologies in education. Unfortunately, the latter frequently is given priority when non-educational governmental agencies are involved, these ill-informed decision-makers being seduced by the superficial appeal of new hardware and software approaches.

519

Software which fails to incorporate testing which is diagnostic and related to remedial strategies aimed at ensuring high levels of mastery performance, will merely continue to produce the same cumulative deficits and failure, the same low standard of performance, as before. If we wish to raise standards of performance then the solution lies in our views about students and their learning. The many years of media research can offer insights into student learning and these may be used to promote the development of more effective information technologies. However, it will only be when the new technologies succeed in promoting more effective learning, that both teachers and students will change their views (see Bloom, 1978).

REFERENCES

•Block, J. H. and Burns, R. B. (1977) Mastery learning. In L. S. Shulman (Ed) **Review Of Research In Education**. Volume 4. Itasca, Ill.: F. E. Peacock, pp. 3-49.

•Bloom, B. S. (1978) New views of the learner - implications for instruction and curriculum. **Educational Leadership**, Vol. 35, pp. 562-76.

•Chu, G. C. and Schramm, W. (1968) **Learning from television: what the research says**. Institute for Communication Research: Stanford University.

•Clark, R. E. (1983) Reconsidering the research on learning from media. **Review of Educational Research**, Vol. 53 (4), pp. 445-459.

•Cohen, P. A., Ebeling, B. and Kulik, J. A. (1981) A meta-analysis of the outcome studies of visual-based instruction. **Educational Communication & Technology Journal**, Vol. 29 (1), pp. 26-36.

•Dwyer, F. M. (1978) **Strategies for Improving Visual Learning**. Pennsylvania: Pennsylvania State University.

•Glass, G. V. et al. (1981) **Meta-analysis in social research**. Beverley Hills, Ca.: Sage Publications.

•Jamison, D. T., Suppes, P. and Wells, S. (1974) The effectiveness of alternative instructional media: a survey. **Review of Educational Research**, Vol. 44 (1), pp. 1-67.

•Kulik, C. L. C., Kulik, J. A. and Bangert-Drowns, R. L. (1990) Effectiveness of mastery learning programmes - a meta-analysis. **Review of Educational Research**, Vol. 60 (2), pp. 265-299.

•Kulik, J. A., Bangert, R. L. and Williams, G. W. (1983) Effects of computer based teaching on secondary school students: a meta-analysis. **Journal of Educational Psychology**, Vol. 75, pp. 19-26.

•Kulik, J. A., Kulik, C. C. and Cohen, P. A. (1979) A meta-analysis of outcome studies of Keller's personalized system of instruction. **American Psychologist**, Vol. 34 (4), pp. 307-318.

•Kulik, J. A., Kulik, C. C. and Cohen, P. A. (1980) Effectiveness of computer-based college teaching: a meta-analysis of findings. **Review of Educational Research**, Vol. 50 (4), pp. 525-545.

•Kulik, J. A., Kulik, C. C. and Smith, B. (1976) Research on the personalized system of instruction. **Programmed Learning and Educational Technology**, Vol. 13, pp. 21-30.

•Levie, W. H. and Lentz, R. (1982) Effects of text illustrations: a review of research. **Educational Communication and Technology Journal**, Vol. 30, pp. 195-232.

•Levin, J. R., Anglin, G. J. and Carney, R. N. (1987) On empirically validating functions of pictures in prose. In D. A. Willow and H. A. Houghton (Eds) **The psychology of illustration** (Vol.1). London: Springer-Verlag.

•May, M. A. and Lumsdaine, A. A. (1958) **Learning from films**. New Haven: Yale University Press.

•Miller, W. A. (1938) Reading with and without pictures. **Elementary School Journal**, Vol. 38, pp. 676-682.

•Samuels, S. J. (1970) Effects of pictures on learning to read, comprehension and attitudes. **Review of Educational Research**, Vol. 40, pp. 397-407.

•Schramm, W.L. (1964) **The research on programmed instruction**. Washington DC: US Government Printing Office.

•Spencer, K. (1981) Alternative media: a review of the instructional effectiveness. **Educational Broadcasting International**, Vol. 14 (2), pp. 90-93.

•Spencer, K. (1990) HyperHeart: does animated illustration contribute to mastery learning? **British Journal of Educational Technology**, Vol. 21 (3), pp. 227-228.

•Spencer, K. (1991) **The psychology of educational technology and instructional media** (2nd edition). Liverpool: United Writers Press.

•Twyford, L.C., Davis, J.S. and Seitz, C.P. (1954) **Evaluation of the film: military police support in emergencies (riot control)**. Technical Report no. SDC 269-7-52. Port Washington-New York: Instructional Film Program. Naval Training Devices Centre; Office of Naval Research.

•Vernon, M. D. (1953) The value of pictorial illustration. **British Journal of Educational Psychology**, Vol. 23, pp. 180-187.

•Walberg, H. J. (1984) Improving the productivity of America's schools. **Educational Leadership**, Vol. 11, May, pp. 19-27

Note on Effect Size and Meta-analysis

Early research studies tended to report comparative statistics which gave no indication of the size of differences, if any, between the types of instruction. Conclusions tended to be based on the presence or absence of a statistically significant result. When groups of researches were reviewed there was a tendency to use a 'box score' tally approach, frequently resulting in a small number of studies favouring the innovation, a similar number favouring the traditional approach, and the vast majority showing NSD.

Criticisms of this approach to summarising research have been accommodated by the new meta-analytic methods of formulating generalisations (Glass et al., 1981). Meta-analysis is simply the statistical analysis of a large collection of results from individual studies for the purpose of integrating findings. It seeks to determine the effect of a specific treatment, influence or intervention from a large variety of individual studies on a particular subject. The meta-analytic procedures yield effect size estimates which are converted to 'percentage of standard deviation gains' due to the more powerful treatment, if any.

The effect size (ES) is a measure of the educational importance of any performance changes produced. An ES of 1.0 means that the innovation increases the performance of the group by an amount equal to one standard deviation unit of the control treatment. This would take the average student from a position in the middle of the traditional group to the position occupied by the top 20% of that group. This is a large effect and is educationally significant. An ES of 0.2 is a small or trivial effect, whereas an ES of 0.5 shows a moderate effect, produced by a treatment which certainly merits further investigation.

11. Spencer, K.A. (1996) Recovering reading using computer mastery programmes, British Journal of Educational Technology, 27(3), 191-203.

Recovering reading using computer mastery programmes

Ken Spencer

Ken Spencer is Lecturer in Educational Technology at the University of Hull. Address for correspondence: Ken Spencer, Department of Educational Studies, The University of Hull, Hull HU6 7RX. tel: +44 1482 465406; fax: +44 1482 465406

Abstract

Mastery learning has been shown to be an effective method of raising the level of performance of pupils on a wide range of tasks. This method has been applied to the teaching of reading and spelling with primary school children with learning difficulties, in the form of a series of computer programmes using the high quality digital sound which is available in the current generation of low-cost portable computers.

Unlike many programmes designed to teach or encourage reading skills, the programmes developed for this project provide instantaneous feedback, and identify and reward correct responses, with high criterion levels set for progression from one task to the next. They also provide complete records of pupil performance, allowing teachers to select teaching sequences appropriate for each individual pupil. Case studies of pupils are presented, showing the performance gains obtained with short but regular use of the programmes, over a 12 week period.

Introduction

When reviewing the literature on educational technologies and methodologies (Spencer, 1991) I suggested that the poor performance associated with a variety of basic skills, such as reading, could be raised by amounts considered to be educationally significant using a number of different media employing mastery methods, such as Learning for Mastery and the Personalised System of Instruction.

Reviews of research into the effectiveness of different instructional strategies consistently demonstrate that these mastery methods offer the best opportunities for raising performance, usually with an effect size of 0.5–1.0. Walberg (1984) concluded that mastery related reinforcement techniques contribute significantly to the most effective

Table 1: Modes, media and methods: summary of effect sizes

Small Effect Sizes (Range 0.1–0.3)	Effect size	
Individualised instruction	0.10	*
Ability groups	0.10	*
Visual-based instruction	0.15	**
Programmed learning	0.16	**
Audio-Tutorial	0.25	***
Assigned homework	0.28	***
Computer Assisted Instruction	0.29	***
Moderate to Large Effect Sizes (Range 0.4–7 and above)		
Tutoring	0.4	****
Video feedback	0.41	****
Computer Simulations	0.49	*****
PSI	0.58	******
Graded homework	0.79	********
LFM	0.80	********

*represents the size of effect in tenths of a standard deviation

teaching strategies, with high criterion levels (90–100%) offering the largest effect sizes (Kulik *et al*, 1990).

Unlike many programmes designed to teach or encourage reading skills, the reading programmes developed for this project provide instantaneous feedback, and identify and reward correct responses, with high criterion levels set for progression from one task to the next. For example, in Figure 1 the pupil has correctly entered the first letter of the word. This is retained on the screen, as are further correct responses. This is crucial in developing pupil self-confidence. Instead of waiting until the pupil has incorrectly spelled the full word, at which point s/he is told that s/he is wrong (as many modern programmes continue to do), each response is monitored and corrected as necessary.

Summary of the underlying theory

The HyperReading programmes are designed to teach reading skills indirectly through programmes which develop spelling skills. A more direct route would be to utilise computer speech recognition features, but as van Daal and van der Leij (1992) suggest, adequate speech recognition facilities to evaluate reading-aloud responses are only available at prohibitively high prices. However, developments have made it possible to provide high quality speech via relatively low cost computers, making them more suited to assist in spelling type programmes rather then in reading exercises. Reading skills benefit from this approach because a critical step in reading acquisition is made when children shift from a visual strategy for reading words to an alphabetic strategy, which is first learned in spelling and later carried over to reading. This is a crucial phase because in the course of spelling development, children learn more about how sounds relate to letters (Frith, 1985).

The concept of mastery can be traced back from Bloom's formulation in *Learning for Mastery* (1968/81) to Carroll's seminal paper *A Model of School Learning* (1963). Central to this is the concept that by using appropriate strategies most children (and by this Carroll means 95%) can attain levels of performance similar to the top 20% in an average classroom. Carroll's model, which forms the basis of Bloom's LFM, places emphasis on the amount of time available for study and the time needed to completely master a given task. Clearly the amount of time needed is related to the ability of the pupil, but also to the ability of the teacher to provide appropriate instruction.

Mastery methods are often considered to be close to the behaviouristic approaches suggested by Skinner (1954) and the computer programme described below may look uncomfortably like a Crowder branching strategy from the 1960s (or even a Skinner linear programme). It is, however, rather more sophisticated and responsive. Skinner suggested that a country which could produce quantities of washing machines as labour-saving devices surely could produce teaching machines on the same scale. My latest washing machine purchase is generations ahead of anything available in the 1950s: it is microprocessor controlled, and I simply select one of a large variety of options and the machine carries out its complex task of adjusting water volume and temperature, spin speeds and drying temperature. The modern computer with appropriate teaching programmes offers similar advances over earlier designs. However, such computer programmes are only one medium out of many that can be employed in the classroom for the teaching of reading eg, paired reading, or any method which places the learner in direct contact, on a regular basis, with an expert reader who encourages mastery

The value of a systematic method of teaching reading may be called into question and it is true that the majority of pupils acquire their reading skills through a variety of less highly structured approaches. However, for children who are failing to get started a highly structured approach may prove to be more effective. The mastery methodology employed may also be considered to place too much emphasis on rote learning, rather than general (problem-solving) strategies. However, the problems of illiteracy must be attributed not only to learner, or even teacher, deficiencies, but also to the written form of the language: written English is profoundly illogical, especially the most common 200 words, and the rules, such as they are, cannot always be applied in a regular manner. The Chinese have a similar problem with their written language and adopt a simple rote learning strategy as the most effective method, as do the Japanese. This approach, I believe, can be adopted as one strategy for pupils who are failing to develop their English reading skills.

Early models of the HyperRead computer programme
The Macintosh computer has built-in sound recording and playback facilities, making it suitable for language work, and runs the HyperCard application which presents information on the computer screen as a series of cards, rather like pages in a book. The first approach was to convert the individual child's reading book into an electronic version. Such a version could read the page to the child, highlighting

each word as it was read. The basic idea behind this was that many of the children who were failing in their reading simply did not get sufficient support outside school hours: the computer "reading" could compensate for this. Davidson and Noyes (1995) confirm this in their study, which was based on a similar premise: "Children who read at home regularly simply get more practice than those who don't" (Topping, 1985). Their programme "read" the child's school book, providing an additional hour of reading each week over a 4 week period. The results showed no significant difference between reading the book to a teacher and reading it with the computer providing help when necessary.

The initial HyperRead programme also included a mastery element. The pupil was required to take part in a short test after each page had been read. This consisted of presenting to the pupil a list of the words that had just been read. The pupil was required to select a word which the computer "spoke" (eg, "Now choose the word car.") Those words which were incorrectly selected were placed in a "mastery loop". The mastery loop was a method which gradually eliminated those words which were correctly identified but kept the incorrect words, which were recycled until all words were correctly identified. Thus, after each page was read by the computer, the pupil was active and could not move on to the next page until all the words had been mastered.

This phase was only partially successful. I have noticed that with many talking books, which run from CD-ROM on modern computers, children are easily distracted from the screen and do not follow the laboriously highlighted words, often only turning to the screen when a dramatic sound effect is heard. Our pupils also became bored with the constant demand of responding to words that they knew, before being allowed on to a new page.

A refinement was to set up a database with a record for each pupil. The record stored the performance of a reading test (with mastery loop) which included all the words to be used in the electronic version of their reading book. When they had mastered the words in the list, the programme took them to the appropriated electronic book and read it to them (no test being applied), providing a reward for their effort and reinforcing their learning. This was moderately successful, but again it was noticed that the pupils tended to be easily distracted when the electronic book was being read—in fact, the part which did fully occupy their attention was the testing phase: they really enjoyed this, especially when appropriate computer game noises were added.

In consultation with a group of primary school teachers it was decided that if a programme could be devised that effectively taught specific words, the children would then be able to read their books without assistance, making the talking computer book feature redundant. The challenge was to devise such a teaching programme.

The teachers consulted felt that many features, which they employed when teaching reading and spelling, should be built into the computer programme:

- The teachers entered lists of words directly into the computer, so that individual schemes of work were developed.
- Two types of test were developed: a recognition (look-and-say type) test and a spelling test.
- The recognition test was designed so that the number of alternative words presented varied from 2 to 6. The pupil saw a group of words in boxes on the screen and was asked by the computer to choose a particular word (Choose "boy"). When the correct word was clicked a small animated jack-in-the-box appeared. Incorrect boxes were marked, gradually eliminating the incorrect responses.
- The spelling test offered control over the way each word was displayed. The word could be shown for several seconds before the pupil responded. This could be varied in 0.5 seconds steps from 0 to 10 seconds.
- The number of letters of the word to be spelled could also be varied, ranging from one letter to the complete word. This feature also allowed for forward or backward chaining (ie, starting from the first or last letter) and for particular groups of letters within words to be targeted.
- A "virtual" keyboard was created on the screen and pupils responded by pointing to a letter and clicking it as appropriate. The letters available on the virtual keyboard could take many different forms (eg. all the letters of the alphabet could be shown either in QWERTY form or ABC form; alternatively, only those letters required to complete the task could be shown).
- Errors in spelling were identified immediately a response was made, and the pupil offered a number of additional attempts, until mastery of the word was achieved. The number of additional attempts could range from 2 to 10, but usually was set at 3. This is shown in Figure 1.
- A mastery loop was applied to each group of words under test. Groups of words which failed the mastery criterion level on the first attempt were repeated until the criterion was reached for the whole group. This is shown in Figure 2.
- A limit could be imposed on the number of attempts permitted in the mastery loop, both for individual words and groups of words. It had been noticed that a mastery loop without such a cut-out response could produce despair in children who were required to work endlessly at the same words.
- A complete record of each pupil's responses was logged on their student record card, to enable the teacher to decide upon further appropriate action.
- A short audio-visual reward, consisting of one of several animated characters dancing to a tune, appeared after each successful session.
- The pupil card could be set into one of two modes: the pupil mode showed minimal information (pupil name and start button) and could not be altered; the teacher mode allowed various setting to be changed.

The computer programme, which included all the above features, was used within the first phase of the Portable Reading Laboratory project supported by the National
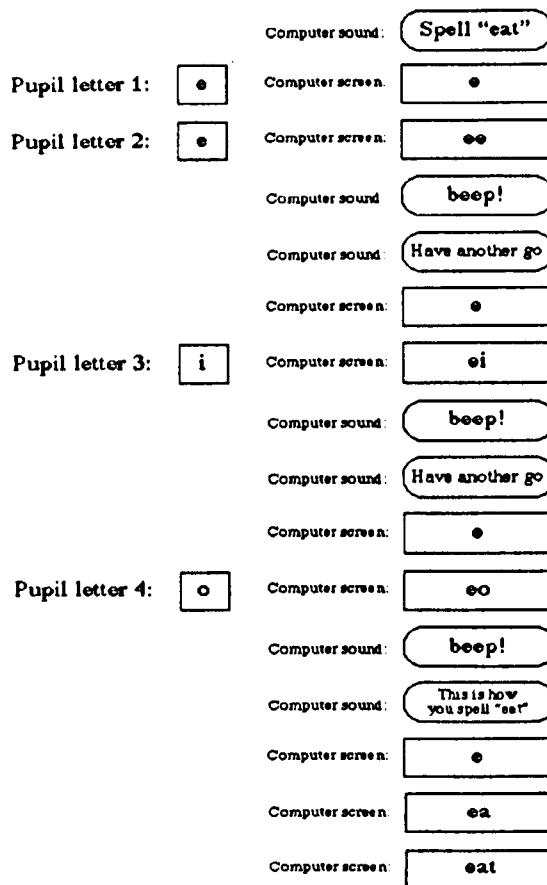
Figure 1: Mastery for individual words

Council for Educational Technology *Portable Computers in Schools* project (Bowell, France & Redfern, 1994).

During this time the programme was in general use in two Humberside schools, running on 8 portable computers, and a small group of children was studied intensively. A major problem emerged early in this phase: teachers had demanded a large array of controls to apply to each individual child's interaction with the computer, but with increasing demands being made on their time they could not always make the necessary adjustments to the individual pupil record cards.

For example, a group of words may be required to be spelled with a brief glimpse of the word before each was spelled (called the "hints" condition). Following this, the words would be required to be spelled without any hint. The teacher had to continually reset the hints control for each pupil, meaning that the pupil was constantly demanding the teacher's attention.

Criterion level: 90% on first attempt.

Failed to reach criterion on first attempt:
repeat group of words

Criterion level: 90% on first attempt.
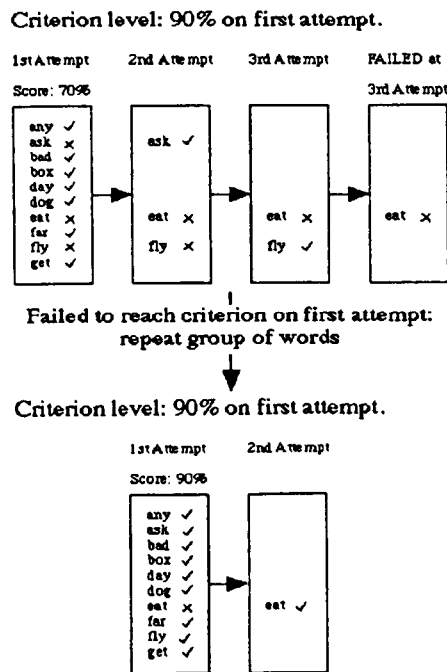
Figure 2: Mastery loop for word groups

However, the results from this initial phase were very promising. For example, one 9-year old child who could not read the most common word in the English language rapidly learned to read and spell the most common 10 English words, using the programme for 12 minutes a day for 1 week.

These results were so encouraging that a new approach to the setting up of the control features for each pupil, designed to save teacher time, was adopted. This consisted of separate files (called sequencer files) which have the details of a particular sequence of control settings which can then be applied to a particular pupil record file (Figure 3). The pupil record file loads up the relevant information from the sequencer file (instead of having the teacher reset it), goes to the relevant test file, which records the pupil performance and calculates a performance score. In most of the early research the pupils started at the first card of each new sequence and gradually moved from simple to more complex tasks. This is the general pattern for the following data.

The introduction of the sequencing feature led to the need for control of the time a pupil spent working on the automated sequence. The time for study was set on the pupil card, and used the computer clock to determine how long the pupil had worked with the materials. After a preset time, usually 10–15 minutes, the session was ended, and another student took over.

The pupil record card logged the performance of the pupil, noting the sequence and
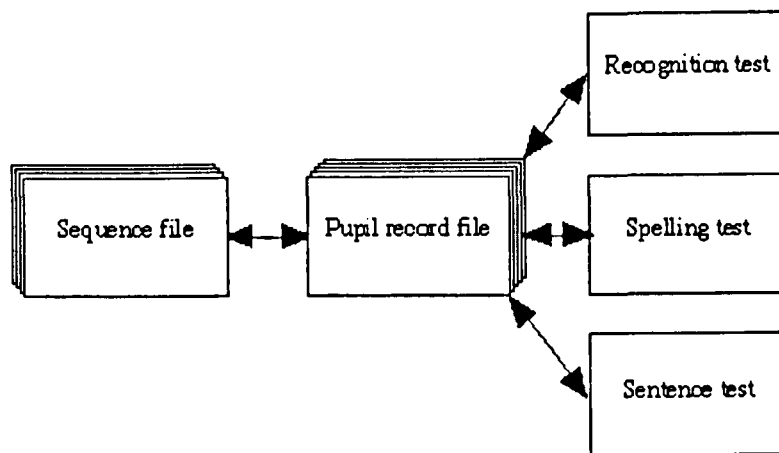
Figure 3: Sequence, pupil and test files

card level at the end of a session. The pupil's next session, usually the following day, started from this point.

Research results
The following case study shows the effectiveness and limitations of the computer mastery approach to the teaching of reading.

Pupil S was mentioned earlier in this report. He was 9.5 years old when he started using the computer programme. At school he had failed to establish any reading skills. He had experienced a very difficult home life and received no help with his work from his family. He was the youngest child of six, the parents had separated and his natural father had died the year before. He was sharing a bedroom with his elder brothers, at least one of whom takes drugs. In school he was frequently tired and was hard to keep on task, his level of interest being difficult to raise or maintain. He was at Stage 3 on the Special Needs Register and is to be considered for Stage 4 (referral for Statement) when he is next seen by the educational psychologist. On the WISC (R) test he was found to have a verbal IQ of 82, placing him in the below average range of ability.

He was first tested on his alphabet and had no problems, and no identifiable letter confusions. When tested on the most common 120 words (Reid, 1989) his score was very low (2 words). He demonstrated his ability to learn several words during an intensive investigation with the early version of the programme and would work for 10–15 minutes before becoming distracted.

In the programme that was devised for him the most common 120 English words were broken down into groups of 10 words, starting with the most common 10. Within these groups word lengths were mixed, varying from 1 to 6 letters, with an average of 3 letters for the first 50 words. The auto time facility was set at 12 minutes, giving
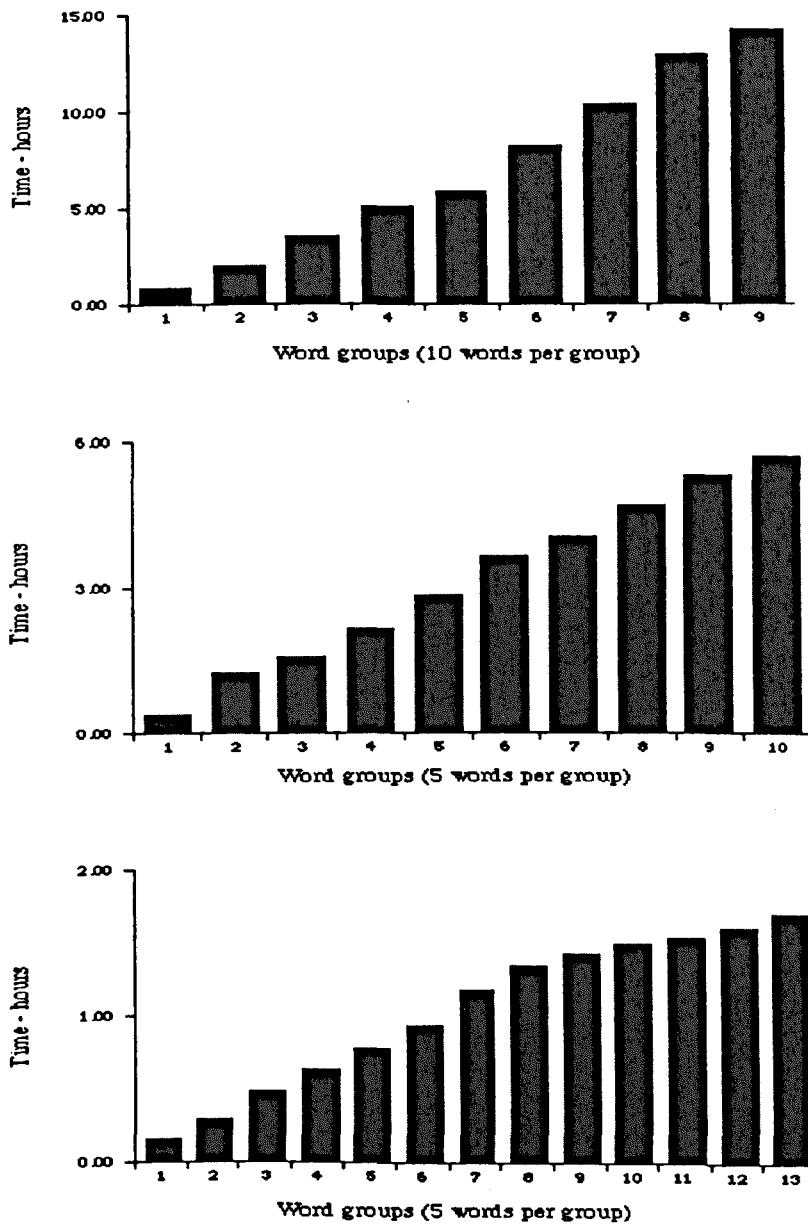
Figure 4: Cumulative time for completion of word groups

12–15 minutes daily work with the computer programme. Twenty cards made up a sequence for each group of 10 words, each card containing a separate set of control features. The class teacher arranged for the pupil to work on a daily basis with the programme and the pupil was given responsibility for setting the equipment up,

switching the computer on, choosing the appropriate pupil record card, and working through the day's sequences. Full details of the pupil's responses were automatically logged, together with times and scores for each sequence.

Figure 4(a) shows the results for the first 9 groups of words. It represents the cumulative time spent working with the programme. Each bar represents the total amount of time taken to learn the particular group of words and the preceding groups. The difference between one bar and the next is the amount of time spent learning that particular group of 10 words.

The cumulative time curve shows a steady work rate, the time being a little longer towards the end of the sequence. The average rate of completion was 6.4 minutes per word for the first fifty words, and 9.5 minutes per word for the first 90 words.

This an impressive acquisition rate for a child who at the start of the teaching programme could read only 2 of the most common 120 words ("I" and "a"). A month later, having worked on the programme for an average of 16 minutes each day, he could read 50 of the words.

Pupil S spent a total of 14.08 hours, during a 12 week period, working with the computer programme and mastered the reading of 90 words. When tested on the whole 90 words his reading score was 89%; and when retested 8 weeks later, after the summer vacation, his score was 67%.

However, after the summer holiday pupil S was placed in a new class. The computer programme was available on a computer exclusively used by the class, and available particularly to pupil S. Although in the previous class the pupil had been responsible for setting up the computer and working on the programme, this did not happen in the new class. Instead of an intensive daily session with the programme, the pupil used it very occasionally (1–2 times per month). After 7 months of normal classroom activities, with sporadic use of the computer programme, his performance had only reached 69%. There is some consolation here in the knowledge that mastery programmes do produce learning that is resilient to forgetting (possibly because of the over-learning of some of the content), as demonstrated by Kulik *et al*'s (1990) meta-analysis.

Pupil C was 6.25 years old when he started working with the computer programme. He was the second of three children in the family and there was no apparent encouragement or stimulus for learning at home. He seemed to spend most of his time watching "endless videos" according to teaching staff. Pupil C and his younger brother have major problems with speech, although their elder sister does not have such problems. The father attended special school in his youth. Educational psychologists noted the discrepancy between pupil C's ability, which was average with an cognitive index of 94, and his poor understanding of language, which was coupled with a severe phonological problem. This had not responded to speech therapy, and the pupil was finding it increasingly difficult to access the National Curriculum at a

level appropriate to his ability. His attention span was limited, especially for activities that involve long periods of listening, which was consistent with his limited understanding of language. A programme of daily speech therapy implemented by a teaching assistant was stopped following financial cut-backs, since which time his progress had been classed as minimal. The pupil was statemented prior to the start of the computer project and it was agreed that he should have a structured language programme on a frequent and regular basis, with the school providing for this by the use of the laptop computer for basic sight vocabulary and spelling.

The results from earlier use of the computer programme scheme led to a new strategy which involved learning groups of 5 words rather than 10. A word recognition task with 100% criterion level was followed by spelling tasks in which the pupil was required to spell the first letter of each word, followed by the second, third and fourth letters. For each of the four letters, two conditions were used: the pupil was first shown the letter (hints condition) before responding; the pupil was then asked to spell the letters without any hints. For pupils with low sight vocabulary this method brought about a gradual learning of the first 4 letters of the words. Pupil C's sight vocabulary test score with the 120 most common words was 62%. Figure 4(b) shows the rate of acquisition of those 50 words with which he was not familiar. Working an average of 13 minutes per day on the programme he made rapid progress, acquiring 50 words in 5.67 hours (a rate of 7 minutes per word). When finally retested on the 120 words his score was 114 (95%).

Pupil J was identified as having learning difficulties with reading, writing and spelling when he was a pupil in Key Stage 1. His class teacher worked with the school's SEN co-ordinator to provide support strategies. He scored poorly in KS SATs obtaining level 1 in reading, writing and spelling. In 1992 a teacher's aide was appointed for 5 hours per week to assist the pupil with learning programmes established by schools staff in liaison with the learning support service. The pupil's parent was very concerned about the lack of progress, given the close liaison between home and school. The educational psychologist had been involved in providing professional advice concerning the pupil's uncooperative behaviour at home. An initial statement of the pupil's special educational needs was made in 1993 and the teacher's aide continued to support the pupil for 5 hours per week (1 hour daily).

The pupil has a history of being moody and uncooperative, making individual targets difficult to implement. He responds well to things he enjoys, and he enjoys working with computers, but otherwise can be extremely difficult to motivate. He is articulate and is capable of expressing himself orally. He has an exceptionally good vocabulary. The most recent psychological tests indicate the discrepancy between his receptive language and literacy. Using the British Picture Vocabulary Scale (short form) his performance is rated as 14.9 years, but his basic reading (word recognition) is 7.3 years and reading comprehension is 7.0 years.

The computer programme was running in the school with all children in the class using the programme, in order to identify the general benefits of such an approach to

teaching reading and spelling to the full range of abilities. Pupil J responded very positively to the programme, in fact, at times more positively than to his teacher's aide. This responsiveness to the computer programme led to him receiving a regular session of 15 minutes per day with the computer, under the supervision of the assistant.

Figure 4(c) shows his performance with the most common 65 words. The initial strategy was the same as for pupil C and resulted in 45 words being covered in 1.33 hours (1.8 minutes per word). This was considered by the teachers to be excellent, given his temperamental nature: he worked consistently and enthusiastically on task for each 15 minute session. However, he did indicate that he was getting bored with the procedure of slowly building up each word and an "express" version was prepared for words 41 to 65, in which he was required to spell the first 4 letters of the word, without working through letters 1 to 3, initially with a 2 second hint and then without a hint. The rate of working through the words increased dramatically, as can be seen in the graph: 25 words in 21.6 minutes (0.9 minutes per word).

Pupil J had a particular difficulty with reading and spelling non-phonetic words such as "people" and "because". When he was given the opportunity to work out a method (sometimes complicated) for memorising these particular words, with the gentle correction from a non-human agency, his approach to his work changed dramatically. Prior to the introduction of the computer programme he was not co-operative when tested for his reading of the 120 most common words and a baseline could not be established. However, on completion of the programme his confidence and co-operation had improved: he could read all 120 words and could spell 93% of them.

Upon completion of the basic sight vocabulary words, a new programme was designed which required the pupil to write out whole sentences, based on the most common 200 words, using a similar strategy to that employed in the spelling of individual words. He has worked on a regular basis with this programme for two terms and now has little difficulty in completing sentences 6-8 words in length.

Discussion: Mastery and SEN
Pupil S's case demonstrates the validity of the mastery approach. For many years appropriate instruction was not provided, and the result was that he failed to master any reading skills and consequently made little progress. By providing an appropriate strategy, and sufficient time on task, rapid progress was made. When this strategy was discontinued progress ceased, although the attained level of performance was substantially maintained, supporting the data provided by Kulik *et al*'s meta-analysis (1979), which clearly demonstrates the strength of the learning under mastery conditions.

Pupil C also had very specific needs and these were also met by the same programmed characteristics as those of Pupil S. Pupil C's problems were identified at an earlier stage than Pupil S, and his needs will continue to be met by a combination of the computer programme support, now featuring 400 basic words (and sentences made from the

basic words), and daily support from a teacher's aide. Under these conditions he has already mastered a basic sight vocabulary and is expected to continue to make good progress.

Pupil J's needs were rather different and required an alternative mastery approach, which was more specifically directed at his need for support at a higher level. Having gained confidence with the mastery of reading and spelling of the basic 120 words he continues to make progress with the most common 200 words by building up complete sentences within a new computer module.

In all the above cases, as predicted by Carroll and Bloom, teachers indicate that pupil self-esteem appears to have been raised, which in turn has had an effect on pupil perseverance, which Carroll identifies as a key factor in the amount of time a pupil will actually spend on task. This sets up a self-sustaining cycle of improvement: appropriate tasks, with corrective feedback, lead to mastery performance, which raises self-esteem and makes the pupil more prepared to invest more effort in learning.

The computer programme is not currently sufficiently intelligent to select the type of structure most suitable for a pupil, but this feature will be built into the system in the near future. With such an algorithmic artificial intelligence in place the programme could contain 12–24 months of work for pupils, with full details of performance available to the teacher, even to the extent of suggesting appropriate reading books for the pupils!

References

Bloom, B S (1968/81) Learning for mastery *in* Bloom B S *All Our Children Learning* McGraw-Hill, London, 153–177

Bowell, B, France, S and Redfern, S (1994) *Portable Computers in Action* National Council for Educational Technology, Coventry

Carroll, J (1963) A model of school learning *Teachers College Record* 64 723–733

Davidson, J and Noyes, P (1995) Computer-generated speech-feedback as support for reading instruction *Support for Learning* 10 1 35–39

Kulik, C L C, Kulik, J A and Bangert-Drowns, R L (1990) Effectiveness of mastery learning programmes—a meta-analysis *Review of Educational Research* 60 2 265–299

Kulik, J A, Kulik, C C and Cohen, P A (1979) A meta-analysis of outcome studies of Keller's personalised system of instruction *American Psychologist* 34 4 307–318

Reid, D (1989) *Word for Word* LDA, London

Skinner, B F (1954) The science of learning and the art of teaching *Harvard Educational Review* 24 2 86–97

Spencer, K A (1991) Modes, media and methods: the search for educational effectiveness *British Journal of Educational Technology* 22 1 12–22

Topping, K (1985) Parental involvement in reading: Theoretical and empirical background *in* K Topping, K and Wolfendale, S (eds) *Parental Involvement in Children's Reading*, Croom-Helm, Beckenham

Treiman, R (1984) Individual differences among children in reading and spelling styles *Journal of Experimental Child Psychology* 37 463–477

Van Daal, V H P and van der Leij, A (1992) Computer-based reading and spelling for children with learning disabilities *Journal of Learning Disabilities* 25 3 186–195

Walberg, H J (1984) Improving the productivity of America's schools *Educational Leadership* May 19–27

12. Spencer, K.A. (1998) Predictive models of spelling behaviour in 7- and 11-year-olds, Journal of the Simplified Spelling Society, 24(2), 23-27.

Ken Spencer

Predictive Models of Spelling Behaviour in 7- & 11-year-olds

Ken Spencer is Lecturer in Media Studies and Educational Technology in the School of Education, Hull University, Hull HU6 7RX, UK. He may contacted by email at <k.a.spencer@educ.hull.ac.uk>. The statistical Tables are given at the end of the article.

0 Abstract

A predictive model for spelling is suggested, based on the results of spelling tests taken by 2,684 seven- and eleven-year olds, in 1996. The tests were part of a national scheme of testing for the School Curriculum & Assessment Authority (SCAA). The factors identified as influencing the number of children correctly spelling a word are: word length (number of letters), frequency of a word, and a measure of the word's phoneticity. A measure of the most infrequent form of representation of the phonemes in a word (the 'trickiest' phoneme) is a strong factor with 7-year-olds, whereas a measure of the average phoneticity of a word is a better indicator of word difficulty for 11-year-olds, who are susceptible to the mitigating effects of high word frequency on irregular spellings.

1 Learning for mastery

Carroll's model of school learning (1963) suggests that most pupils are capable of reaching levels of performance more usually associated with the top 10-20% of the school population. He proposed that a number of variables could be manipulated to increase levels of performance, such as the quality of teaching materials and the amount of time available for learning. Some variables associated with the pupil, such as perseverance, may be difficult to manipulate.

Carroll's model has been applied to teaching by Bloom and his associates, in the methodology known as Learning for Mastery, which places great emphasis on formative testing in order to determine deficiencies in either learning or teaching. An essential within this system is the requirement that high levels of performance are demanded at the early stages of learning, which ultimately result in higher overall levels of performance. Required criterion levels are as high as 100%, although more usually they range between 80–90%.

2 Case-study in remedial literacy

This approach has recently been applied to the teaching of reading (Spencer, 1996), using computer-based learning techniques. In the case of one pupil, who had reached the age of 10.5 years without being able to read the most common word in the English language, decisions had to be made concerning the teaching of the most common 100 words: should a phonics approach be adopted, or a method based on gradually increasing the demands in a simple spelling exercise. Many of the most common words fail to obey even the most rudimentary rules and so the simple spelling approach, with increasing mastery demands, was adopted. Practising for 10-15 minutes per day the pupil mastered 80% of the words over a period of 12 weeks.

It was clear from the performance of this pupil that much of his problem was associated with the vagaries of English spelling — he simply gave up when he applied rules to common English words and was told that he was wrong: the rules didn't work, and what should have been a simple task proved impossible. With the continuing concern of parents, teachers and politicians about the levels of literacy in the UK, the question arises: are we disadvantaging our children by making a simple task incomprehensible?

3 Searching for models

Of particular interest to researchers investigating the application of computer-based literacy systems is the search for a predictive model of spelling performance. Knowledge of such a model would indicate the factors that make words difficult to spell; determine if they are the same for all ages; and may indicate how strategies change with age, to make spelling more accurate. This, in turn, has implications for reading. Frith's (1985) six-step model of literacy development suggests that there is an initial period when children use a logographic strategy to read and a phonological strategy to spell, ie, they read and spell in different ways. According to this model, the emergence of phonemic representations in spelling leads to advances in later reading. Rego (1991) demonstrated that the ability to spell non-words is strongly related to progress in reading, and this has been confirmed by Lazo et al. (1997), who show that early attempts to read words are strongly related to the progress made in spelling, as early attempts to spell words influence later reading.

The following analysis, which identifies several models for learning to spell, is based on data from national tests carried out by SCAA in 1996. SCAA's activities have recently been incorporated into the Qualifications and Curriculum Authority (QCA) whose statutory functions were set out in the Education Act, 1997. Principally, by its forthcoming review of the school curriculum, QCA hopes to define the structure and content of teaching and learning that will enable all pupils to develop and demonstrate their knowledge, skills and understanding. QCA's functions and responsibilities include: developing learning goals for the under-fives; accrediting assessment schemes for children entering primary school; monitoring and reviewing the National Curriculum and its assessment; continuing with the development of national assessment at the ages of 7, 11 and 14.

4 Factors that may be relevant

When attempting to build a predictive model of behaviour the researcher usually has in mind factors that may be relevant. In the case of spelling-accuracy, the present approach looked at the following factors:

> number of letters in the word
> phoneticity of the word
> frequency of the word

4.1   Length of the word

This is a simple measure and was included because in the initial stages of spelling (and reading) 7-year-olds are still developing short term memory strategies, and any lapses in memory are likely to manifest themselves with longer words. Longer words also give more opportunities for errors.

4.2   Phoneticity of the word

This is seen as a major factor, but one that can be defined in a variety of ways. There is no standardized way of measuring this factor and a number of approaches were adopted and refined.

1. Phonic Ratio

The first approach is to look at the individual letters of the word and measure the degree to which they correspond to a simple alphabetic representation (as in the word *hat*). This is expressed as a ratio of the number of letters pronounced as in the simple alphabet, divided by the total number of letters. The word *hat* has a simple phonic ratio value of 1; *boat* has a value of 0.5; and *shout* a value of 0.2 (see Table 1). This is a crude method, only accounting for sounds represented by single letters, so it will be less powerful at predicting than other measures. The astonishing thing is that, for 7-year-olds, it is a predictor at all.

2. Phoneme frequency measures

It was recognized that simple phoneticity might be a factor with younger children. A more sophisticated measure was also developed which could be applied to both 7- and 11-year-old age groups. Children learn at an early age that a variety of representations can be used for the same sound and, as SCAA recognized, the difficulty is less knowing the patterns than knowing which pattern to use in each individual word. In order to establish the range of representations of the phonemes that make up the English language, and the frequency of each representation, the 3,500 most common words from the LOB Corpus (Hofland and Johansson, 1982) were analysed.

The phonetic representation of each of the 3,500 words was determined from the *Oxford English Dictionary* (Second Edition, CD-ROM version) enabling the standard alphabetic representation of each phoneme to be determined for each word. With each phoneme coded, tables showing the various forms of representation for each phoneme were extracted. The average number of representations per phoneme was 5.95 (a total of 262 for the 44 English phonemes used in the O.E.D.). Of course, some phonemes have relatively few forms of representation, while others have many more. As for frequency, the common phoneme /ɪ/ (as in *hill* /hɪl/) represents 9.64% of the sample phonemes (total = 20,197); and the infrequent /ʒ/ (as in *visual* /vɪʒʊəl/) represents only 0.13%.

Knowing the different representations of each phoneme allows two measures of the frequency for each form of representation. The first is the proportion of the particular representation for that phoneme (PhR); the second is the frequency of the particular representation in relation to the total number of phonemes in the LOB corpus (PhT), thus showing how often it occurs in running text.

2.i   Representation as a proportion
> *of all representations of the phoneme (PhR)*

This measure considers a particular representation of a phoneme only in relation to other representations of that phoneme. Percentage values for all representations of the phoneme total 100%.

Table 3 shows the values for the /e/ phoneme (as in *den* /den/). This phoneme represents 3.36% of all the phonemes in the sample. The most common alphabetic representation is E, and this is found in 90% of cases (PhR value of E representation for /e/ phoneme). The rare form AI has a PhR value of 0.6%.

2.ii   Representation as a proportion
> *of all phonemes in the LOB sample (PhT)*

This measure is necessary because phonemes occur at different frequencies, and the difference between the most common phonemes (/ɪ/ at 10%) and least common (/ʒ/ at 0.1%) is considerable. In terms of the total number of phonemes in the sample, an infrequent form of a common phoneme may be encountered more often than the usual representation of a less common phoneme. The percentage values for PhT for a particular phoneme will add up to the frequency of that phoneme in the total sample. Table 3 illustrates this: the total for PhT is 3.36%, which represents the frequency of the /e/ phoneme in the total sample.

Knowing the frequency of each form of representation for each phoneme allows an average phonetic value to be calculated for each word. This value can be calculated for both PhR and PhT (see Tables 1 & 2).

In addition, particularly unusual phonemic representations can be identified. In the case of the data for 7-year-olds, the most infrequent form of representation was determined within each word, giving a value for the 'tricky' phonemes, in terms of PhR values, eg, /ɒ/ as represented by AU only occurs in 0.32% of cases and is the trickiest phoneme representation in the word *because*, since all the other representations have a higher value than this.

4.3   Frequency of the word

The frequency of word occurrence also seems likely to influence the spelling and reading of words: the more common a word, the more likely it is that the form will be internalized by the learner. The LOB corpus provides an ordered list of the most common 7,000 words. The total number of occurrences of a word within the entire corpus (1,000,000 words) is also given; this absolute frequency was used as a factor in the analysis.

4.4 Spelling scores and probabilities

The spelling test for 7 year-olds (Key Stage 1) was taken by 1,184 children working at level 2 from SCAA's Schools Sampling Project, a national representative sample of schools taking part in a longitudinal monitoring survey. The test for 11-year-olds (Key Stage 2) was taken by 1,500 pupils from the University of Bath's sample for the 1996 Standard Assessment Tasks (SATs). The data available from SCAA were in the form of percentage correct scores for each word. This score was converted to a probability value for use in the regression analysis. The following formula was used:

Log_{10} (probability right/probability wrong).

5 Analysis of data

The analysis of the data was undertaken with the multiple regression module in the Statistical Package for the Social Sciences (SPSS, version 6.1.1 for Macintosh computers). Regression methods utilize the presence of an association between two variables to predict the values of one from those of another. The regression analysis attempts to predict the spelling behaviour of the two age groups from characteristics (frequency, length, phoneticity) of the words.

5.1 Results of the Multiple Regression Analysis for 7-year-olds

Initial consideration was given to the more obvious factors that are likely to affect the spelling performance of 7-year-olds: the number of letters in word (LETTERS) and the simple phonic ratio (PHONIC), as given in Table 1. The results are given in Table 4, which shows highly significant correlations between the standardized spelling score (LOGPROB) and the two factors. The regression analysis shows that the more powerful predictor is the number of letters in the word. When the absolute frequency of words was included in the analysis no significant correlations were found for that factor; for 7-year-olds, frequency of the selected words does not appear to influence spelling-accuracy.

A second analysis, using more detailed information about the phonetic structure of the words (Table 1: Average PhR; Average PhT; and Tricky phonemes), was conducted. Significant correlations were not found for either PhR or PhT, but the so-called "tricky" phonemes factor was highly correlated (0.77) with the standardized spelling score (Table 4). The analysis demonstrated that the 'tricky' phoneme factor was a more powerful predictor than the simple phonic ratio used in the initial analysis. Both factors contribute in an equal but opposite way in the prediction of spelling scores. The words selected for the 7-year-old tests are not as complex as those for the 11-year-olds; and the 'tricky' phonemes measure identifies those words with particularly unusual spelling features. This factor is exemplified in the contrast between the word *hat*, in which the greatest uncertainty is in the T representation (T=95.90%; TT=3.6%; ED=0.5%), and *friends*, in which the IE is a unique representation (E=90.6%; EA=7%; A=1.2%; IE=0.6%; AI=0.6%). The greater the uncertainty in the representation of the

phoneme, the lower the spelling score. The results of the test for 7-year-olds show that the predictive model has 2 factors: number of letters in the word and degree of difficulty of representation (as measured by relative frequency of occurrence) of key phonemes.

5.2 Results of the Multiple Regression Analysis for 11-year-olds

The words used in the 11-year-old test (Table 2) are more complex than those in the 7-year-old test: they have, on average, 2 additional letters; and some words have several phonemes with rare forms of representation. Table 4 shows those factors which have significant correlations with the standardized spelling scores (LOGPROB) for 11-year-olds: absolute frequency of occurrence (FREQABS) in the LOB corpus; number of letters (LETTERS); and the average frequency of phonetic representations as a function of the total number of phonemes (PhT). The predictive value is almost identical to that obtained with the model for the 7-year-olds. The regression equation shows all factors contributing to spelling behaviour in a similar manner, but with number of letters acting in the opposite direction. This model suggests that the spelling behaviour of older pupils, when responding to more complex words, will deteriorate for less common words that are longer and use unusual forms of phonemic representations.

6 Discussion

Working from data collected by SCAA for more than 2,500 children in 1996, factors have been identified which predict the percentage of pupils likely to correctly spell the given words at ages 7 and 11. The factors identified are those which are arrived at by any common sense view of the level of difficulty that words present to pupils: number of letters, frequency of usage, and the presence of unusual forms of phonemic representation.

There is often criticism of poor spelling in schools and even at University level. This study has clearly indicated that a major factor in poor spelling, which will also be reflected in poor reading, is the failure of English spelling to conform to specific rules for the representation of phonemes. For 7-year-olds, words with unusual written forms are much more difficult, and the more unusual the written form, the more difficult they are to spell. For 11-year-olds, the words tested were longer, less frequently used, and more likely to have several unusual forms of representation. In this case, because 11-year-olds have acquired much greater experience with words, unusual representations may be mitigated by more frequent use. Even bizarre representations are learned by 11-year-olds if they are frequently encountered.

The analyses of data presented here clearly indicate that a major cause of poor spelling is to be found in the form of representation of the words, and not solely in the students. The main problem is that for many words the form has to be known and remembered, because the imperfect patterns which govern English cannot always be applied to give the correct result. Instead of using coherent patterns that always give

correct answers, written English has developed as a system which requires a great deal of rote learning. This takes time and energy that could be better employed in other educational activities.

By the age of 11 years, most students are able to deal successfully with all but the most unusual written forms for word sounds. By regularizing such highly irregular forms (eg, *friends*, *stretched*), spelling, reading and the self-confidence of these students would be greatly enhanced.

If we do not develop a rational system of English spelling, we must accept the consequences: to eradicate poor spelling and reading at a national level, much more time must be devoted to learning the idiosyncratic written forms. The extra time needed will be at the expense of other subjects such as maths, science and technology.

Table 1: Word values for 7-year-old test

Word	Phonetic rendering	No. of letters	Phonic ratio	Absolute frequency	% Score	Logprob score	Average PhR	Average PhT	Tricky phonemes
because	bɪkɒz	7	0.29	777	35	-0.27	42.84	1.55	0.32
boat	boːt	4	0.5	72	55	0.09	66.94	3.03	5.48
bucket	bʌkɪt	6	0.67	6	23	-0.52	58.6	2.51	5.2
family	fæməli	6	0.83	281	29	-0.39	72.8	2.15	3.4
fish	fɪʃ	4	0.5	121	84	0.72	60.06	2.97	25.08
flag	flæg	4	1	8	83	0.69	91.87	2.45	84.44
friends	frendz	7	0.71	177	25	-0.48	73.05	3.14	0.6
hand	hænd	4	1	460	85	0.75	97.17	3.52	91.68
hat	hæt	3	1	56	97	1.51	98.63	3.29	95.9
holiday	hɒlədeɪ	7	0.57	74	40	-0.18	64.4	1.94	3.4
house	haʊs	5	0.4	571	62	0.21	59.32	0.42	4.41
morning	mɔː(r)nɪŋ	7	0.71	233	41	-0.16	72.87	3.65	48.53
net	net	3	1	53	91	1	94.43	5.97	90.4
pictures	pɪktʃə(r)z	8	0.38	83	13	-0.83	57.14	2.47	2.8
road	roːd	4	0.5	205	54	0.07	64.15	2.9	5.48
shout	ʃaʊt	5	0.2	10	39	-0.19	64.85	2.67	25.08
smile	smaɪl	5	0.4	76	32	-0.33	70.24	3.23	32.89
sock	sɒk	4	0.75	1	61	0.19	53.92	1.96	5.2
spade	speɪd	5	0.6	2	39	-0.19	74.1	3.07	48.22
wait	weɪt	4	0.5	82	27	-0.43	58.18	2.75	14.97

Table 2: Word values for 11-year-old test

Word	Phonetic rendering	No. of letters	Absolute frequency	% Score	Logprob score	Average PhR	Average PhT
beautiful	bjuːtɪfəl	9	85	48	-0.03	60.83	2.94
crept	krept	5	5	71	0.39	88.92	4.34
disturbed	dɪstɜː(r)bd	9	26	63	0.23	63.93	3.48
echoed	ekoːd	6	12	55	0.09	38.21	1.76
heard	hɜːd	5	239	74	0.45	67.05	1.55
honest	onɪst	6	33	89	0.91	56.21	5.18
notice	noːtɪs	6	103	83	0.69	66.06	4.63
piece	piːs	6	63	64	0.25	36.14	1.38
remained	rɪmeɪnd	8	103	64	0.25	55.67	3.05
replace	rɪpleɪs	7	20	84	0.72	47.65	2.9
shook	ʃʊk	5	53	54	0.07	20.96	0.37
silence	saɪləns	7	92	68	0.33	50.68	3.25
slipped	slɪpt	7	32	65	0.27	46.2	3.3
sneeze	sniːz	6	1	67	0.31	43.63	2.99
sprawling	sprɔːlɪŋ	9	4	39	-0.19	67.31	3.51
still	stɪl	5	823	97	1.51	60.05	4.72
stretched	stretʃt	9	23	34	-0.29	58.91	3.22
tallest	tɔːlest	7	1	85	0.75	60.37	3.75
uncoiled	ʌnkɔɪld	8	1	57	0.12	65.19	2.91
visitors	vɪzɪtəz	8	37	71	0.39	63.8	3.44

Table 3: Representation for /e/ phoneme

/e/ phoneme = 3.36% of total phonemes in sample

Word	Phonetic Rendering	Spelling	% of /e/ phoneme (PhR)	% of total phonemes (PhT)
dental	dentl	e	90.60	3.04
heather	heðə(r)	ea	7.00	0.24
anybody	enιbɒdi	a	1.20	0.04
friendship	frendʃιp	ie	0.60	0.02
against	əgenst	ai	0.60	0.02
			100.00	3.36

Table 4: Correlation values for 7- & 11-year-old tests

Correlation, 2-tailed significance, 7-year-olds:

	LETTERS	PHONIC	TRICKY
LOGPROB	−0.76*	0.16*	0.77*
LETTERS		−0.402	−0.56*
PHONIC			0.70*

* p<0.01

Correlation, 2-tailed significance, 11-year-olds:

	LETTERS	FREQABS	PHONIC
LOGPROB	−0.55*	0.66*	0.51*
LETTERS		−0.37	0.13
FREQABS			0.22

* p<0.01

References

Bloom, B.S. (1981) All our children learning. Chapter 8:learning for mastery: pp. 153-177. McGraw-Hill: London.

Carroll, J. (1963) 'A model of school learning', Teachers College Record, 64, 723-733.

Frith, U. (1985) Beneath the surface of developmental dyslexia. In K.E. Patterson, J.C. Marshall & M. Coleath (Eds.), Surface Dyslexia (pp. 301-330). London: Erlbaum.

Hofland, K. and Johansson, S, (1982) Word Frequencies in British and American English. Bergen: The Norwegian Computing Centre for the Humanities.

Lazo, M.G., Pumfrey, P.D. & Peers, I. (1997) 'Metalinguistic awareness, reading and spelling: roots and branches of literacy', Journal of Research in Reading, 20(2), 85-104.

School Curriculum & Assessment Authority (1997) Standards at Key Stage 1 English & Mathematics. Report on the 1996 National Curriculum Assessments for 7-year-olds. London: QCA.

School Curriculum & Assessment Authority (1997) Standards at Key Stage 2 English & Mathematics. Report on the 1996 National Curriculum Assessments for 11-year-olds. London: QCA.

Rego, L.B. (1991) The role of early linguistic awareness in children's reading and spelling. University of Oxford, unpublished DPhil Thesis..

Spencer, K.A. (1996) 'Recovering Reading Using Computer Mastery Programmes', British Journal of Educational Technology, 27(3), 191-203.

◆

Updated edition of SSS pamphlet

Until very recently, every member of the SSS received a copy of Bob Brown's *Spelling Reform in Context: a Typology, List and Bibliography of English Spelling Reforms* (2nd edition, 1992). This pamphlet by the Society's Secretary from 1991 to 1996 has always been highly regarded as an introduction to the history of and the various possible approaches to spelling reform in English. Stocks of the pamphlet being exhausted by the summer of 1998, the SSS Committee decided on a new edition, updated to take account of some of the latest developments. This 3rd edition is now available (ISBN 0 9506391 7 6) at £2 or US$4, or free to new members who never received an earlier edition. Orders to the Editor (see p1 for address).

13. Spencer, K.A. (1999a) Predicting spelling behaviour in 7 to 11 year-olds, Reading, July, 72-77.

Ken Spencer

Predictive Models of Spelling Behaviour in 7- & 11-year-olds

Ken Spencer is Lecturer in Media Studies and Educational Technology in the School of Education, Hull University, Hull HU6 7RX, UK. He may contacted by email at <k.a.spencer@educ.hull.ac.uk>. The statistical Tables are given at the end of the article.

0 Abstract

A predictive model for spelling is suggested, based on the results of spelling tests taken by 2,684 seven- and eleven-year olds, in 1996. The tests were part of a national scheme of testing for the School Curriculum & Assessment Authority (SCAA). The factors identified as influencing the number of children correctly spelling a word are: word length (number of letters), frequency of a word, and a measure of the word's phoneticity. A measure of the most infrequent form of representation of the phonemes in a word (the 'trickiest' phoneme) is a strong factor with 7-year-olds, whereas a measure of the average phoneticity of a word is a better indicator of word difficulty for 11-year-olds, who are susceptible to the mitigating effects of high word frequency on irregular spellings.

1 Learning for mastery

Carroll's model of school learning (1963) suggests that most pupils are capable of reaching levels of performance more usually associated with the top 10-20% of the school population. He proposed that a number of variables could be manipulated to increase levels of performance, such as the quality of teaching materials and the amount of time available for learning. Some variables associated with the pupil, such as perseverance, may be difficult to manipulate.

Carroll's model has been applied to teaching by Bloom and his associates, in the methodology known as Learning for Mastery, which places great emphasis on formative testing in order to determine deficiencies in either learning or teaching. An essential within this system is the requirement that high levels of performance are demanded at the early stages of learning, which ultimately result in higher overall levels of performance. Required criterion levels are as high as 100%, although more usually they range between 80-90%.

2 Case-study in remedial literacy

This approach has recently been applied to the teaching of reading (Spencer, 1996), using computer-based learning techniques. In the case of one pupil, who had reached the age of 10.5 years without being able to read the most common word in the English language, decisions had to be made concerning the teaching of the most common 100 words: should a phonics approach be adopted, or a method based on gradually increasing the demands in a simple spelling exercise. Many of the most common words fail to obey even the most rudimentary rules and so the simple spelling approach, with increasing mastery demands, was

adopted. Practising for 10-15 minutes per day the pupil mastered 80% of the words over a period of 12 weeks.

It was clear from the performance of this pupil that much of his problem was associated with the vagaries of English spelling — he simply gave up when he applied rules to common English words and was told that he was wrong: the rules didn't work, and what should have been a simple task proved impossible. With the continuing concern of parents, teachers and politicians about the levels of literacy in the UK, the question arises: are we disadvantaging our children by making a simple task incomprehensible?

3 Searching for models

Of particular interest to researchers investigating the application of computer-based literacy systems is the search for a predictive model of spelling performance. Knowledge of such a model would indicate the factors that make words difficult to spell; determine if they are the same for all ages; and may indicate how strategies change with age, to make spelling more accurate. This, in turn, has implications for reading. Frith's (1985) six-step model of literacy development suggests that there is an initial period when children use a logographic strategy to read and a phonological strategy to spell, ie, they read and spell in different ways. According to this model, the emergence of phonemic representations in spelling leads to advances in later reading. Rego (1991) demonstrated that the ability to spell non-words is strongly related to progress in reading, and this has been confirmed by Lazo et al. (1997), who show that early attempts to read words are strongly related to the progress made in spelling, as early attempts to spell words influence later reading.

The following analysis, which identifies several models for learning to spell, is based on data from national tests carried out by SCAA in 1996. SCAA's activities have recently been incorporated into the Qualifications and Curriculum Authority (QCA) whose statutory functions were set out in the Education Act, 1997. Principally, by its forthcoming review of the school curriculum, QCA hopes to define the structure and content of teaching and learning that will enable all pupils to develop and demonstrate their knowledge, skills and understanding. QCA's functions and responsibilities include: developing learning goals for the under-fives; accrediting assessment schemes for children entering primary school; monitoring and reviewing the National Curriculum and its assessment; continuing with the development of national assessment at the ages of 7, 11 and 14.

Teachers can often figure out the percentage frequencies for the various representations, but are sometimes dramatically wrong. We are introduced to the "how now brown cow" rhyme early in life, and it does colour our view of the frequency of occurrence of the sound. In fact, the ratio is:

ow	cow	26%
ou	loud	74%

But do these different frequencies or probabilities matter? McGuinness says, most emphatically: yes. By taking account of them she argues she has had dramatic results with a range of clients. I agree, and the results from experiments with children in a Hull school indicate that these probabilities have a profound effect on the children's abilities in spelling.

Analysis of 1996 National Spelling Test Results

In order to determine the contribution that code probabilities have on spelling, which Lazo et al (1997) have shown to be linked to reading, I analysed the results for the 1996 national spelling test (Spencer 1998). A predictive model of spelling behaviour was developed which suggests that readily identifiable characteristics of words, including the probabilities associated with different representations of sounds, can be used to determine word difficulty in spelling tasks. The spelling test for 7 year-olds (Key Stage 1) was taken by 1,184 children working at level 2, in SCAA's Schools Sampling Project, a national representation sample of schools taking part in a longitudinal monitoring survey; and the test for 11 year-olds (Key Stage 2) was taken by 1,500 pupils from the University of Bath's sample for the 1996 SATs. The model identified three factors which contributed to word difficulty: number of letters in the word, phoneticity of the word and frequency of occurrence of the word. The factors accounted for more than 60% of the variability in pupil scores.

The simple word length measure consists of the number of letters in the word, and was included because in the initial stages of spelling (and reading) 7 year-olds are still developing short term memory strategies, and any lapses in memory are likely to manifest themselves with longer words. Longer words also give more opportunities for errors. A significant negative correlation (-0.55) was obtained between word length and the scores for each word: the longer the word, the fewer the pupils who can spell it.

Phoneticity is a major factor when considering how easy a word is to spell or read, but one that can be defined in a variety of ways. There is no standardised way of measuring this, but the approach used for the present study is similar to those used by Carney (1994) and McGuinness (1998). Children learn at an early age that a variety of representations can be used for the same sound and, as SCAA recognise, the difficulty is less knowing the patterns than knowing which pattern to use in each individual word. In order to establish the range of representations of the phonemes that make up the English language, and the frequency of each representation, the 3,500 most common words from the Lancaster/Oslo-Bergen Corpus (Hofland and Johansson 1982) were analysed. The phonic representation of each word was determined from the Oxford English Dictionary (Second Edition, CD-ROM version) enabling the alphabetic representations of each phoneme to be determined for each word. With all phonemes coded (20,000), tables showing the various forms of representation for each phoneme, and their frequency of occurrence (FO) were extracted. The average number of representations per phoneme was 5.95 (a total of 262 for the 44 English phonemes used in the O.E.D.). The phoneticity factor used here was determined by adding up the FO value for the representation of each phoneme in a word and calculating the average value for that word. A significant correlation of 0.50 was obtained between the average phoneticity of the word and the percentage correct scores issued by SCAA.

The word frequency was also found to influence the spelling of words; and it seems reasonable to suggest that the more common a word is, the more likely it is that unusual representations of sounds will be internalised by the learner. A significant correlation of 0.66 was obtained, showing that word frequency ameliorates the effects of uncommon forms of representation, so that even very unusual forms are learned if the word is frequently in use.

The Whole-school Test

The same words used by SCAA in the 7 and 11 year-old tests were combined into a single list of 40 items and administered to all pupils in school at a local primary school, which performs at average national SAT levels in English, maths and science. The collected data of spelling performance spanned five year groupings (Y2–Y6), for a total of 236 pupils. The mean percentage correct value for all words for each year group increases steadily from 30.89% to 67.13%. When controlled for word length and frequency of a word, there is a high average correlation (0.62) over the five year groups between the difficulty of a word, as shown in the number of students in a year group spelling it correctly, and the phoneticity of the word, which ranges from an average phoneticity value of 20.96 (shook) to 98.63 (hat). The correlations are similar across the five year groups, showing that the same factors are at work for all pupils in the primary school.

Table 1. Percentage correct values, Age and Reading Quotients, Year Groups 2–6

	Year 2	Year 3	Year 4	Year 5	Year 6
Mean % score	30.89	39.16	51.25	56.25	67.13
(S.D.)	25.30	29.96	30.68	27.99	23.20
Mean Age	7.08	8.07	9.15	10.14	11.08
(S.D.)	0.28	0.34	0.31	0.21	0.29
Mean Reading Quotient	99.79	102.24	104.05	99.44	96.44
(S.D.)	13.26	19.58	13.17	10.93	13.37

The factors which influence pupils' spelling abilities, in this study and the SCAA work, are those which are arrived at by any common-sense view of the level of difficulty that words present to pupils: number of letters, the presence of unusual forms of phonemic representation (phoneticity), and frequency of occurrence of the word in the language. The results indicated that between KS1 and KS2, as children have more contact with words and expand their vocabulary, the frequency of occurrence exerts an influence: increased frequency ameliorates the effects of obscure phoneme representations (i.e. words with low average phoneticity). These factors work in a very similar way across ages 7 to 11, in a typical primary school.

The probability of a particular representation works in another way. The results described above use the traditional marking method for spelling. A word is either absolutely correct or it is incorrect, irrespective of the number of errors: it is a form of mastery test, with a high criterion level of 100%! But you can also mark spellings on a phoneme by phoneme basis, and when you do so the probability factor associated with a particular phoneme representation is the largest determinant of correct spelling. I used the 1997 SCAA spelling test words and again presented the words to years 2 to 6 in the Hull school that had provided data for the spellings of the 1996 SCAA words. The results showed the same factors influencing spelling of individual phonemes throughout the five years: phoneme representation probability was the strongest factor (in the THRASS example above, this would be 50 for "o" and 5 for "oa"); phoneme position in the word and the number of letters in the word were significant but weaker factors.

Discussion

Knowing that the frequency of a particular form of representation of a word sound has an effect on spelling and reading has profound implications for the teaching of reading. McGuinness (1998) indicates that initially only those letters that have a one-to-one relationship with sounds should be taught. She calls

this the basic code. Gradually more complex representations are introduced, with children mastering the code at each stage before going on to the next. Successful literacy strategies must teach students that the alphabetic code is reversible (speech-text-speech). They must also recognise that the complexity of the code, which not only has less letters (26) than sounds (44), but compounds this by having many different letter combinations for some sounds, demands carefully graded steps with the most probable representations being introduced in the first instance. McGuinness (1998) indicates that there may not be such a condition as dyslexia, and that a system of instruction, which employs phoneme awareness training, together with a phased introduction to the complexity of the code for English, can have quite dramatic results for failing students. She reports that 87 children, aged 6 to 16 years, with reading difficulties, gained an average of 13.7 standard score points on word recognition and 19.34 points on nonsense word decoding after only 12 hours of training. This is clearly an extreme view, and one based on a relatively small sample of clients.

A corollary from this position is that with full orthographic transparency there should be no dyslexia, and this can be tested by looking at orthographies other than English. There is, indeed, a growing body of evidence which shows that languages such as Turkish, Italian, and German, with their more regular orthographies, are much less difficult for beginning readers than English. But even with such languages there will always remain a group of children that still have reading difficulties. Landerl et al (1997) conducted a stringent comparison of these difficulties in typical German and English dyslexics, and found that "English dyslexic children suffered from much more severe impairments in reading than German dyslexic children ... the differences between the English and the German dyslexic children were comparatively small for the short, high frequency words. For all other conditions, however, the differences were enormous." (p. 328) Similar results were also reported for normal children in a later study (Frith 1998). These results lead to a less extreme view than that suggested by McGuinness: even in more perfect writing systems

there will be some individuals who, for a variety of reasons, have reading/writing difficulties; but the burden is much heavier for those individuals who have to deal with the unnecessary complexities of codes in such languages as English.

A perfect spelling system is one that has no alternative spellings for the same sound, and no overlap in the code where one spelling pattern stands for different sounds. The main problem for readers is that English does not have such a system. It follows from this that spelling errors, as well as reading errors, will be related to some measure of the phoneticity of words, where phoneticity is a measure of the deviation of a word from a truly phonetic rendition.

To help teachers consider the sequencing of sounds and their representations I have produced three tables which give the 44 English phonemes and their different representations. Those in Table 2 have a strong candidate for a single representation that can be taught to children. This representation occurs in more than 51% of cases. There is no intermediate (21–50%) category for these sounds, although there is a small number of unusual forms, which can be ignored in the early stages of learning. Table 3 shows those phonemes that have a strong representation coupled with an intermediate representation, which can be confusing for children and should clearly be taught by introducing the more common element first. There are also, in some cases, words with unusual forms. Phonemes in Table 4 do not have a strong form of representation. These are the really tricky phonemes,

often with several forms in the frequency range 20–50%, plus many different forms in the range 1–20%. Teaching the representations for these phonemes will be a challenge for any teacher because there is often no way of predicting how the sound will be represented with any degree of certainty.

When teaching phonics with an indication of the different frequencies of representation we can actually begin to make predictions concerning which words our pupils are likely to have difficulty with. Table 5 is an attempt to do this: the most common 150 words in the English language have been classified according to their most difficult phoneme. The first column indicates those words that do not have a low frequency representation and should present little difficulty for children in primary school. There are some interesting points to note: "th" is a very stable form of representation; "ou" is the most common form of representation for the au sound (as in "loud"), not "ow". The final column shows those words that have unusual forms of representation (<5%) and these will often present problems for children, although because they are high frequency words their forms will be learned eventually. Perhaps the best strategy in dealing with words in the last two columns is learning by rote, simply because there are no rules which will enable children to derive spellings from the sounds. Indeed, perhaps we will have to accept that English, like Japanese, may have to be divided into 2 groups of words: those that can be decoded from a simple alphabetic principle; and those that have representations of phonemes that are so infrequent that they can only be learned by rote.

Table 2. *Phonemes with single high probability representation*

phoneme	51–100%	example	phonic	21–50%	5–20%	<5%
h	h	hut	hʌt			
æ	a	personality	pɜːsənælɪtɪ			
ð	th	the	ðə			
θ	th	three	θriː			bb
b	b	be	biː			bb
p	p	part	pɑːt		pp	
d	d	and	ænd		ed	dd
g	g	ground	graʊnd			x, gg
m	m	from	frɒm			mm, mb
r	r	from	frɒm			rr, wr
t	t	it	ɪt			tt, ed
f	f	forbidden	fəbɪdn		ff, ph	gh
n	n	grant	grɑːnt			gn, nn, kn
e	e	dental	dentl		ea	a, ie, ai
ɒ	o	prominent	prɒmɪnənt		a	ow, o+e, au
k	c	calm	kɑːm		k, x, ck	qu, ch
l	l	like	laɪk		ll, al, le	el, il
j (juː)	u	music	mjuːzɪk		y	ue, ew, eau, ui, i
s	s	successful	səksesfəl		ce, ss, c	se, x, sc

Table 3. Phonemes with high probability representation and medium probability representation

phoneme	51–100%	example	phonic	21–50%	5–20%	<5%
aʊ	ou	loud	laʊd	ow		
ɔɪ	oi	exploit	ɪksplɔɪt	oy		
ŋ	ng	proceedings	prəsi:dɪŋz	n		
v	v	verdict	vɜ:dɪkt	ve		f
tʃ	ch	cheese	tʃi:z	t	tch	
ʌ	u	summit	sʌmɪt	o	ou	oo
ʊ	u	fulfilled	fʊlfɪld	oo		oul, o
w	w	upward	ʌpwəd	u	wh	o
ʒ	s(ion)	explosion	ɪkspləʊʒn	s		t(ion), ge
z	s	is	ɪz		se	ze, z, x, ss
ɑ:	ar	partner	pɑtnə	a		au, al, er, ear, are
ʃ	ti(o)	execution	eksɪkju:ʃn	sh	ci, si	s, sci, x, ss, ch
ɪ	i	hill	hɪl	e		a, u, ey, ui, o, ia, ee, ai

Table 4. Phonemes with medium probability representations and low probability representations

phoneme	51–100%	example	phonic	21–50%	5–20%	<5%
i:		pleaded	pli:dɪd	ea, ee, e, y at end of word	ie	i, ei, ey, eo, ae
dʒ		emerge	ɪmɜ:dʒ	ge, j, g	dge	d, gg, dg
əʊ		romance	rəʊmæns	o, o+e	ow	oa, ou, oe
eə		unaware	ʌnəweə	are, ai(r)	ere, a(r)	ear, eir, ayor, ayer
eɪ		decade	dekeɪd	a+e, a	ai, ay	ei, ea, ey, et, e
aɪ		suicide	sju:ɪsaɪd	i+e, i	y, igh	ie, ei, uy, ui, ir, eye, eigh
ʊə		visual	vɪʒʊəl	ua	ure, u, ue, our, ou	oor
ɜ:		her	hɜ:	er	ur, ir, ear, or	our, olo, ere
u:		fluid	flu:ɪd	u	u+e, oo, o, ue, ew	ui, ou, ough, eau, wo, oe
ɔ:		corns	kɔ:nz	or	ore, au, a(ll), aw, a, our	ar, ough, oor, oa, augh, oar, al
ə		dramatic	drəmætɪk	a	er, o, e	u, i, or, ure, ar, our, iou, y, ur, ier, ei
ɪə		disappear	dɪsəpɪə		ear, iou, ia, ea, e, io, ere, ier, ie	iu, eer, ior, iar, eu, eo, eou

References

CARNEY, E. (1994) *A Survey of English Spelling.* London: Routledge.

DAVIES, A. and RITCHIE, D. (1996) *THRASS: Teacher's Manual.* London: Harper Collins.

FRITH, U., WIMMER, H. and LANDERL, K. (1998) 'Differences in Phonological Recoding in German- and English-Speaking Children', *Scientific Studies of Reading*, 2(1), 31–54.

HANNA, P.R., HANNA, J.S. and HODGES, R.E. (1966) *Phoneme-grapheme Correspondences as Cues to Spelling Improvement.* OE-32008, US Department of Health, Education, and Welfare.

HOFLAND, K. and JOHANSSON, S. (1982) *Word Frequencies in British and American English.* Bergen: The Norwegian Computing Centre for the Humanities.

LANDERL, K., WIMMER, F. and FRITH, U. (1997) 'The impact of orthographic consistency on dyslexia: A German–English comparison', *Cognition*, 63, 315–334.

LAZO, M.G., PUMFREY, P.D. and PEERS, I. (1997) 'Metalinguistic awareness, reading and spelling: roots and branches of literacy', *Journal of Research in Reading*, 20(2), 85–104.

McGUINNESS, C., McGUINNESS, D. and McGUINNESS, G. (1996) 'Phono-Graphix: A new method for remediation reading difficulties', *Annals of Dyslexia*, 46, 73–96.

McGUINNESS, D. (1998) *Why Children Can't Read: And What We Can Do About It.* London: Penguin.

MITTON, R. (1996) *English Spelling and the Computer.* London: Longman.

SCHOOL CURRICULUM & ASSESSMENT AUTHORITY (1997) *Standards at Key Stage 1 English & Mathematics. Report on the 1996 National Curriculum Assessments for 7-year-olds.* London: QCA.

SCHOOL CURRICULUM & ASSESSMENT AUTHORITY (1997) *Standards at Key Stage 2 English, Mathematics & Science. Report on the 1996 National Curriculum Assessments for 11-year-olds.* London: QCA.

Table 5. Most frequent 150 English words, sorted by difficult phonemes (% frequency of representation); rank order of word frequency 1–150 (RO) is also given

100–60%	RO	59–20%	RO	19–10%	RO	9–5%	RO	<5%	RO
an	32	such	83	there	34	year	127	course	150
than	62	much	86	over	77	he	11	would	41
at	17	so	44	other	67	be	14	could	63
that	6	no	45	after	84	we	38	should	76
man	85	go	122	never	125	she	28	little	101
on	15	old	130	under	137	these	69	new	81
not	21	only	56	another	131	being	94	you	30
and	3	most	88	the	1	might	114	thought	144
had	19	both	129	into	60	right	146	great	123
get	120	long	119	own	118	may	72	their	39
men	143	or	29	first	75	way	97	any	73
them	58	for	10	like	80	say	140	many	90
then	66	made	74	make	113	day	145	they	31
from	23	same	117	take	139	to	4	through	116
out	51	came	138	when	42	do	71	one	36
found	141	her	27	which	26	work	110	who	48
in	5	last	105	where	89	world	147	even	92
it	9	time	61	while	149	used	136	though	142
if	43	life	126	too	99	more	49	know	112
him	47	have	24	each	135	before	87	are	25
did	103	now	70	by	18	all	37	were	33
can	59	how	104	my	57	here	121	of	2
good	100	down	102	will	46	also	93	our	79
is	7	about	52	well	91	was	8	said	50
as	12	just	108	still	109	what	53	again	133
his	16	those	95	see	111	your	107	two	65
has	40			been	35	back	98	some	55
up	54			three	124			come	128
us	134			between	106			because	115
but	22							people	96
me	64								
very	78								
with	13								
went	148								
without	132								
its	68								
this	20								
must	82								

REGO, L.B. (1991) The Role of Early Linguistic Awareness in Children's Reading and Spelling. University of Oxford, unpublished DPhil Thesis.

SPENCER, K.A. (1996) 'Recovering reading using computer mastery programmes', *British Journal of Educational Technology*, 27(3), 191–203.

SPENCER, K.A. (1998) 'A predictive model of spelling behaviour in 7 to 11 year-olds', *Journal of the Simplified Spelling Society*, 24(2), 23–27.

CONTACT THE AUTHOR:
Ken Spencer, The Centre for Educational Studies, The Institute for Learning, The University of Hull, Hull HU6 7RX.

14. Spencer, K.A. (1999b) Predicting word-spelling difficulty in 7- to 11-year-olds, Journal of Research in Reading, 22(3) 283-292.

Predicting word-spelling difficulty in 7- to 11-year-olds

Ken Spencer

Centre for Educational Studies, University of Hull, UK

ABSTRACT

Some languages have simple grapheme-phoneme codes in which there is a one-to-one mapping, making them easy to teach and learn, while others have more complicated structures and are more difficult for teachers and students. There is now an increasing number of studies which demonstrate that readers in more transparent orthographies such as Italian, Spanish, Turkish, Greek and German have little difficulty in decoding written words, while English children have many more problems. Increasingly, lack of orthographic transparency in English is seen as having a powerful negative effect on the development of literacy skills in English-speaking children.

In the present study, the word factors associated with poor spelling in 5 school year-groups (ages 7 to 11 years) are identified as: (a) frequency of the word in the English language, (b) length of the word and (c) a measure of the phoneticity of the word. The concept of word phoneticity is explored and data is presented illustrating the predictive model of spelling. The implications the model has for teaching and learning English are discussed, with particular reference to the beneficial effects that greater orthographic transparency would have for children.

INTRODUCTION

A recent approach to the teaching of reading (Spencer, 1996), using computer-based learning, enabled one student, who had reached the age of 10½ years without being able to read the most common word in the English language, to master 80% of the most common 100 words over a period of 12 weeks, practising for only 10–15 minutes per day. It was clear from the performance of students in the research programme that their problems with spelling and reading were often associated with the vagaries of English spelling.

Knowledge of the factors that make words difficult to spell has implications for reading. Frith's (1985) six-step model of literacy development suggests that there is an initial period when children use a logographic strategy to read and a phonological strategy to spell, i.e. they read and spell in different ways. The emergence of phonemic representations in spelling leads to advances in later reading. Rego (1991)

demonstrated that the ability to invent spelling is strongly related to progress in reading, and this has been confirmed by Lazo et al (1997), who show that early attempts to read words are strongly related to the progress made in spelling, as early attempts to spell words influence later reading.

McGuinness (1998) has suggested that successful literacy strategies must teach students that the alphabetic code is reversible (speech–text–speech). They must also recognise the complexity of the English orthographic code, which not only has fewer letters (26) than sounds (about 44 phonemes), but compounds this by having many different letter combinations for some sounds. For successful learning this demands carefully graded steps, with the most probable representations being introduced in the first instance. McGuinness indicates that a system of instruction that employs phoneme awareness training, together with a phased introduction to the complexity of the code for English, can have quite dramatic results for failing students. She reports that 87 children (aged 6 to 16) years, with reading difficulties, made gains averaging 13.7 standardised score points on word recognition and 19.34 points on nonsense word decoding after only 12 hours of training. McGuinness defines a perfect spelling system as one that has no alternative spellings for the same sound, and no overlap in the code where one spelling pattern stands for different sounds. She suggests that the main problem for readers is that English does not have such a system. It follows from this that spelling errors, as well as reading errors, will be related to some measure of the phoneticity of words, where phoneticity is a measure of the deviation of a word from a truly phonetic rendition.

Using data from spelling tests administered to 1,184 7-year-olds and 1,500 11-year-olds as part of a national survey carried out by SCAA (School Curriculum and Assessment Authority) in 1996 (SCAA, 1997a, 1997b) a predictive model of spelling behaviour was developed by the author (Spencer, 1998). This model suggests that readily identifiable characteristics of words can be used to determine word difficulty in spelling tasks. Three factors that contributed to word-spelling difficulty were identified: number of letters in the word, the phoneticity of the word and the frequency of occurrence of the word.

The predictive model of student performance from national data sets indicates factors that are related to the specific population tested. If it is to be a useful educational tool, the model must be shown to be sufficiently robust to account for the behaviour of students in other contexts. The present research has been undertaken to determine the extent to which the factors identified in the national studies are relevant in predicting spelling behaviour in a typical school situation, across ages 7 to 11 years (school years 2 to 6).

METHOD

Participants

In the present study, spelling data were collected from all children attending a primary school in an urban part of Hull during the last week of February, 1998. The school has a total of 246 children and performs at average UK national levels in English, maths and science. The collected data of spelling performance spanned five year groupings (ages 7 to 11 years) for a total of 236 students (10 of the 246 were absent on one or both of the two days of testing). Data on the average ages and

average reading quotients of each year group are shown in Table 4. Reading quotients (obtained from the school records) were based on the administration of several standardised reading tests that are widely used in UK schools .

Materials and procedure

The SCAA 1996 national spelling data had been collected using 2 sets of 20 words (see Table 1), one set that was administered to 7-year-olds, the other to 11-year-olds. For the present study, the 2 sets were combined and administered to each of five consecutive year-groups (ages 7 to 11 years). Words from both sets of 20 words were randomly assigned to either list 1 or list 2, each list being administered to all the children present during the morning session on two consecutive days. Class teachers administered the spelling tests, saying the word followed by an example sentence, and a further repetition of the word. There was no time limit for the test. Children wrote their answers on forms with word numbers clearly marked.

Measures employed in the study

Word length (WL)

This is a simple measure, consisting of the number of letters in the word, and was included because in the initial stages of spelling (and reading) 7-year-olds are still developing short term memory strategies, and any lapses in memory are likely to manifest themselves with longer words. Longer words also give more opportunities for errors.

Phoneticity (PhR)

Phoneticity can be defined in a variety of ways. There is no standardised way of measuring this factor and a number of approaches were adopted and refined for the predictive model (see Spencer, 1998). Children learn at an early age that a variety of representations can be used for the same sound and the difficulty is not so much a matter of knowing the patterns than knowing *which pattern* to use in each individual word. In order to establish the range of representations of the phonemes that make up the English language, and the frequency of each representation, the 7,000 most common words from the Lancaster-Oslo-Bergen (LOB) Corpus (Hofland and Johansson, 1982) were analysed. The LOB Corpus of 1 million words provides statistical data for each word, including the total number of occurences of each word (i.e. the frequency per million words). The 7,000 words were lemmatized by removing inflected forms (such as: cleaned, cleaning, cleans) if the base form (clean) also appeared in the list; other inflected forms were retained. This process resulted in a reduced sample of 3,500 words.

The phonemic representation of each of the 3,500 words was determined from the Oxford English Dictionary (Second Edition, CD-ROM version). The alphabetic representations of each phoneme were determined by selecting all words, within the 3,500 word sample, containing a particular phoneme. Each word was then scanned (initially by a computer programme designed to indicate likely forms of orthographic representation) to determine the particular written form of the phoneme. With each

Table 1. Word values for 7-year-old (words 1–20) and 11-year-old (words 21–40) from the SCAA 1996 Spelling Tests.

Word	Phonic representation	Word length	Absolute frequency	Average PhR*
because	bɪkɒz	7	777	42.84
boat	bəut	4	72	66.94
bucket	bʌkɪt	6	6	58.6
family	fæməlɪ	6	281	72.8
fish	fɪʃ	4	121	60.06
flag	flæg	4	8	91.87
friends	frendz	7	177	73.05
hand	hænd	4	460	97.17
hat	hæt	3	56	98.63
holiday	hɒlədeɪ	7	74	64.4
house	haus	5	571	59.32
morning	mɔːnɪŋ	7	233	72.87
net	net	3	53	94.43
pictures	pɪktʃəz	8	83	57.14
road	rəud	4	205	64.15
shout	ʃaut	5	10	64.85
smile	smaɪl	5	76	70.24
sock	sɒk	4	1	53.92
spade	speɪd	5	2	74.1
wait	weɪt	4	82	58.18
beautiful	bjuːtɪfəl	9	85	60.83
crept	krept	5	5	88.92
disturbed	dɪstɜːbd	9	26	63.93
echoed	ekəud	6	12	38.21
heard	hɜːd	5	239	67.05
honest	ɒnɪst	6	33	56.21
notice	nəutɪs	6	103	66.06
piece	piːs	6	63	36.14
remained	rɪmeɪnd	8	103	55.67
replace	rɪpleɪs	7	20	47.65
shook	ʃuk	5	53	20.96
silence	saɪləns	7	92	50.68
slipped	slɪpt	7	32	46.2
sneeze	sniːz	6	1	43.63
sprawling	sprɔːlɪŋ	9	4	67.31
still	stɪl	5	823	60.05
stretched	stretʃt	9	23	58.91
tallest	tɔːlest	7	1	60.37
uncoiled	ʌnkɔɪld	8	1	65.19
visitors	vɪzɪtəz	8	37	63.8

* Phoneticity of the word (for explanation see text).

word containing a given phoneme coded for the particular form of representation, the various written forms for a particular phoneme were then extracted. The frequency for each form of representation was determined as a proportion of all the representations of that phoneme (PhR). This measure considers a particular representation of a phoneme only in relation to other representations of that phoneme. As an illustration, Table 2 shows the values for the /e/ phoneme. This phoneme represents 3.36% of all the phonemes in the sample. The most common alphabetic representation is <e>, and this is found in 90% of cases (the PhR value of <e> representation for the /e/ phoneme). The unusual form <ai> has a PhR value of 0.6%. The average number of representations per phoneme was 5.95 (a total of 262 for the 44 English phonemes used in the Oxford English Dictionary) for the 3,500 word sample.

Table 2. Forms of representation for 'e' phoneme.

/e/ phoneme = 3.36% of total phonemes in sample

Word	Phonic Representation	Form of Representation	% of "e" phoneme (PhR)
dental	dentl	e	90.60
heather	heðə	ea	7.00
anybody	enɪbɒdɪ	a	1.20
friendship	frendʃɪp	ie	0.60
against	əgenst	ai	0.60
			100.00

Carney (1994) completed a similar analysis of the English language for 26,000 words, which were lemmatized, and also provides statistical information on the various representations of individual phonemes. This process has also been adopted by McGuinness (1998), using the most common 3,000 English words. McGuinness points out that the probability structure based on 3,000 common words will be different to larger corpus works (Carney, 1994; Hanna et al, 1966) because of the different probabilities associated with Latin-origin words.

The phoneticity factor for whole words, which formed the basis of the predictive model, was determined by calculating the average PhR value for each word. As an illustration, Table 3 indicates the calculation of PhR values for a single word ('because'). The average PhR values for all the words used in the study are given in Table 1. Carney does not provide individual values for all representations of phonemes, but it is possible to use Carney's figures to compute average PhR values for the 40 sample words used in this study. The correlation between average PhR values obtained by the method described above and those derived from Carney's tables is statistically highly significant (r=0.85, p<0.001), a result that challenges McGuinness's assertion concerning word sample size.

Table 3. PhR values for phoneme representations and average PhR value for the word 'because'.

Phoneme	Form of Representation	PhR of representation (%)
b	b	99.43
ɪ	e	24.80
k	c	69.80
ɒ	au	0.32
z	se	19.85
Average PhR value of word:		**42.84**

Word frequency (WF)

The frequency of occurrence was also found to influence the spelling of words in the national tests: and it seems reasonable to suggest that the more common a word is, the more likely it is that unusual representations of sounds will be internalised by the learner. The total number of occurrences of a word within the entire LOB Corpus (frequency per million words) was used as the frequency factor in the analysis.

Probability value

The words used in the SCAA national spelling tests for 1996 are shown in Table 1. The data available from SCAA were in the form of percentage correct scores for each word (not shown in Table 1). This score was converted to a probability value for use in the regression analysis. The following formula was also used in the present study to convert percentage correct scores for each word into a probability value (PV):

$$PV = Log_{10}(\text{probability right/probability wrong})$$

RESULTS

The analysis of the data was undertaken with the simultaneous multiple regression module in the Statistical Package for the Social Sciences (SPSS, version 6.1.1 for Macintosh computers). Regression methods utilise the presence of an association between two variables to predict the values of one from those of another (Norusis, 1991). The regression analysis of the SCAA data predicted the spelling behaviour on the basis of three factors: frequency, length and phoneticity. The present analysis was undertaken to test the general applicability of these factors within a whole school context, across the five year-groups, for the 40 words used in the SCAA tests for 1996.

The percentage correct values for each of the 40 words are given in Table 4. The mean percentage correct value for all words for each year group increases steadily from 30.89% to 67.30%. Statistically significant correlations were obtained between two of the three factors (average phoneticity; number of letters) and the standardised spelling score, across all year-groups (see Table 5). The third factor (word frequency) was statistically significant across year-groups 3, 4 and 5, as predicted by the SCAA

Table 4. Percentage correct values, age and reading quotients for each year group.

	Year 2	Year 3	Year 4	Year 5	Year 6
because	17.78	69.57	71.05	79.17	87.8
boat	62.22	65.22	78.95	83.33	87.8
bucket	17.78	26.09	55.26	45.83	70.73
family	31.11	52.17	65.79	68.75	63.41
fish	66.67	78.26	94.74	97.92	100.00
flag	82.22	84.78	94.74	91.67	97.56
friends	35.56	60.87	68.42	43.75	68.29
hand	73.33	82.61	97.37	97.92	95.12
hat	95.56	97.83	97.37	97.92	97.56
holiday	28.89	65.22	65.79	75.00	63.41
house	73.33	78.26	73.68	89.58	92.68
morning	22.22	56.52	71.05	72.92	85.37
net	88.89	84.78	94.74	97.92	97.56
pictures	11.11	13.04	47.37	45.83	78.05
road	48.89	65.22	94.74	97.92	95.12
shout	31.11	65.22	65.79	83.33	82.93
smile	33.33	50.00	76.32	77.08	82.93
sock	44.44	63.04	84.21	70.83	90.24
spade	35.56	45.65	63.16	60.42	78.05
wait	22.22	43.48	73.68	66.67	65.85
beautiful	4.44	23.91	13.16	20.83	34.15
crept	48.89	54.35	57.89	70.83	70.73
disturbed	4.44	4.35	15.79	31.25	63.41
echoed	6.67	1.00	13.16	25.00	43.90
heard	15.56	34.78	42.11	54.17	63.41
honest	11.11	2.17	28.95	27.08	51.22
notice	13.33	17.39	26.32	47.92	56.10
piece	6.67	4.35	21.05	18.75	56.10
remained	4.44	10.87	15.79	33.33	39.02
replace	22.22	39.13	52.63	64.58	80.49
shook	24.44	4.35	18.42	16.67	29.27
silence	15.56	10.87	23.68	39.58	60.98
slipped	13.33	4.35	7.89	8.333	34.15
sneeze	15.56	4.35	15.79	25.00	41.46
sprawling	6.67	2.17	5.26	25.00	31.71
still	44.44	54.35	76.32	77.08	87.80
stretched	6.67	2.17	5.26	16.67	14.63
tallest	15.56	19.57	34.21	43.75	63.41
uncoiled	11.11	13.04	21.05	29.17	29.27
visitors	22.22	10.87	21.05	31.25	53.66
Mean % score	30.89	39.16	51.25	56.25	67.13
(S.D.)	25.30	29.96	30.68	27.99	23.20
Average age	7.08	8.07	9.15	10.14	11.08
(S.D.)	0.28	0.34	0.31	0.29	0.29
Average reading quotient	99.79	102.24	104.05	99.44	96.04
(S.D.)	13.26	19.58	13.17	10.93	13.37
N	49	45	44	46	52

data analysis. There is a smaller, but statistically significant, negative correlation between phoneticity and word length, suggesting that there is a tendency for longer words to have lower phoneticity.

Table 5. Correlation coefficients for Probability Value (PV), Phoneticity (PhR), Word Length (WL) and Word Frequency (WF) for the five year-groups (Y2–Y6).

	PV	PhR	WL	WF
PV	1.000	Y2) 0.63**	Y2) −0.81**	Y2) 0.21
		Y3) 0.65**	Y3) −0.65**	Y3) 0.35**
		Y4) 0.60**	Y4) −0.79**	Y4) 0.32*
		Y5) 0.63**	Y5) −0.73**	Y5) 0.33**
		Y6) 0.46**	Y6) −0.69**	Y6) 0.28
PhR			−0.35**	0.01
WL				−0.13

Number of cases :40. Correlation: 1-tailed significance (**$p < 0.01$; *$p < 0.05$).

The multiple regression models, based on the 3 factors, show impressive multiple R values across the year groups (see Table 6). The multiple regression analysis enables predictions to be made about the number of pupils in a class spelling a word correctly on the basis of the three characteristics of the word: number of letters, phoneticity and frequency. The R^2 value indicates the proportion of the total variation in spelling behaviour explained by the three factors. For years 2 to 5 at least 70% of the variation is accounted for by the factors.

Table 6. Multiple R values for the five year-groups (Y2 – Y6).

	Y2	Y3	Y4	Y5	Y6
Multiple R	0.90	0.84	0.89	0.87	0.75
R Square	0.81	0.71	0.80	0.76	0.57
Adjusted R Square	0.79	0.69	0.78	0.74	0.53
Standard Error	0.28	0.47	0.36	0.38	0.49

DISCUSSION

From data collected by SCAA for more than 2,500 children in 1996, factors were identified which predict the percentage of students likely to correctly spell words at ages 7 and 11. The factors identified are those which are arrived at by any common-sense view of the level of difficulty that words present to students: number of letters, the presence of unusual forms of phonemic representation, and frequency of

occurrence. The results indicated that between ages 7 and 11 years, as children have more commerce with words and expand their vocabulary, the frequency of occurrence of a word exerts an influence: increased frequency ameliorates the effects of obscure phoneme representations (i.e. words with low average phoneticity).

The present study confirmed the importance of these factors across ages 7 to 11 years, in a typical primary school, demonstrating that word frequency is not an important factor in the very early years of schooling, but rapidly begins to influence spelling behaviour at age 8 years.

The results are relevant to the teaching of reading and spelling, and also the design of spelling tests. Rego (1991) and Lazo, Pumfrey and Pear (1997) have demonstrated the reciprocal nature of spelling and reading, supporting McGuinness's (1998) suggestion that teaching the reversibility of the code is an essential feature of any effective literacy programme. With a perfect spelling system, phonemic awareness training alone should ultimately lead to near perfect spelling and reading for most of the population. With an imperfect code this will never be the case. Only carefully sequenced procedures that introduce representations of phonemes according to their frequency of use, such as the Phono-Graphix method (C., D. and G. McGuinness, 1996) or the Auditory Discrimination in Depth method used by Howard (1982), may be wholly successful.

The model derived from this study clearly indicates that phoneme representation probabilities play a crucial role in spelling. Words with phonemes that are represented by low probability forms are more likely to be spelled incorrectly. Controlling for word length and frequency the average correlation between the phoneticity measure and word scores is 0.62 ($p < 0.0001$). To an extent, low probability representations are ameliorated by higher frequency in the language.

There are a number of educational implications from the proposed model of spelling behaviour. For example, the model could be used to derive difficulty weightings for words, and this would be relevant to designers of teaching and test materials. Clearly, further research is necessary to corroborate the predictive value of such a model, and a start has been made on this using data provided by Mitton (1996), which was obtained from 1,000 adults forming a quota sample stratified by sex, age, and social class. Predicted probability values (PV) were calculated for each word using average regression coefficients for the three factors from the present study, and these were compared with the actual values obtained from the sampled population. The Pearson correlation coefficient was statistically significant ($r = 0.9$, $p = 0.014$), confirming the predictive value of the model.

There is often criticism of poor spelling in schools and even at university level. This study confirms the conclusions from the analysis of the national test data (Spencer, 1998) that a major factor in poor spelling, which may also be reflected in poor reading, is the failure of English orthography to conform to specific rules for the representation of phonemes. Across ages 7 to 11 years, the more unusual the written forms, the more difficult they are to spell, although unusual representations may be mitigated by more frequent occurrence of the words. Even bizarre representations are learned if they are frequently encountered.

By the age of 11 years, most students are able to deal successfully with all but the most unusual written forms for word sounds. It can be argued from the data presented that by regularising such highly irregular forms, spelling, reading and the self-confidence of these students could be greatly enhanced.

REFERENCES

CARNEY, E. (1994) *A Survey of English Spelling*. London: Routledge.

FRITH, U. (1985) Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall and M. Coltheart (eds.) *Surface Dyslexia*. London: Erlbaum, 301–330.

HANNA, P. R., HANNA, J. S. and HODGES, R. E. (1966) *Phoneme-grapheme correspondences as cues to spelling improvement*. Washington: US Department of Health, Education, and Welfare, OE-32008.

HOFLAND, K. and JOHANSSON, S. (1982) *Word frequencies in British and American English*. Bergen: Norwegian Computing Centre for the Humanities.

HOWARD, M. (1982) Utilizing oral-motor feedback in auditory conceptualization. *Journal of Educational Neuropsychology*, 2, 24–35.

LAZO, M. G., PUMFREY, P. D. and PEERS, I. (1997) Metalinguistic awareness, reading and spelling: roots and branches of literacy. *Journal of Research in Reading*, 20(2), 85–104.

McGUINNESS, D. (1998) *Why children can't read: and what we can do about it*. London: Penguin.

McGUINNESS, C., McGUINNESS, D. and McGUINNESS, G. (1996) Phono-Graphix: a new method for remediating reading difficulties. *Annals of Dyslexia*, 46, 73–96.

MITTON, R. (1996) *English spelling and the computer*. London: Longman.

NORUSIS, M. J. (1991) *The SPSS guide to data analysis*. Chicago: SPSS Inc.

REGO, L. B. (1991) *The role of early linguistic awareness in children's reading and spelling*. Unpublished DPhil Thesis. Oxford: University of Oxford.

SCAA (1997a) *Standards at Key Stage 1 English & Mathematics. Report on the 1996 National Curriculum Assessments for 7-year-olds*. London: Qualifications and Curriculum Authority.

SCAA (1997b) *Standards at Key Stage 2 English, Mathematics & Science. Report on the 1996 National Curriculum Assessments for 11-year-olds*. London: Qualifications and Curriculum Authority.

SPENCER, K. A. (1996) Recovering reading using computer mastery programmes. *British Journal of Educational Technology*, 27(3), 191–203.

SPENCER, K. A. (1998) A predictive model of spelling behaviour in 7 to 11 year-olds. *Journal of the Simplified Spelling Society*, 24(2), 23–7.

Address for correspondence: KEN SPENCER, Centre for Educational Studies, Institute for Learning, University of Hull, Hull, HU6 7RX, UK.
E-mail: *k.a.spencer@educ.hull.ac.uk*

15. Spencer, K.A. (2000) Is English a dyslexic language? Orthographic transparency and multilingual teaching, Dyslexia (Journal of the British Dyslexia Association), 6, 152-162.

Is English a Dyslexic Language?

Ken Spencer*

University of Hull, UK

McGuinness has suggested that there 'is no diagnosis and no evidence for any special type of reading disorder like dyslexia', and that poor teaching accounts for low levels of English literacy performance, rather than inherent personal deficits. Implicit in this is the assumption that some languages have simple grapheme–phoneme codes in which there is a one-to-one mapping, making them easy to teach and learn, while others have more complicated structures and are more difficult for teachers and students. There is now an increasing number of studies which demonstrate that readers in more transparent orthographies such as Italian, Spanish, Turkish, Greek and German have little difficulty in decoding written words, while English children have many more problems. Increasingly, lack of orthographic transparency in English is seen as having a powerful negative effect on the development of reading skills in English-speaking children. There is evidence that English-speaking children who fail to acquire reading skills may fall into two distinct categories: those who would succeed in languages, other than English, that have greater orthographic consistency; and those who would still have problems even with perfect orthographic transparency. The first, larger, group is let down by the interaction of poor teaching methods and an incomprehensible system of orthography. The present study examines word factors associated with poor spelling and reading that have been identified. Three factors account for the relative ease with which pupils can spell words: frequency of the word in the English language; length of the word; and the presence of 'tricky' letters or letter combinations. Data are presented illustrating the predictive model of spelling and reading which enables word difficulty to be calculated from the characteristics of English words. The implications the model has for teaching and learning English are elaborated, with reference to the possible benefits to be derived from mother-tongue teaching in British schools. Copyright © 2000 John Wiley & Sons, Ltd.

* Correspondence to: Ken Spencer, Centre for Educational Studies, The Institute for Learning, The University of Hull, Hull HU6 7RX, UK. E-mail: k.a.spencer@educ.hull.ac.uk

INTRODUCTION

Technology in education has a longer history than is often realized. First, there was language, the powerful tool that allowed accumulated knowledge to be passed across the generations, speeding up the human evolutionary process. Then, written language appeared about 5000 years ago, allowing human thoughts and ideas to be transmitted across almost limitless vistas of time. Again, the process of human evolution was speeded up:

> The invention of writing was highly significant for the development not only of language, but of society, and favoured the progress of commerce. It confirmed the power of the priests through the trained scribes, and even more the might and prestige of the ruler (Singer, Holmyard and Hall, 1954).

McLuhan (1964) introduces the Greek myth of Cadmus and the alphabet to suggest that alphabetic writing spells the end of the stationary temple bureaucracy and the priestly monopolies of knowledge and power.

> Unlike pre-alphabet writing, which with its innumerable signs was difficult to master, the alphabet could be learned in a few hours. The acquisition of so extensive a knowledge and so complex a skill as pre-alphabetic writing represented, when applied to such unwieldy materials as brick and stone, insured for the scribal caste a monopoly of priestly power. The easier alphabet and the light, cheap, transportable papyrus together effected the transfer of power from the priestly caste ... The phonetic alphabet is a unique technology. There have been many kinds of writing, pictographic and syllabic, but there is only one phonetic alphabet in which semantically meaningless letters are used to correspond to semantically meaningless sounds (pp. 82–83).

Robert Calasso (1993), in his rewriting of the Greek myths, suggests that Cadmus had:

> ... brought to Greece 'gifts of mind': vowels and consonants yoked together in tiny signs, 'etched model of a silence that speaks'—the alphabet ... no one could erase those small letters, those fly's feet that Cadmus the Phoenician had scattered across Greece ... (p. 390).

McLuhan and Calasso, however, are writing in two completely different types of orthography (English and Italian), although both are generally considered to use the same technology: the phonetic alphabet. It is becoming clear that the implementation of the phonetic alphabet varies, and this produces a range of writing technologies of varying degrees of effectiveness. McLuhan suggests that, unlike the more cumbersome pre-alphabetic systems, the alphabet could be learned in a short time, breaking the monopoly of the scribes and priests: making written language accessible to the whole population. Written language, in alphabetic form, was being democratized.

The basic idea of alphabetic systems is a one-to-one mapping of phonemes to graphemes. Some languages manage this (e.g. Finnish and Turkish), some deviate a little (e.g. Italian, Greek and Spanish), and for others the deviation is considerable (English).

McGuinness (1998) has suggested that there 'is no diagnosis and no evidence for any special type of reading disorder like dyslexia', and that

poor teaching accounts for low levels of English literacy performance, rather than inherent personal deficits. Implicit in this is the assumption that some languages have simple grapheme–phoneme codes in which there is a one-to-one mapping, making them easy to teach and learn, while others have more complicated structures and are more difficult for teachers and students.

There is, indeed, a growing body of evidence which shows that languages such as Turkish, Italian and German, with their more regular orthographies, are much less difficult for beginning readers than English.

Oney and Goldman (1984) confirmed that this was the case with Turkish, which was radically modernized in the 1920s. They compared first- and third-grade pupils in American and Turkish schools (Figure 1). Turkish pupils were faster and more accurate at reading words than their American counterparts, and there was little difference between Turkish first- and third-grade pupils. McLuhan's claim for the ease of acquisition of alphabetic writing seems to be vindicated in the case of Turkish, but not for English. Turkish first-grade children actually outperformed American third-grade children on most words, and especially on three-syllable words. Turkish, with its almost perfect grapheme-to-phoneme correspondence, resulting in orthographic transparency, is clearly a more efficient writing technology than the system used for English.

But even with languages with almost perfect orthographic transparency there will always remain a group of children that still have reading difficulties. Landerl, Wimmer and Frith (1997) conducted a stringent comparison of these difficulties in typical German and English dyslexics and found that:

> English dyslexic children suffered from much more severe impairments in reading than German dyslexic children . . . the differences between the English and the German dyslexic children were comparatively small for the short, high frequency words. For all other conditions, however, the differences were enormous (p. 328).

This is dramatically evident from the data in Figure 2. Similar results were also reported for normal children in a later study (Frith, Wimmer and Landerl, 1998). These results lead to a less extreme view than that suggested by McGuinness: even in more perfect writing systems there will be some
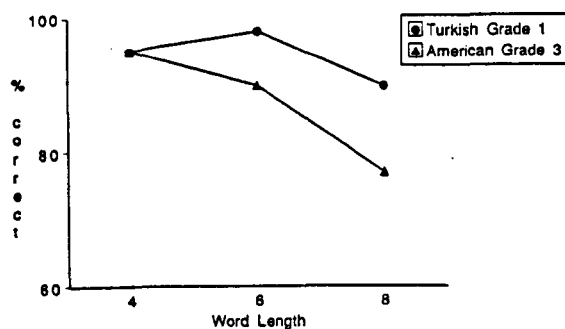
Figure 1. Comparison of reading performance on four-, six- and eight-letter words for American third-grade and Turkish first-grade children (based on Oney and Goldman, 1984).
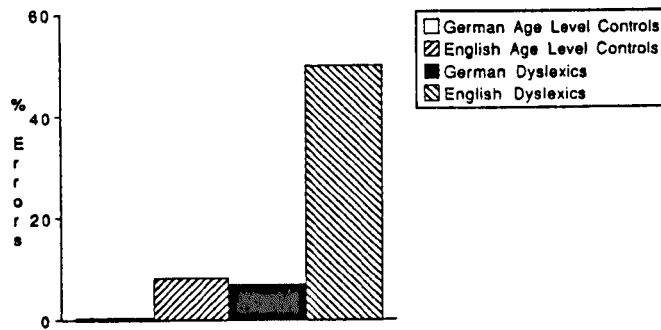
Figure 2. Comparison of reading performance on low-frequency words for German and English dyslexic children and age level controls (based on Landerl, Wimmer and Frith, 1997).

individuals who, for a variety of reasons, have reading/writing difficulties; but the burden is much heavier for those individuals who have to deal with the unnecessary complexities of codes in such languages as English. As Frith, Wimmer and Landerl (1998) point out:

> Clearly, the success of phonological recoding depends on whether the word to be identified conforms to the code. In the case of irregular words, which are frequent in English, phonological recoding, if strictly applied, leads by definition to an incorrect word sound . . . Learning to read and write in a consistent orthography should not lead to such difficulties. Here, grapheme–phoneme recoding is reliable, and the assembly of phonemes results in pronunciations close to that of the target word (p. 32).

A perfect spelling system is one that has no alternative spellings for the same sound, and no overlap in the code where one spelling pattern stands for different sounds. The main problem for readers is that English does not have such a system. It follows from this that spelling errors, as well as reading errors, will be related to some measure of the phoneticity of words, where phoneticity is a measure of the deviation of a word from a truly phonetic rendition.

MEASURING THE PHONETICITY OF WRITTEN ENGLISH

English is known to be a deviant language, but just how deviant is it? There is no commonly accepted index of phoneticity for the English language. However, Spencer (1998, 1999a,b) has suggested a method for deriving phoneticity measures for English words, which has been shown to be related to pupil spelling performance. In particular, the representation of a phoneme as a proportion of all representations of that phoneme—PhR—showed very significant correlations with pupil spelling performance. This measure considers a particular representation of a phoneme only in relation to other representations of the phoneme. The procedure used to derive these values is similar to those described by McGuinness (1998) and Carney (1994). Knowing the frequency of each form of representation for each phoneme

allows an average phoneticity value to be calculated for words. The correlation (0.85) between phoneticity values for words using PhR values derived from Spencer (1998) and Carney (1994) is statistically highly significant.

In addition to average PhR values for words, particularly unusual phonemic representations can be identified within each word. The most infrequent form of representation within each word gives a value for the 'tricky' phoneme, in terms of PhR values. The 'tricky' phoneme value can be seen as representing the most difficult element of a particular word.

Figure 3 demonstrates the remarkable constancy of the average phoneticity of words in the English language over a wide range of frequencies. The number of phonemes climbs steadily as words become less frequent, from a mean value of 3 for the most common words, to 7 for the random 1% sample taken from the 20000-word dictionary. The mean value of the most difficult phoneme falls from 34 as the number of phonemes rises, to a plateau of 18–20 in the mid-range of frequencies, and dips further to a low of 13 in the random sample. This means that although the average phoneticity remains relatively constant throughout the frequency range, the most difficult or unpredictable element within words, as represented by the 'tricky' value, becomes even more unpredictable as the words become less frequent.

Using data from national tests carried out by the SCAA in 1996 (School Curriculum and Assessment Authority, 1997a,b), a predictive model of spelling behaviour was developed (Spencer, 1998) which suggested that readily identifiable characteristics of words can be used to determine word difficulty in spelling tasks. The model identified three factors which contributed to word difficulty: number of letters in the word; phoneticity of the word; and frequency of occurrence of the word. However, if it is to be a useful educational tool, the model must be shown to be sufficiently robust to account for the behaviour of pupils in many contexts. The present research has been undertaken to determine the extent to which the factors identified in the national studies were relevant in predicting spelling behaviour in a typical school situation, across ages 7–11 years.
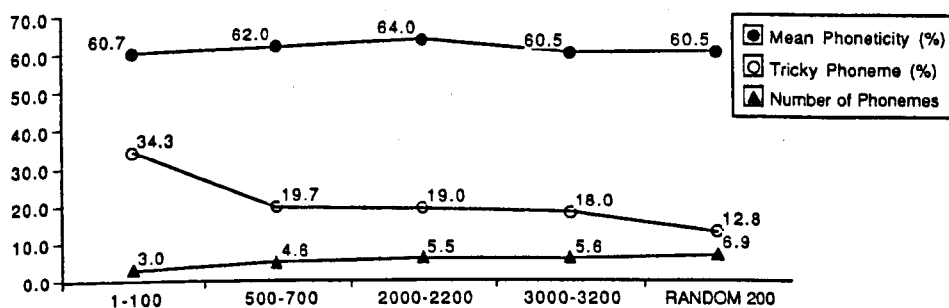
Figure 3. Mean phoneticity (PhR), mean 'tricky' phoneme value and mean number of phonemes for English words in the frequency range 100–3200 words and random 1% sample of 20000 words.

METHOD

Subjects

In the present study, data were collected during February 1998 for all pupils at a local primary school which performs at average national SAT levels in English, maths and science. The collected data of spelling performance spanned five year groupings (Y2–Y6), for a total of 236 pupils. Mean scores for the 90 words used in the test, age and reading quotients for each year group are given in Table 1.

Table 1. Percentage correct values for 90 words, age and reading quotients, year groups 2–6

	Y2	Y3	Y4	Y5	Y6
Mean % score	32.98	43.24	52.27	68.53	75.33
(SD)	27.31	30.95	31.80	26.33	25.41
Average age	7.08	8.07	9.15	10.14	11.08
(SD)	0.28	0.34	0.31	0.29	0.29
Average reading quotient	99.79	102.24	104.05	99.44	96.04
(SD)	13.26	19.58	13.17	10.93	13.37

Procedure

The SCAA 1996 national spelling data were collected from two sets of 20 words, one set being administered to each Key Stage (KS). The 1997 national spelling data were collected from two sets of words, a set of 20 being administered to KS1 and 30 to KS2. The 90 words from the 1996 and 1997 national tests for KS1 and KS2 were administered to pupils during the morning session on five consecutive days. A total school population of 246 was reduced to 236 because of subject attrition over the five days.

Class teachers administered the spelling tests, giving the word followed by an example sentence, and a further repetition of the word. There was no time limit for the test. Pupils wrote their answers on forms with word numbers clearly marked.

The 1996 data available from the SCAA were in the form of percentage correct scores for each word. This score was converted to a probability value for use in the regression analysis. The following formula was used:

$$\text{probability value (PV)} = \log_{10}(\text{probability right}/\text{probability wrong})$$

Percentage correct scores for each word in the present study were converted to probability values using the above formula.

Word Characteristics

Values for the factors which had been found to predict spelling performance in the national tests (number of letters; phoneticity—PhR and 'tricky' phoneme; and word frequency) were obtained for each of the 90 words

Table 2. Number of letters, average phoneticity, 'tricky' phoneme and frequency for 90 test words

	Number of letters	PhR	Tricky phoneme	Frequency
Mean	6.09	59.63	15.66	201.92
SD	1.80	15.67	23.79	381.14

taken from the 1996/7 national tests for 7- and 11 year-olds. Table 2 gives the summary values for the 90 words used in this study.

RESULTS FOR WHOLE WORD SCORING METHOD AND ANALYSIS OF DATA

The words were scored as either correct or wrong for each student. This is the traditional way of marking spelling in schools, and the method used for the national test data for 1996/7. This method requires pupils to achieve a 100% criterion level for each word for that word to be counted as correct. The mean percentage value for the number of correct spellings for each word for each year group was then calculated. This percentage score was converted to a probability value using the formula given above.

The analysis of the data was undertaken with the multiple regression module in the Statistical Package for the Social Sciences (SPSS, version 6.1.1 for Macintosh computers). Regression methods utilize the presence of an association between two variables to predict the values of one from those of another. The regression analysis of the SCAA data predicted the spelling behaviour on the basis of three factors. The present study was undertaken to test the general applicability of these factors within a whole school context, across five year groups, for the 90 words used in the SCAA tests for 1996/7.

Results of the Multiple Regression Analysis for Year Groups 2–6

The mean percentage correct values for the 90 words are given in Table 2. The mean values for all words for each year group increase steadily from 32.78% for year 2 to 75.33% for year 6.

A summary of the results for the multiple regression analyses for the five year groups is given in Table 3. The regression analyses were undertaken with both measures of phoneticity: PhR and the value of the 'tricky' phoneme. In each year group, the 'tricky' value for each word increased the predictive power of the model when compared with the average phoneticity (PhR) value.

Statistically significant results were obtained between two of the three factors (number of letters; 'tricky' phoneme) and the spelling scores across all year groups. The third factor (frequency) was statistically significant for three year groups. The multiple regression models, based on the three factors, show impressive multiple R values across the year groups: mean multiple $R = 0.76$, mean adjusted $R^2 = 0.59$.

Table 3. Multiple *R* and adjusted *R²* values for spelling regression model, and significance of *t*-test values for variables: number of letters, value of 'tricky' phoneme and frequency of word

	Y2	Y3	Y4	Y5	Y6
Multiple *R*	0.79	0.70	0.82	0.76	0.77
Adjusted *R²*	0.61	0.47	0.67	0.60	0.58
Number of letters (*p* =)	0.00	0.00	0.00	0.00	0.00
'Tricky' phoneme (*p* =)	0.00	0.03	0.00	0.00	0.00
Frequency (*p* =)	0.04	0.05	0.36	0.10	0.02

RESULTS FOR PHONEME SCORING METHOD AND ANALYSIS OF DATA

An alternative to the traditional method of scoring spelling is to score the representation of each phoneme in a word as correct or wrong. This method is not usually used in classrooms, but it does give a clearer picture of pupil performance in that it identifies the degree of error within a word, and the nature of that error. In this case the mean percentage value for the number of correct spellings of each phoneme in each word was calculated for each year group (Y2–Y5). Again, this percentage score was converted to a probability value using the above formula. The regression analysis considered the same variables as for the whole word scoring method, and, in addition, the position of the phoneme in the word.

A summary of the multiple regression analyses for the four groups is given in Table 4. Statistically significant results were obtained for three of the variables: the PhR value for the phoneme; the position of the phoneme in the word; and the number of letters in the word. The multiple regression model, based on the three variables, shows consistent high multiple *R* values across all year groups: mean multiple *R* = 0.66, mean adjusted *R²* = 0.40.

Figure 4 shows the debilitating effect that increasingly infrequent forms of representation of phonemes have on pupil spelling. Phonemes whose representations are highly predictable (81%–100% range) are correctly spelled by most pupils, irrespective of age. Performance deteriorates as the form of representation becomes less predictable, and the decline is more pronounced for younger children. This result supports the data from Oney and Goldman (1984), showing that for representations that are stable and predictable (i.e.

Table 4. Multiple *R* and adjusted *R²* values for spelling regression model, and significance of *t*-test values for variables: PhR for phoneme, position of phoneme and number of letters in word

	Y2	Y3	Y4	Y5
Multiple *R*	0.58	0.71	0.58	0.68
Adjusted *R²*	0.33	0.50	0.33	0.46
PhR for phoneme (*p* =)	0.00	0.00	0.00	0.00
Phoneme position (*p* =)	0.00	0.00	0.00	0.00
Number of letters in word (*p* =)	0.00	0.00	0.00	0.00
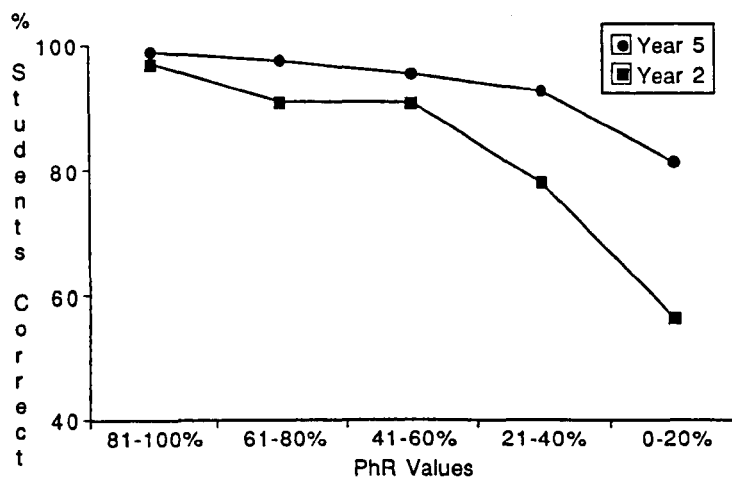
Figure 4. Mean percentage of students correctly spelling phonemes (year groups 2 and 5) vs. PhR values for phonemes.

showing high orthographic transparency, as is found for most Turkish words), performance will be high and there will be little difference between pupils of different ages.

DISCUSSION

The results from many sources are clearly indicating the deficiencies of the writing technology for English. The whole word scoring method results confirmed those obtained from the national tests: phoneticity is a consistent, major factor in predicting pupil spelling across school years 2–6. The phoneme analysis results from this study demonstrate the role that arbitrary, unpredictable forms of representation play in hindering pupil achievement in this basic educational skill. The results show that if all phonemes had highly predictable, consistent forms of representation, English children would show the same high level of performance, irrespective of age, that is found with subjects using more regular orthographies.

Children certainly ought to be made aware of the fact that English is a difficult written language, and that most of the problems they will face will be the fault of the poor technology they are using rather than personal deficits. This will be immediately apparent to children who are multilingual, because whatever additional languages children speak, they are unlikely to be as inconsistent as English in their orthography. A case can certainly be made to encourage multilingual children to learn to read in languages other than English, because their progress, generally, will be faster. Children who are being educated in an English language environment, but who have access to another language in the home environment, may well benefit from the provision of text translations into the home language if that language is more orthographically transparent than

English. This further strengthens the case for the provision of interpreters and translations of materials in key curriculum areas.

The great challenge here is to determine the extent to which English orthography exacerbates dyslexic tendencies in children. McGuinness, McGuinness and McGuinness (1996) argue that the orthography causes most of the problems that their clients face, and that their methods of highly structured teaching can overcome these problems. However, Landerl, Wimmer and Frith (1997) show that even with regular orthographies dyslexia is present, although the difficulties are mitigated by the orthography. German dyslexic children make slightly less errors than their English age level controls, and this may well suggest the scale of the damage done to English pupils by the inconsistent writing technology. If English pupils are so damaged by their orthography that their performance is worse than dyslexic pupils in other more orthographically transparent languages, then English can truly be said to be a dyslexic language.

In order to make the English system of writing a more efficient and effective technology, there must be reform. However, there have been so many attempts at reform that have failed, that it seems likely that the present system will remain in place for some considerable time, probably until international demands are made for International English to be standardized. The consequences of this must be accepted. Children using English as an educational medium will be disadvantaged; dyslexic children will be greatly disadvantaged; and the most disadvantaged group of all may be dyslexic children for whom English is an additional language.

References

Calasso, R. (1993) *The Marriage of Cadmus and Harmony*. New York: Alfred A. Knopf.

Carney, E. (1994) *A Survey of English Spelling*. London: Routledge.

Frith, U., Wimmer, H. and Landerl, K. (1998) Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2, 31–54.

Landerl, K., Wimmer, F. and Frith, U. (1997) The impact of orthographic consistency on dyslexia: a German–English comparison. *Cognition*, 63, 315–334.

McGuinness, C., McGuinness, D. and McGuinness, G. (1996) Phono-Graphix: a new method for remediation reading difficulties. *Annals of Dyslexia*, 46, 73–96.

McGuinness, D. (1998) *Why Children Can't Read: and What We Can Do about It*. London: Penguin.

McLuhan, M. (1964) *Understanding Media*. London: Routledge and Kegan Paul.

Oney, B. and Goldman, S.R. (1984) Decoding and comprehension skills in Turkish and English: effects of the regularity of grapheme–phoneme correspondences. *Journal of Educational Psychology*, 76(4), 557–568.

School Curriculum and Assessment Authority (1997a) *Standards at Key Stage 1 English and Mathematics. Report on the 1996 National Curriculum Assessments for 7-year-olds*. London: QCA.

School Curriculum and Assessment Authority (1997b) *Standards at Key Stage 2 English, Mathematics and Science. Report on the 1996 National Curriculum Assessments for 11-year-olds*. London: QCA.

Singer, C., Holmyard, E.J. and Hall, A.R. (1954) *A History of Technology, Volume 1, Part 1, Chapter 4: Speech and Language.* Oxford: Oxford University Press.

Spencer, K.A. (1998) A predictive model of spelling behaviour in 7 to 11 year-olds. *Journal of the Simplified Spelling Society,* 24, 23–27.

Spencer, K.A. (1999a) Predicting spelling behaviour in 7 to 11 year-olds. *Reading,* 33(2), 72–77.

Spencer, K.A. (1999b) Predicting word-spelling difficulty in 7- to 11-year-olds. *Journal of Research in Reading,* 22, 283–292.

3. DISCUSSION OF THE PAPERS

Papers in Series 1

In Chapter 1 the failure of many new media technologies to deliver significant learning gains was discussed and related to the first series of papers presented in this submission. These papers (Spencer, 1977, 1981a, b, c & d, 1988 Chapter 5) represent a coherent strand of research which sought to understand the nature of the 'no significant difference' effect, and to place it into context, particularly the context of cost and effectiveness.

Spencer (1977) concludes that the results of the study, with university physics students, "tend to confirm the general 'no significant difference' results in other media comparisons". It was noted that although students preferred the high cost television productions, the additional costs of either high- or low-cost videotaped presentations could not be justified. Spencer (1981a) provides similar evidence, but in a different context: the subjects were mentally handicapped adults. In this study the comparison was between film and tape-slide presentation, and led to the conclusion that "there is little difference between the two media presentations for the identification or information tests, a result which confirms those previously obtained with normal subjects, which run counter to the 'realism' theories."

Both studies show that the subjects do learn from the programmes, but that they learn no more from the higher cost media. The comparisons in both studies were between different media (videotape and audiotape; film and tape-slide); Spencer and Clarke (1981b), again working with mentally handicapped adults, compared teachers working with and without television programmes as teaching aids, and also with television as the sole means of instruction. The conclusion was that the "results do not indicate a single instance of an overall superior performance for the Ell [in which teachers had access to the television aids] group... These findings are not in fact surprising given the results of previous experiments involving the efficiency of television aids." The study also demonstrated that television could act as a substitute for teachers, and that the subjects learned as much from the videotapes as they did from the normal teaching situation, thus confirming Hoban and Van Ormer's (1950) conclusions.

These results were set in context in Spencer (1981c), which reviewed many media studies and concluded that the "results clearly demonstrate the effectiveness of instruction via the media, and while not lending full support to Wilkinson's suggestion that the long history of no significant difference between instructional strategies reduces the cost-effectiveness question to one of finding the cheapest method, it does lead to the inevitable conclusion that 'small is beautiful'." This sentiment was echoed in Spencer (1981d), which dealt with media attribute research (colour, motion and pictorial complexity) in an effort to illuminate the gap between audio-visual theorists, using 'realism theories' to justify the use of such aids, and the accumulating history of 'no significant difference' in studies evaluating different forms of mediated instruction.

The emerging pattern in media research could be described as 'counter-intuitive': it appeared obvious that the more realistic a visual aid was the more a student would learn from it, but the results from many research studies indicated that this was not necessarily the case. Cox (1976) reviewed many studies investigating the value of colour in visual displays and concluded that there was no marked difference in learning from colour or monochrome television or film; Dwyer (1978) demonstrated that on four measures of learning about the human heart, the most effective of the illustrations, which included realistic photographs and drawings, was the simple line drawing.

The key to understanding the results, particularly for media attribute research, was the relatively new area of cognitive psychology, especially the concept of human information processing capacity. This central concept, introduced in Spencer (1981d), indicated that "man can only handle information coming from one sensory input at a time, and that only a fraction of that sensory input can be processed" and therefore a combination of audio and visual presentations would not necessarily be beneficial because attention could only be directed to one sensory channel at a time and that, as the human processing system can handle only a fraction of the information passed to it by the sense organs, simplified displays, such as black and white drawings, would prove to be equally as useful as more complex, realistic displays.

However, it became apparent that the Travers-Broadbent model of human information processing could not account for all the available data, and this resulted in Spencer (1988, Chapter 5) claiming that "the single-channel theory proposed by Broadbent (1958) and elaborated by Travers (1967) is no longer wholly tenable, and it is reasonable to assume that varying allocations of attention can be made to incoming information." In addition, there was good evidence to support two complementary representation systems, along the lines suggested by Piavio (1971), with extensive interconnections between verbal and image systems. This is a position acceptable to multi-channel advocates and supports the early claims of the audio-visual theorists.

Spencer (1988, Chapter 5) offers a resolution of the results from media comparisons which demonstrate no significant differences, and the inter-modal comparisons that demonstrate that audio-visual presentations are often superior to single channel presentations. In particular, the results of Levie and Lentz's (1982) review, which demonstrated the benefits of illustrated text, had to be accommodated.

The conclusions, based on reviews of research and the experimental data presented in the first series of papers, are:

- The human information system has limited capacity and much of the information available to the senses is either not attended to by the processing system or is compressed in the early stages of perception.
- There are two complementary representation systems for verbal and image information, with extensive interconnections between the systems.
- Activation of both systems will generally result in higher levels of learning than activation of a single system.

- Increasing the complexity of visual input (e.g the addition of colour, visual motion) may not bring about commensurate increases in learning because the human information system has limited capacity.
- Audio-visual media capable of presenting complex visual stimuli (e.g. film and television) may not bring about increases in learning when compared with those capable of presenting less complex visual stimuli (tape-slide and printed materials) because the human information system has limited capacity and complex details are screened out at an early stage in processing.
- Students from a wide range of abilities learn as much from audio-visual presentations, and visual presentations that activate verbal and visual representation systems (e.g. illustrated texts), as from traditional methods.

It appears from the above results that there is little to choose between competing media. The cumulative 'no significant difference' results are currently being given an interesting spin in the drive to provide distance education: if there really are no differences, cost becomes the driving force.

Papers in Series 2

The second series of papers (Spencer, 1988 Chapters 1-4, 1990a, 1990b, 1992, 1996) are concerned with identifying methods which have been shown to produce measurable increases in achievement when compared with traditional classroom methods.

Chapter 1 of "The Psychology of Educational Technology and Instructional Media" identifies two strands in educational technology: the tools-technology, sometimes known as the hardware approach, but more commonly known under the title audio-visual aids, which has been deal with in the preceding section of this conclusion; and the second strand, dealing with the educational application of knowledge from the behavioural sciences, which underpins the second series of papers. The origins of a science of learning are presented within the context of Thorndike, Pavlov and Watson, and related to Skinner's development of programmed instruction (Chapters 1 and 2). The characteristics of programmed learning are evaluated and the conclusion suggests that the "results of these early research studies do seem to cast doubt on most of the characteristics of linear programmes, which were derived from Skinner's psychological theories", although the review demonstrates that students, from many different disciplines and age groups, do learn from programmed instruction. Meta-analytic reviews demonstrate a small effect size (0.24) for programmed instruction in higher education; and an even smaller effect size (0.08) when it is used in secondary education. The overall results of 30 years of experimental evaluations showed that the expectations set by Skinner for his technological revolution had not been fulfilled.

Skinner's work with programmed learning may not have fully succeeded, but it did have a profound effect on educational thought, particularly on the educational objectives movement (Chapter 3), and mastery learning approaches including Bloom's Learning for Mastery (LFM) programme, and Keller's Personalised System of Instruction (PSI) (Spencer, 1988, Chapter 4).

Bloom's (1968) LFM strategy is based on the idea of breaking courses down into small, manageable units, and the requirement of passing criterion referenced, formative tests at a high level before moving from one course unit to the next. An essential ingredient is the provision of optimal instruction per learner: students are provided with sufficient opportunities to enable them to reach high levels of performance in the formative tests. The system results in impressive learning gains when compared to traditional methods, with a large average effect size of 0.83.

Keller's (1968) PSI approach is similar in its demands for "unit-perfection" before students move on to the next module. Again, the students receive opportunities for additional support if they fail to reach the set criterion level (at least 80% performance), and they may take as many terminal tests as are required for them to learn to the criterion standard.

Both systems advocate a mastery approach to learning, and it is this component that seems to make a significant contribution to the superior performance of the systems when compared with traditional courses. The average retention effect size (for PSI and LFM studies) of 0.67 reflects this.

The results of researches demonstrating the effectiveness of mastery systems forms the basis for Spencer's (1991) conclusion concerning raising standards of performance:

> These results indicate that the technology for substantially enhancing learning is available and can be delivered using a variety of different techniques. Keller Plan uses traditional media vehicles, human tutors and print materials, to implement an interactive mastery approach; it represents a simple and effective method for raising student performance. Bloom's Learning for Mastery offers a similar approach within a classroom context. Effective computer programmes, which simulate human activities and are efficient at handling mastery requirements, are also available. What seems to be crucial is the application of the technology OF education rather than the provision of technology IN education. Unfortunately, the latter is frequently given priority when non-educational agencies are involved, ill-informed decision-makers being seduced by the superficial appeal of new hardware approaches.

This position was reflected in Spencer (1990a, 1990b), which presented the theoretical background to mastery learning techniques and demonstrated how it could be applied to the teaching of biological concepts:

> Teachers felt that a criterion of 95 per cent would deter most students and that they would simply abandon the program. This has not been the case. Even students of low ability have persevered and mastered the material, gaining a positive self-image in the process, to the surprise of their teachers. (1990a)

The teaching programme incorporated the optimal learning per pupil concept, central to Bloom's LfM, with the requirement for 95% success on the criterion test. The mean performance after exposure to the instructional materials (based on Dwyer's 1967 materials) was 60%; the average additional time required to shift performance from 60% to 95% was a further 80%. Thus, to achieve mastery, pupils required almost twice as much time.

This promising line of investigation in the teaching of biology formed the basis for the more substantial project which was supported by the National Council for Educational Technology "Portables in the Classroom" project (Spencer, 1996).

The HyperReading programmes were designed to teach reading skills indirectly through programmes which develop spelling skills. A more direct route would have been to utilise computer speech recognition features, but at the time, as van Daal and van der Leij (1992) suggested, adequate speech recognition facilities to evaluate reading-aloud responses were only available at prohibitively high prices, although this is no longer the case.

However, developments in the early 1990s made it possible to provide high quality speech via relatively low cost computers, making spelling type programmes possible. Reading skills actually benefit from this approach because a critical step in reading acquisition is made when children shift from a visual strategy for reading words to an alphabetic strategy, which is first learned in spelling and later carried over to reading. This is a crucial phase because in the course of spelling development, children learn more about how sounds relate to letters (Frith , 1985).

However, such computer-based programmes are only one medium out of many that can be employed in the classroom for the teaching of reading e.g. paired reading, or any method which places the learner in direct contact, on a regular basis, with an expert reader who encourages mastery.

The value of a systematic method of teaching reading may be called into question because the majority of pupils acquire their reading skills through a variety of less highly structured approaches. However, for children who are failing to get started a highly structured approach may prove to be more effective. Hence the project title: The Portable Reading Recovery Laboratory.

The mastery methodology employed may also be considered to place too much emphasis on rote learning, rather than general (problem-solving) strategies. However, the problems of illiteracy must be attributed not only to learner, or even teacher, deficiencies, but also to the written form of the language: written English is profoundly illogical, especially the most common 200 words, and the rules, such as they are, cannot always be applied in a regular manner. This problem is dealt with in greater detail in the third series of papers and will be commented upon in the next section of this conclusion. The Chinese have a similar problem with their written language and adopt a simple rote learning strategy as the most effective method, as do the Japanese.

The approach was successful for many children. In particular, Pupil S, who could not read the most common word in the English language after 5 years in school, had mastered 80 words within 14 hours of instruction with the mastery-based computer programme.

A key feature of the programme was the pupil management file, which kept a record of all responses made by the pupil, calculated performance scores and used these to move students through sequences of exercises. An 80% criterion level was set for promotion to more difficult sequences (e.g. requiring pupils to spell more letters in a word).

The application of mastery concepts to the computer-based learning medium clearly demonstrated that literacy deficits could be rapidly compensated for within the confines of the most common 100 words. But there was a conceptual weakness in the programme: it did include features that were known to have a beneficial effect on learning (recognition tasks, restricted response options, high criterion levels, backward/forward chaining etc.), but there was no theoretical underpinning for the actual sequencing of words presented. The words were simply grouped together in their relative frequency order, derived from Reid (1989). This meant that words in a group to be learned (e.g. the 10 most common words for the first group) varied in two main dimensions: the number of letters, and the regularity of grapheme-phoneme correspondences.

Papers in Series 3

The third series of papers (Spencer, 1998, 1999a, 1999b, 2000) investigate factors that affect pupil spelling, in order that pupil difficulties can be understood and that insights gained can inform the development of a mastery-based system for teaching literacy skills. As a starting point, it seemed logical that pupils experiencing difficulties should not be confronted with words that were too difficult, and that a mastery-based literacy programme would introduce words of increasing difficulty, rather than on the basis of frequency alone (as was the case with the common 100 words).

A word list of the English language indicating the relative difficulty of words was required. Unfortunately, no such list exists! This does lead to the very serious question: why are some words more difficult to spell (or read) than others?

Spencer (1998) indicates that there are identifiable characteristics of words which contribute to spelling difficulties experienced by children. The data were collected by SCAA in 1996, and represent the percentage of children (aged 7 and 11 years) correctly spelling groups of words. This percentage was used as the basis for defining the ease or difficulty of a word: the easier a word is, the larger the group of children who can spell it. Three factors were identified as independently contributing to the number of pupils correctly spelling a word: the frequency of the word in the English language (Hofland and Johanssen, 1982); the number of letters in the word; and a measure of the phoneticity of the word. Phoneticity is a new concept in the study of reading and spelling. It is related the idea of "depth of orthographic transparency" or

"orthographic consistency" (Landerl et al, 1997), but actually provides a numerical value based on an analysis of the language under investigation. For example, in a recent analysis (Spencer and Xing, in preparation) Italian has an average phoneticity value of 96%, Spanish 92%, English 70% and French 57%. The phoneticity values used in the third series of studies were calculated from a set of 3,000 words. Carney's (1994) data can also be used to calculate phoneticity values which correlated highly with Spencer's data [r = 0.85, p<0.001].

The general applicability of the model described in the initial paper was tested in Spencer (1999b). Children in years 2 through 6 were tested in a local school, using the 40 words from the 1966 SCAA spelling tests. The three factors accounted for 70% of the variation in spelling, confirming that such factors were generally applicable, and hence could be used to develop a mastery-based literacy programme. In theory, knowing the values of the 3 factors (word length, frequency, and phoneticity) for any word, a probability value could be calculated for the word, indicating how difficult it would be (in terms of how many children in a class could spell it). This idea was developed in Spencer (1999a), which criticised the THRASS (Davies and Ritchie, 1996) approach to phonetics on the grounds that teaching children that phonemes have different graphemes associated with them was only part of the task: the different representations also vary in their frequency, and it may be seen as unnecessarily confusing to teach children obscure forms of representation in the early stages of reading and writing. This criticism is implicit in McGuinness's (1998) basic code formulation. The paper (1999a) presented the 44 English phonemes in 3 groups associated with different representation probabilities, and introduced the hypothesis that the most common 150 words could be divided into difficulty groups sorted by their most difficult phoneme (i.e. the phoneme with the least frequent form of representation). Some words are made up entirely of stable, frequent forms of representation (e.g. man, not, must), whereas others contain a very obscure form of representation for one or more phonemes (e.g. could, through, because). Data obtained to test this hypothesis are available, and will be discussed in the next section.

The final paper in the series (Spencer, 2000) was originally presented at the BDA First International Multilingualism and Dyslexia Conference (Manchester, June, 1999). Data indicated that with a larger set of words (90 words from the SCAA 1996 and 1997 tests) the general model held, but that the predictive power of the model was improved when the value of the most difficult phoneme (i.e. least frequently represented - termed the "tricky" phoneme value) was entered rather then the average phoneticity of the word. A further analysis demonstrated that when individual phonemes in a word were scored, the most powerful predictor was the phoneticity value for the phoneme. There was little difference between year 2 and year 5 children for phonemes of high frequency values (81-100%), and most children performed at a high level of accuracy. However, phonemes with graphemes in the 0-20% range presented severe problems for year 2 and year 5 children. Further, the analysis of the structure of the English language suggested that the average value of the most difficult component of words across the frequency range fell in the 12-20% range. In other words, the structure of the language was such that the average word in any frequency range (apart from the most frequent 100 words) had at least one phoneme with a form of representation that was in the 0-20% range, which the data indicated

was the range that produced severe difficulties for children, especially younger children (year 2).

This leads inexorably to the conclusion that the English orthographic technology is inefficient and actually causes pupils to fail in their writing, as Landerl et al (1998) had suggested was the case for reading. The data on individual phonemes indicated that the regular phoneme representations (81-100%), which is the norm for orthographically transparent languages such as Italian and Turkish, can be spelled correctly by most pupils. Orthographically transparent languages do not have representations in the range 0-20% which produced most difficulty for the English children: all representations are in the "easy" range, so reading and spelling are easy.

Before discussing the implications, for a mastery-based literacy programme, of the conclusion that "English is a dyslexic language" the unpublished data of a study which has tested the hypothesis expressed in the 1999a paper will be described.

The 150 words that were categorized according to the value of their most difficult phoneme were presented in a spelling test throughout a local school (238 pupils), and the correlation between probability values and "tricky" values calculated. The correlations for all year groups are highly significant, as are the correlations between years (Table 3.1).

Correlations

		TRICKY	Year 2 PV	Year 3 PV	Year 4 PV	Year 5 PV	Year 6 PV
TRICKY	Pearson Correlation						
	Sig. (2-tailed)						
	N						
Year 2 PV	Pearson Correlation	.510**					
	Sig. (2-tailed)	.000					
	N	150					
Year 3 PV	Pearson Correlation	.579**	.828**				
	Sig. (2-tailed)	.000	.000				
	N	150	150				
Year 4 PV	Pearson Correlation	.572**	.770**	.780**			
	Sig. (2-tailed)	.000	.000	.000			
	N	150	150	150			
Year 5 PV	Pearson Correlation	.521**	.735**	.789**	.720**		
	Sig. (2-tailed)	.000	.000	.000	.000		
	N	150	150	150	150		
Year 6 PV	Pearson Correlation	.572**	.700**	.740**	.725**	.747**	
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	150	150	150	150	150	

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3.1 Correlations between probability scores (Years 2 to 6) and most difficult phoneme score (Tricky value).

The hypothesis that the relative frequency of the form of representation of a grapheme can be used to predict the difficulty of a word is demonstrated in Figure 3.1. The effects are quite dramatic for year 2 children: only 40% can spell words that have

unusual graphemes (i.e. frequencies < 5%); whereas 91% can spell the more regular words (representation frequency range 60-100%). Of course, this is only one factor, as the previous research had indicated. The regression model, using factors identified in previous studies, again was found to be highly predictive (Table 3.2), accounting for more than 60% of the variability for year 2 pupils. Its predictive power falls, as would be expected with such a restricted range of word frequency, but still accounts for 40% of the variability in year 6. The inter-correlations between years is consistently high and legitimizes the collapsing of data across the year groups. When the probability values for the entire school (years 2 to 6) are entered into the regression equation the model accounts for 62% of the variability.

	"Tricky"	Word length	Frequency	Adjusted R Square
Year 2	0.00	0.00	0.00	0.62
Year 3	0.00	0.00	0.00	0.53
Year 4	0.00	0.00	NS	0.46
Year 5	0.00	0.00	0.02	0.41
Year 6	0.00	0.00	0.02	0.46
Year 2 to 6 combined	0.00	0.00	0.00	0.62

Table 3.2 Statistical significance values (p<) for 3 factors and predictive model adjusted R square values.
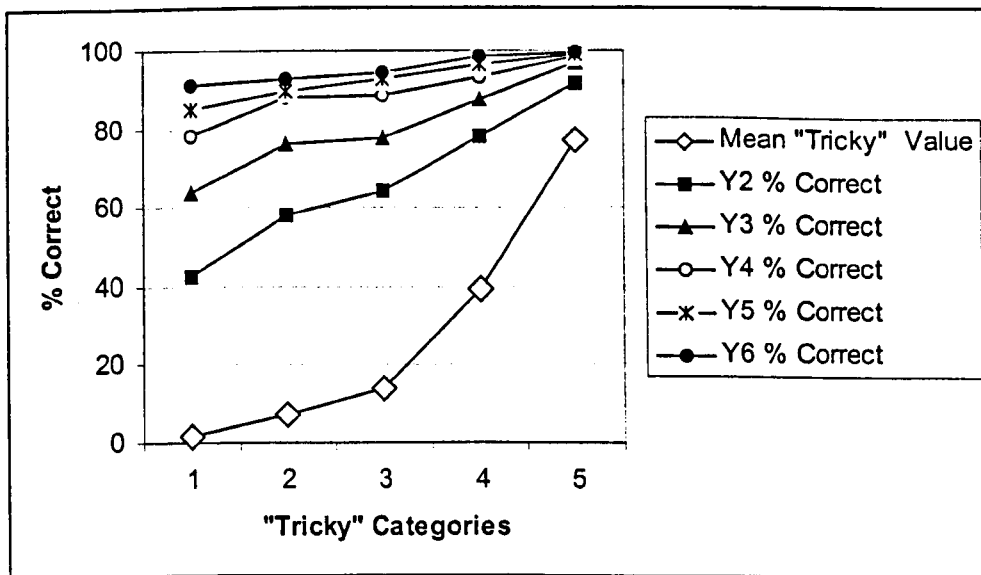
Figure 3.1 Percentage correct scores (years 2 to 6) for 5 categories of word difficulty ("Tricky" values)

Finally, the data from this study allow a reverse view of spelling performance and word characteristics, and the mistake of perceiving this as a tautological approach must not be made! The characteristics of the words are compiled before testing and
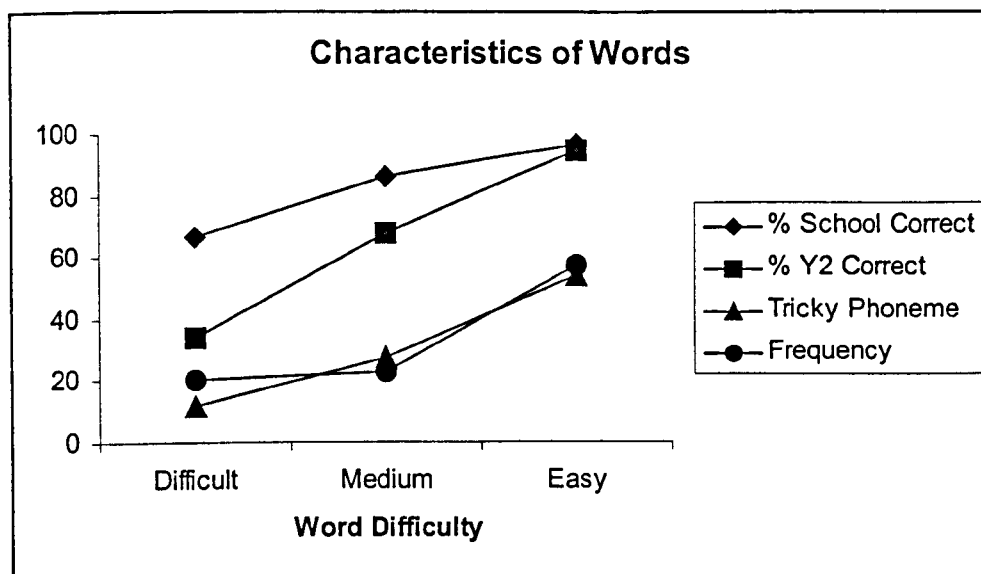
Figure 3.2 Characteristics of Words – Tricky Phoneme and Frequency (per 10,000)

are used as factors to predict spelling performance. The performance figures are then used to confirm the relationship with the various factors, indicating the variability within the behaviour that can be accounted for by the word factors. Knowing how many children can spell a word correctly also indicates how powerful the various factors are (a measure of which is obtained from the beta values in the regression equation). This can be expressed graphically, as in Figures 3.2 and 3.3. In Figure 3.2 the factor "word length", which in previous studies has simply been the number of letters in the word, has been further broken down into the number of phonemes in the word, and the difference between this value and the number of graphemes used to represent the word (this factor is called the "phonetic difference"). The word "cat" has 3 phonemes and 3 letters, hence the phonetic difference is 0; the word "through" also has 3 phonemes, but 7 letters, and the phonetic difference is 4. When these factors are used in the regression equation it is clear that the number of phonemes (the size of the word in terms of the number of sounds) does have an effect on spelling, but only a weak effect (beta value –0.2); whereas the phonetic difference factor has a much stronger effect (beta value –0.4). Similarly, the "tricky" factor is stronger (beta = 0.36) than the frequency factor (beta = 0.24).

The word groups in Figures 3.2 and 3.3 are based on how many children (years 2 to 6) can spell them. There are 3 groups of 50 words: the categories are the easiest 50 words, the most difficult 50 words, and an intermediate group of 50. The relationship between the value of the most difficult phoneme (Tricky) is made explicit in Figure 3.2: as performance falls, so does the value of the tricky phoneme; frequency also falls, but there is little difference between the medium and hard words. Figure 3.3 illustrates the relationship with the number of phonemes and the additional letters added to the words. The number of phonemes is a weak factor, and changes only slightly across the difficulty range. The number of additional letters, the phonetic
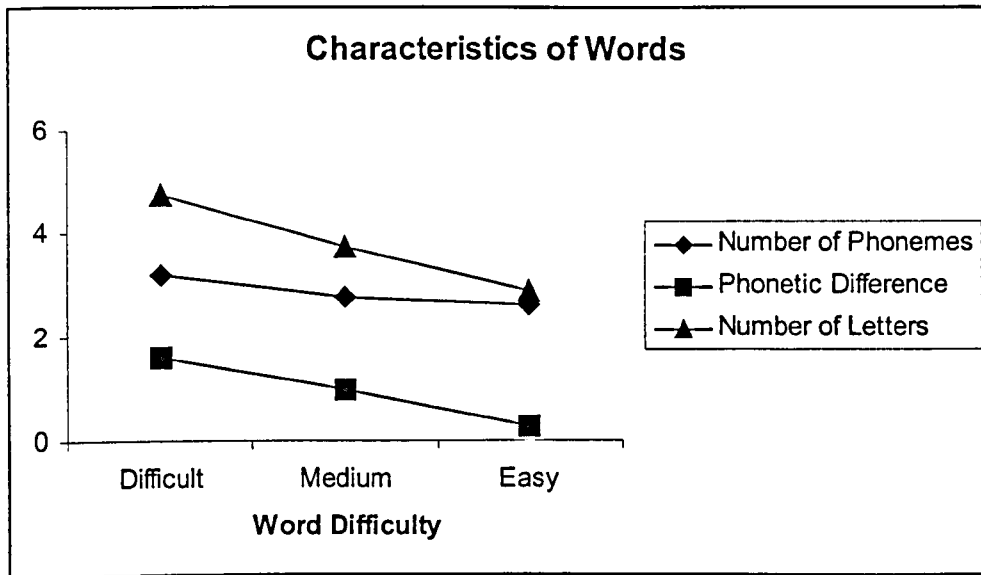
Figure 3.3 Characteristics of Words – Number of Phonemes, Number of Letters and Phonetic Difference.

difference, a much stronger factor, shows a definite relationship with word difficulty: easy words have almost one-to-one mapping; difficult words have, on average, two additional letters.

The question "Is English a Dyslexic Language?", put to the BDA First International Multilingualism and Dyslexia Conference, was greeted by many participants as confirming the general pattern in papers presented at the meeting: the English system of writing technology, because of its complexity, produces inflated levels of illiteracy. If this is the case, then a successful mastery-based literacy system must take this into account.

The most obvious method for dealing with the English literacy problem is reform of the system. This is unlikely to happen, but it would undoubtedly improve educational performance in the UK. Thorstad (1991) has demonstrated the beneficial effects of the initial teaching alphabet (i.t.a) in her comparison with Italian pupils.

Given that the orthography is unlikely to change, an understanding of the factors that make words difficult for English pupils will enable designers of mastery-based literacy programmes to prepare materials in sequences of increasing difficulty. Many reading programmes do this by gathering experimental data to indicate word difficulty, the current research enables predictions to be made about word difficulty without the need to gather data from pupils. There is confirmatory data available which suggests that the model is highly correlated with data obtained by testing. Spencer (1999b)demonstrates a high correlation (r=0.9) with Mitton's data. Recently, performance values, based on the regression equation for the above unpublished data, were calculated for the Schonnels' (1950) graded spelling test. The Pearson correlation between list position and the calculated PV was –0.88 (p<0.001), and between list group and PV was -0.98 (p<0.001).

This approach forms the basis for McGuinness et al's success: pupils who are failing in their attempts to acquire literacy skills can be taught what McGuinness refers to as the "basic code" in a relatively short space of time. Increasingly complex elements of the code are introduced as easier elements are mastered. This approach seems to offer the best way forward for a system that will minimise the powerful effects of a defective writing technology.

Overall Conclusion

The first series of papers explain why much research that has compared different media presentations has shown "no significant" difference. Implicit in these papers was a search to discover how improvements in pupil performance can be achieved. This is the theme of the second series of papers, which results in the successful application of mastery methods to literacy issues. The third series of papers evaluates the factors associated with English orthography that impinge on word difficulty, and indicates a rational strategy for a graded series of mastery-based literacy programmes.

Future research is clearly mapped out:

- Evaluation of the factors associated with spelling difficulty in the context of reading difficulty;
- Preparation of a compendium of English words indicating their probable reading and spelling difficulty (i.e. a difficulty index);
- Evaluation of the difficulty index for reading and spelling;
- Preparation of a mastery-based literacy programme for reading and spelling, based on the results of the above research.

4. CONCLUSION

Finally, I shall present an overview of the relevance of the research in Part 2 to current issues, and extend the arguments presented in the early work to the modern world, illustrating how the 'effectiveness of media forms in a learning context is affected by pervasive digitisation of those media forms in an electronic environment universally exploitable'.

The medium versus method debate
As technologies develop and new technologies emerge it is natural to anticipate that the new technologies may represent a fundamental shift in human endeavour, irrevocably changing our point of view. The telescope certainly had this effect in the hands of Galileo, and contributed to the shift from the orthodox geocentric to the Copernican heliocentric perspective on the solar system, although the Catholic church could not be swayed by the accumulating evidence. To an extent Kozma (1994) raised a challenge to what he saw as the new orthodoxy, rooted in Clark's (1983) article, that indicated that "media do not influence learning under any conditions" because media are "mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition."

Kozma felt that after 10 years it was time to revisit or reframe the question, arguing that the question was not so much **do** media influence learning, but **will** they? He suggested that there was no relationship between media and learning "because we have not yet made one" and claimed that there was a certain urgency because:

> We will be faced with a situation where telephone, cable television and digital computer technologies will merge... This capability presents the prospect of interactive video integrated with access to large multimedia data bases distributed among people in offices, classrooms, and living rooms all over the world. (Kozma, 1994, p 8)

The essence of Kozma's critique was that the research paradigm for much of the data that formed the basis of Clark's contention was constrained by the vestiges of the behavioural roots from which educational technology had emerged. It was a criticism of the concept of learning as the receptive response to instruction's delivery. Kozma wished to see learning as "an active, constructive, cognitive and social process by which the learner strategically manages available cognitive, physical and social resources to create new knowledge by interacting with information in the environment and integrating it with information already stored in memory." This was a powerful challenge to neo-behaviouristic approaches to learning from the emerging cognitive-constructivist camp that had been evolving during the 1980s (see Spencer, 1988, Chapters 6 & 7). By the mid-1990s cognitive theory had been absorbed, leading the major review of the period (Winn & Snyder, 1996, p134) to conclude that:

The evidence that educational technologists have accepted cognitive theory is prominent in the literature of our field (Gagne & Glaser, 1987; Richey, 1986; Spencer, 1988; Winn, 1989a).

This cognitive approach aims to provide a different perspective on instructional design, rooted in knowledge schemata actively constructed by learners. Kozma challenges the "no media effect" because he sees the new interactive and adaptive environments as being agents that enable students to think, and is in effect developing Salomon's contention that media are "tools for thought" (see Spencer, 1988, p 184-7).

Clark's (1994) reply to the new challenge was unrepentant:

> The claim of "no learning benefits" from media has been made and substantiated many times in the past. Many researchers have argued that media have differential economic benefits but no learning benefits. (Clark, 1994, p21)

Clark also points out that Kozma actually agrees that there is no compelling evidence "in the past 70 years of published and unpublished research" that media, rather than method, cause learning increases. He further claims that media research is a "triumph of enthusiasm over substantive examination of structural processes in learning and instruction." The principal issue here is Clark's commitment to the empirical research methodology, and Clark maintains his position that there simply is no empirical evidence to support media effects.

Updating the evidence

To what extent has this position changed in the intervening years? It could be claimed that, because a research indicator such as Russell's "No Significant Difference" web site (http://teleeducation.nb.ca/nosignificantdifference/) now links to a "Significant Difference" site (http://teleeducation.nb.ca/significantdifference/), new media approaches to learning are showing superior results. However, such results may be caused by statistical and/or methodological errors. This tendency, resulting in a small number of studies favouring the innovation, a similar number of studies favouring the traditional approach, but the vast majority showing no significant difference, has been noticeable since the early media research on film and television (Chu and Schramm, 1968). Russell (1999) indicates he "did not use any scientific sampling method but instead listed every study found that showed no significant difference" and that along the way found a comparatively small number of significant results:

> The point remains that such studies are practically nonexistent and the very few that do exist are offset by a like number which show negative results for technology-based instruction. The good news is that these no significant difference (NSD) studies provide substantial evidence that technology does not denigrate instruction. (p xiii)

Although statistical significance remains of interest to empirical researchers there is a move to consider *educational* significance as an indicator of the value of new media

or methods, often as a result of the meta-analysis of large numbers of individual studies. This is translated into an "effect size", which illustrates any change brought about by the new approach in terms of the distribution of scores (standard deviation) in the comparison group. This trend was acknowledged by McIssac and Gunawardena (1996, p420):

> Most of these media comparison studies found no significant differences (NSD) in learning... Critiquing these early media comparison studies, Spencer (1991) points out that they tended to report comparative statistics that gave no indication of the size of differences, if any, between the types of instruction. Conclusions tended to be based on the presence or absence of a statistically significant result. "When groups of research were reviewed there was a tendency to use a 'box score' tally approach, frequently resulting in a small number of studies favoring the innovation, a similar number of studies favoring the traditional approach, and the vast majority showing NSD" (p. 13)

The value of effect sizes over the 'box score' tally approach was noted in the Publication Manual of the American Psychological Association (1994) and there were even calls for mandatory reporting of effect sizes (Thompson, 1996).

Spencer (1991) attempted to deconstruct the available reported research effect sizes in order that a clearer understanding of the components would inform decision makers in the future. The valuable research of many decades could then have an influence on future attempts to either improve performance levels, broaden access or both.

What was apparent was that teaching methods, rather than media, could influence the level of performance of students, and that different media could be used without impairing the quality of learning. Clark (1987) had also identified this aspect, calling it the "confusion of technologies". Instructional technologies were engaged in the task of designing suitable methods and environments to enhance achievement, but a very different technology, delivery technology, provides access to those methods and environments.

Computer-based learning: an overview
When Kozma (1994) called for the debate on media to be re-visited the WWW had been in existence for 5 years but an easy-to-use interface for it (the Mosaic browser) had only recently become available. While the WWW was a relatively new medium for the distribution of digital information, the technology making it possible had been available for several decades, and there was an emerging consensus, from empirical research using meta-analyses, that computer-based technologies could enhance teaching and learning.

Kulik and Kulik (1987) had applied meta-analytic techniques to computer-based learning and found that the average effect size was 0.31, slightly larger than the 0.25 figure previously obtained for computer-based college teaching (Kulik et al, 1980)

and similar to the 0.32 obtained with secondary school students (Kulik et al, 1983). Their updated analysis (Kulik and Kulik, 1991), with a larger sample of 254 controlled evaluation studies, obtained an average effect size of 0.30. Although consistently statistically significant in outcome, an effect size of this magnitude is classed as small (Cohen, 1977), compared to the large effect sizes (0.8) obtained in mastery learning outcomes (see Spencer, 1991).

Khalili and Shashaani (1994) reported a slightly larger average effect size (0.38) for studies in their analysis of research on computer applications, which covered the years 1988-92. What is apparent from these results is that effects are dependent on the way in which the computer is used. Clark had earlier challenged Kulik on the underlying cause of the difference, leading Kulik (1985) to agree that it was the teaching method, rather than the computer as medium, that accounted for the learning gains.

When looking at the methodology embedded in the computer medium Khalili and Shashaani found that computer simulation had the largest effect (0.79) and drill and practice the smallest (0.11). A substantial number of studies used the computer for problem solving, and this method produced the second greatest effect size (0.41). They also confirmed Kulik and Kulik's (1991) finding that duration of treatment produced different achievement outcomes. Shorter duration (4-7 weeks) produced larger effects than semester or whole year studies. Kulik and Kulik (1991) suggested that as the treatment grows familiar, it loses its potency: the Hawthorne or novelty effect. In addition, there was evidence supporting a reduction in learning time, with CBI students requiring about two-thirds as much instructional time as did students who were taught conventionally (Kulik and Kulik, 1991). However, it must be borne in mind, once again, that this is probably not an exclusive effect of the medium, Vernon (1940) had observed a similar effect with films, which had cut instructional time to one-third of the conventional approach.

From the above it can be seen that the 'modes, media and methods' issue raised by Spencer (1991) parallels Clark's position that even with the newer technologies, such as computer-based learning, what is producing the beneficial achievement effects is not the medium per se, but the methodology implemented within the confines of the medium: drill and practice methods generally produce small (educationally insignificant) effects, whereas simulation methods, providing rich feedback mechanisms, produce large effect sizes (a shift in performance of almost 1 standard deviation). When problem solving methods are incorporated in the CBI model moderate effects are observed. None of these methods is the exclusive preserve of computers, although computers may be the most cost-effective medium of delivery of the instructional method.

Distribution technologies

Lockee et al (1999) when viewing the no significant difference results for distance learning warn that research into delivery media must avoid duplication of previous studies:

For example, as studies involving the WWW are designed, researchers must ask themselves what attributes of the system differentiate it from pre-existing delivery mechanisms such as computer-based instruction (CBI) or computer-mediated communication (CMC). CBI has all of the same message display abilities as the WWW (text, video, graphics), can offer practice and give immediate feedback, and can be engaged in at the student's pace. CMC allows for interaction with others, synchronous and asynchronous, much like Web-based chats or threaded discussions, and provides a forum for collaborative learning strategies not accessible in CBI.

The implication here is that if web-based instruction merely duplicates pre-existing systems then the same experimental results will inevitably occur. Again, as Russell (1999, see above) points out, for distance learning this can be seen in a positive light: if web-based courses can provide the same methodologies that make CBI successful, then they will also be successful in raising performance levels. Lockee et al are simply saying that such research is an unnecessary duplication.

The WWW was developed by Berners-Lee, in 1989, as an internet-based hypermedia initiative for global information sharing (Berners-Lee, 2001: http://www.w3.org/People/Berners-Lee/). It was essentially a publishing medium, allowing rapid dissemination of information in the form of text and graphics. More recently this has included multi-media displays (including real time audio and video) and a higher level of interactivity (both synchronous and asynchronous). In essence the WWW encompasses all modes, as defined by Rowntree (1982) and developed by Spencer (1991), except realia. The logical conclusion from this is that the WWW can incorporate all pre-existing media (except realia), and as such can be used to deliver pre-existing methods and will therefore, in terms of achievement, have the same effect as the pre-existing methods and media vehicles. As Clark (1994) pointed out, and this is a point of view subscribed to by Spencer (1999), ineffective methodologies will continue to be ineffective irrespective of the delivery medium. And, conversely, methodologies which have been shown to be more effective than traditional approaches, such as mastery learning, simulations and problem-solving, will continue to be more effective when delivered through a new medium such as the computer or the components which form the WWW, provided that the new medium has features which allow the method to be duplicated.

It has been suggested by educational decision-makers that using the internet as a distribution medium can result in economies of scale and reduce per capita costs as student numbers increase (Inglis, 1999; Whalen and Wright, 1999; Kearsley, 2000). This is a restatement of the claim often made for other media, such as television (see, for example, Greenhill, 1959), which may show savings, but only if there are sufficiently large numbers of students and courses.

Inglis (1999) considered online education from the perspective of its original purpose (the distribution of text and graphics) and calculated a number of scenarios (50-200 students) in order to look at cost differentials when moving from a print-based medium for distance education. The online delivery medium was found to be more

costly than print over the range of student intakes considered. The assumption behind this analysis was that the distribution medium would not affect level of performance, (a main justification for print-based distance education when compared with face-to-face/on-campus education), and that it would be appropriate to move to online distribution if this were less costly. However, because the economy of scale had already been achieved in the print-based medium (design, development, delivery) the additional costs associated with online access (e.g. ISP charges) actually made it less cost-effective.

Bates (1995) has addressed the issue of costs and benefits of online learning within his ACTIONS model, which has been applied in the Canadian Federal Government's NCE-Telelearning project through its Office of Learning Technologies, 1997-1999. Two studies have been reported (Bartolic-Zlomislic and Brett, 1999; Bartolic-Zlomislic and Bates, 1999), paying attention to three areas of cost (capital and recurrent; production and delivery; fixed and variable). The NCE-Telelearning project indicated that for a course using the internet to distribute materials, which were essentially text- and graphics-based using standard HTML, and as a means of providing tutor support via synchronous and asynchronous means, the breakeven point was 19 students per course over 5 years for one course (Bartolic-Zlomislic and Brett, 1999) and 44 students over 4 years for the other (Bartolic-Zlomislic and Bates, 1999). The largest cost was the tutor support and grading time within what was classed as a 'constructivist' environment (emphasising learning through active discussion). It was acknowledged that start-up costs were higher than had been anticipated. The study did however show that courses which were not viable in the immediate locality could become cost-effective when opened to a global market.

The preliminary report from the Korean National Open University (Jung and Leem, 2000) shows that development and delivery costs for online courses fall over time and that when compared with distance courses, using texts and television, higher completion rates for the online courses lowered the cost per completer. It is acknowledge that more rigorous, experimental studies are required before any firm conclusions can be drawn about the cost-effectiveness of online education.

Mastery methods and computer-based media
There can be no doubt that as effective methodologies are translated into new media for distribution they will retain their effectiveness. This is the claim made by Spencer (1996), when mastery methods were included in a computer-based system for teaching reading and spelling. The programme was developed initially to provide a 'talking book' facility, using the then latest computer technology, as a compensation programme for pupils who were not read to at home. However, pupils responded more positively to the minor testing phase of the programme rather than the automated reading phase, and this testing mode became the focus of future developments. As Block (1971) has pointed out, mastery learning approaches evolved early in the 20th Century, but only made a significant impact when aligned, in the 1950s, with Skinner's behaviouristic programmed learning methodology (which was implemented in a variety of media forms). However, there were severe technological restraints on the development of programmed learning in the 1950s and 1960s, especially the crucial issue of providing timely reinforcement, shown to critically affect outcome in early experimental studies (see Spencer, 1988, chapter 2). Although teachers can

provide sufficient levels of corrective feedback to pupils who are failing to learn to read, and thus significantly reduce reading deficits (as in Reading Recovery programmes), such levels were difficult to achieve within the early mechanical devices. The digital computer afforded more opportunities for not only the provision of corrective feedback, but the precision with which it was provided. For example, in the programme described by Spencer (1996), responses were logged and reinforced letter by letter, and alternative choices for responses were restricted in order that pupils were given ample opportunities to succeed and that tasks were achievable, an essential feature of successful reading programmes (McGuinnesss, 1998). The matching of task to level of performance is a crucial feature of teaching that can be simulated by a computer system that has a method of managing and using individual pupil data to make decisions concerning the most appropriate strategies for the pupil. Within mastery learning systems this is usually the preserve of human agents (proctors in the case of PSI) and was difficult to implement by machine before the advent of digital computers.

Teachers using the initial programme developed by Spencer were required to reset task parameters after the completion of each task by the pupil, but this was superseded by a computer-based management system which meant that the students completed the course without teacher intervention, although details of the pupils' performance were readily available within the management database. However, it must be recognised that, although the programme was very effective at remediation for the small group of subjects tested, it is, in essence, a drill and practice methodology, and will therefore have limits of applicability similar to those exemplified by recent evaluations of Integrated Learning Systems (ILS). ILSs also incorporate a management system which analyses student responses to exercises, records progress and adjusts the level of difficulty of work presented (National Council for Educational Technology, 1994).

Although, overall, drill and practice CBI implementations produce the smallest effects (see Khalili and Shashaani, 1994, above) there is evidence from Niemiec et al's (1987) synthesis of 48 studies in elementary schools that drill and practice methods produce the greatest effect size (0.47) in such schools, especially for younger and low achieving pupils. Drill and practise is a major methodology in the Computer Curriculum Corporation (CCC) system evaluated in the UK's Integrated Learning Systems trials.

In a meta-analysis of 22 evaluation studies of CCC reading and maths programs, Kulik (1994) reports that effects are especially large and consistent in well-designed programs such as the 'CCC-Stanford program' and gains of 1.4 years are likely with a year-long program, with gains of 2.0 years also quite possible and gains of less than 1.0 years highly unlikely. Kulik's analysis found an average effect size of .40 for the CCC studies, in which the external measures for each study were one of several commonly used standardized tests.

The results of the initial UK ILS trials (NCET, 1994) confirmed substantial maths gains, of 20 months progress within a six month period (equivalent to an effect size of 0.4), but with no discernable difference in reading results. In the second phase of research (NCET, 1996) pupils who remained on the system continued to improve in

maths (ES=0.8) and those who did not continue on the system maintained their advantage over the control group (ES=0.35), although after 18 months on the system there were signs that pupils were becoming demotivated. During this phase of the research effect sizes of 0.5 and 0.6 were also found for the literacy components; however, there was also a -0.4 effect size for the remaining school. The final phase which used standardised tests and evaluated learning outcomes in terms of impact on external examinations led the Durham team to concluded that:

> The fact that samples in both Years 9 and 11 ILS showed lower than the predicted levels of attainment may imply that whatever the non-ILS pupils were learning in the period up to Key Stage 3 and GCSE proved more beneficial than ILS experience on these measures of learning outcomes... What this finding implies is that whilst ILS pupils (if one assumes that these are a comparable group) were learning something of use to them on the system, this was less beneficial than what their peers were learning off-ILS. (Becta, 1998, p 19)

The discrepancy, particularly in maths, seems to relate to the type of test used and the overall conclusion was that the CCC system was good at supporting the learning of basic or 'core' knowledge, skills and procedures, but less effective when dealing with more demanding tasks requiring interpretation or translation. It seems reasonable to assume that this also applies to the system described by Spencer (1996): the efficacy of drill and practice teaching methods varies with the age of the student, successfully promoting mastery at elementary level, but inhibiting mastery if used inappropriately with older (secondary) pupils. Further research is clearly needed to study this interaction between age, teaching method and task difficulty for computer-based media.

Systematising reading acquisition
It was the realisation of the limitations of the computer system described in 1996 that led to the investigations into the structure of the English spelling system. The computer-based reading recovery programme demonstrated that children who had failed to learn how to read the most common word (the) in the English language in more than four years of schooling could learn to read the most common 90 words in the language in a reasonable length of time (14 hours). If an understanding of the structure of the written language could show that the language itself was having a significant effect on the acquisition of literacy skills, then it may be used to inform the development of a more scientific approach to teaching the reading and writing of English; and if the orthography itself was responsible for some measure of literacy failure then by acknowledging the particular difficulty of English not only the methodology of teaching, but also the time-frame for the acquisition of literacy skills, could be re-visited, leading to a more reasonable acquisition demand.

Share (1995) has suggested that reading acquisition is crucially dependent on phonological recoding acting as a self-teaching mechanism, enabling the learner to independently acquire an autonomous orthographic lexicon which is necessary for skilled word recognition and reading comprehension. Regular or shallow orthographies would therefore offer an advantage because 'successful decoding encounters with novel letter strings provide opportunities to learn word-specific print-

to-meaning connections' and this leads to the successful construction of a lexical store which enables fast whole word recognition, leading to text comprehension. Share elucidates the problem of direct rote instruction:

> Either by direct teaching of new words in the classroom, or through less formal assistance in other settings from parents or peers who supply the identity of visually unfamiliar words, a child may be able to acquire reading vocabulary by direct rote association... The problem with this approach is that it ignores the vast number of unfamiliar words continually being encountered in printed text... In the face of this orthographic avalanche, direct instruction is unlikely to offer a feasible acquisition strategy. (Share, 1995, p152-3)

In logographic written languages such as Chinese or the Kanji form of Japanese direct item-by-item instruction is the principal methodology, and its limitations are acknowledged, so that only a few hundred items per year are required to be learned (Mason et al, 1989). When this method is applied to English, McGuinness (1998) demonstrates that a memory ceiling effect is observed of about 1500 words and suggests that although the code for written English is complex, more complex than most alphabetic languages, this semi-alphabetic system still offers learners advantages over purely logographic systems, and these advantages must be built into any system of instruction for reading English (for such a methodology, Reading Reflex, see McGuinness and McGuinness, 1998). The method suggested by McGuinness is based on a similar analysis of the English language to that undertaken by Spencer (1998) and indicates that the basic code, which is formed from the most invariant features of the English orthography, must be taught to a high level of mastery before less predictable features are introduced.

Effects of orthographic depth: comparative data

Implicit in the Reading Reflex method is the assumption that the less predictable elements, those that move the orthography from shallow to deep, cause problems in the acquisition of phonological recoding skills resulting in reading and spelling deficits. A better understanding of the contribution that deep orthographies make to literacy failure will help to provide more efficient methods of remediation. Recent research shows that deep orthographies, such as English and French, make greater demands on brain processing powers and inhibit the development of reading skills in children.

Paulesu et al (2000), using positron emission tomography (PET) scans, demonstrated that orthography can actually have an effect on brain function.

> We present behavioural and anatomical evidence for a multi-component reading system in which different components are differentially weighted depending on culture-specific demands of orthography. Italian orthography is consistent, enabling reliable conversion of graphemes to phonemes to yield correct pronunciation of the word. English orthography is inconsistent, complicating mapping of letters to word sounds.

> In behavioural studies, Italian students showed faster word and non-word reading than English students. In two PET studies, Italians showed greater activation in left superior temporal regions associated with phoneme processing. In contrast, English readers showed greater activations, particularly for non-words, in left posterior inferior temporal gyrus and anterior frontal gyrus, areas associated with word retrieval during both reading and naming tasks. (Paulesu et al, 2000, p91)

Such a finding has profound implications for the emergence of reading deficits in different cultures, particularly the manifestation of dyslexia. If deep orthographies make demands on brain functioning that are not found with shallow orthographies, it follows that for individuals with impairments in brain functioning the transparency of the orthography could affect the manifestation of reading deficits. Paulesu et al (2001) followed up their work with normal subjects to investigate the cultural diversity and biological unity of dyslexia.

> The recognition of dyslexia as a neurodevelopmental disorder has been hampered by the belief that it is not a specific diagnostic entity because it has variable and culture-specific manifestations. In line with this belief, we found that Italian dyslexics, using a shallow orthography which facilitates reading, performed better on reading tasks than did English and French dyslexics. However, all dyslexics were equally impaired relative to their controls on reading and phonological tasks... We conclude that there is a universal neurocognitive basis for dyslexia and that differences in reading performance among dyslexics of different countries are due to different orthographies. (Paulesu et al, 2001, p 2165)

For some time it has been apparent that orthography contributes to reading difficulties for children (see Spencer, 2000, for a review). There is also accumulating evidence that orthographies make differing demands on brain functioning. Spencer (2001, in press[a]) has synthesised these results into a model that indicates the demands that common English words will make on brain functioning, recognising that they fall along a transparency continuum with some words demanding fewer brain resources than others (see Figure 4.1). For spelling, the extra demands of common but irregular words has been estimated (Spencer, 2002, in press[b]) to be equivalent to 3 additional years of schooling, and this is similar to Seymour's (2001) claim that

> ... a learning period of 2 ½ to 3 years is needed to match the competence which is achieved in less than one year in most languages... The conclusion is that English-speaking children acquire foundation literacy much more slowly than other language groups.

The study involved 15 European languages of varying syllable structure and orthographic depth. Danish and English, both showing complex syllable structure and
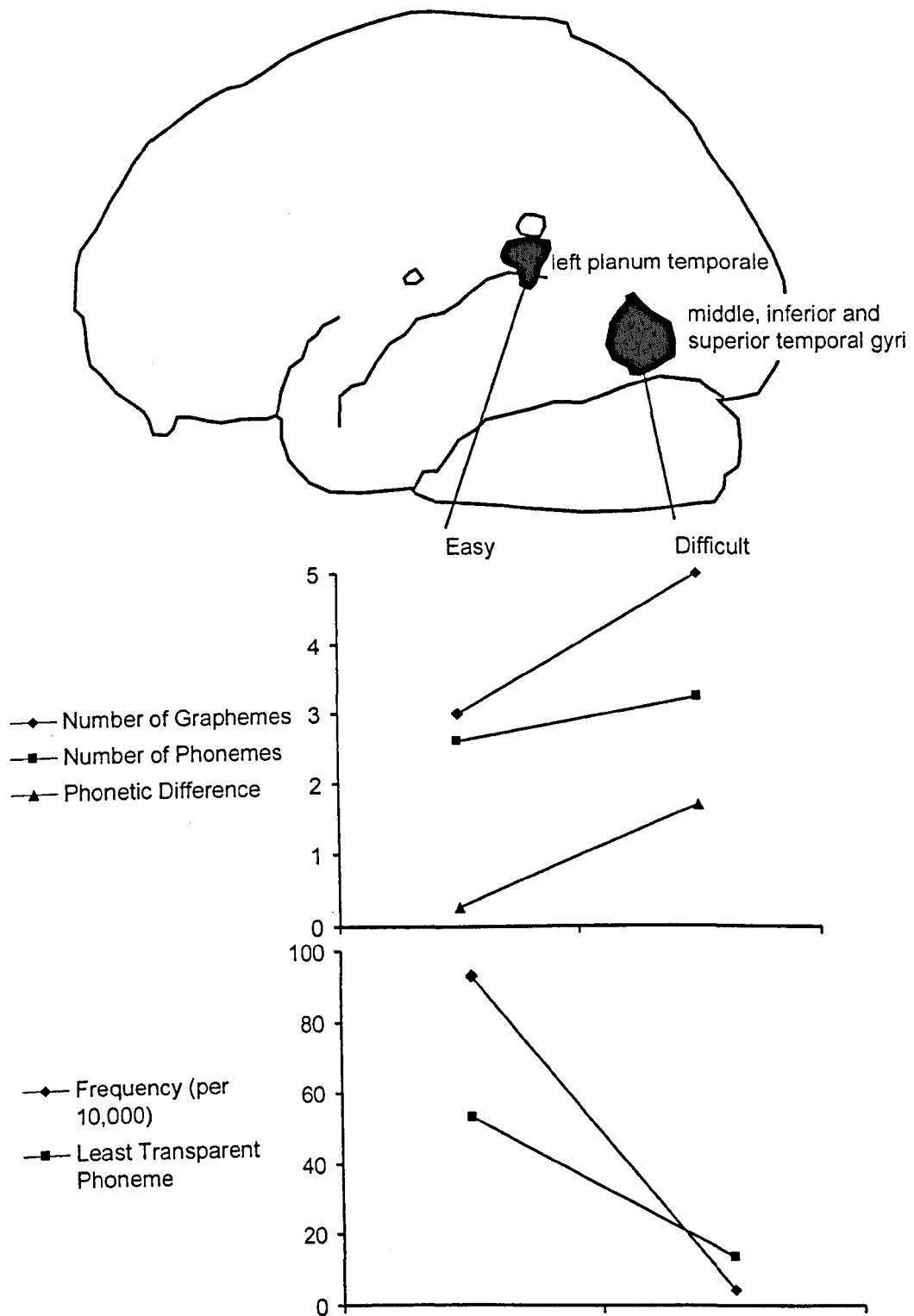
Figure 4.1 Easy and difficult words mapped to reading centres (Spencer, 2001, in press[a])

deep orthography, produced greater error rates for both words and non-words than other languages. French children also performed at a lower level than average, but not as low as the English and Danish children. Seymour and Duncan are currently (2001-2003, see http://www.dundee.ac.uk/psychology/phkseymour/) investigating 'foundation literacy acquisition in European orthographies' and the 'manifestation of dyslexia in European orthographies of varying depth and complexity', which should lead to a more complete picture of the effect that orthography has on literacy. The initial results, however, indicate that:

> The teaching methods could be important. But the teaching method for a deep orthography such as English probably needs to be different from the method for a shallow orthography such as Finnish. So, although it might be possible to accelerate progress somewhat in the UK (by appropriate emphasis on letter knowledge and decoding) it is unlikely that this could eliminate the differences... We would therefore like to retain the suggestion that learning to read in English is simply a slower process than learning in other languages. We suggest that syllable complexity affects decoding and that orthographic depth affects both word learning and decoding. It is possible that the psychological mechanism developed to encompass the task of learning to read differs between shallow and deep orthographies. (Seymour, 2001)

Although orthography is implicated in the acquisition of foundation level literacy, deeper languages requiring different instructional methodologies and longer acquisition times than the shallow languages, it does not automatically follow that shallow languages will always manifest high levels of literacy as measured, for example, by the instruments developed by the International Adult Literacy Survey (IALS). A transparent orthography may lead to rapid acquisition of basic literacy skills, but cultural or political elements may inhibit the achievement of full literacy. However, care must be taken when reviewing the evidence on national levels of literacy, not to be unduly influenced by personal predilections. It may appear that the sun moves around the earth, but appearances can be deceptive! For example, it could be argued that English children fail to learn to read because they watch too much television or do not spend sufficient time talking to their parents, unlike for example the French, who are seen by some as epitomising the very essence of a literate society. Even if French has a deep orthography, it could be argued that because French culture is inextricably linked to literacy this would ameliorate any inherent deficits. It is worth, briefly, considering the available data on this interesting question. The results of the first International Adult Literacy Survey, conducted in 1994, were published in 1995 under the title "Literacy, Economy and Society" published jointly by the Organisation for Economic Co-operation and Development and Statistics Canada.

> In 1994 nine countries took part in the first ever IALS - Canada, the US, Ireland, Germany, France, Sweden, the Netherlands, Poland and Switzerland. These countries worked collaboratively to develop an assessment instrument that

would be used in different languages and cultures taking as
the starting point the conceptual and methodological
developments in measuring literacy in the US in the 1980s.
The results from seven of the countries that took part in IALS
in 1994 were published by the OECD in 1995. (Carey, 2000,
p2)

Nine countries took part, but only the results of seven were published, although all
countries had worked co-operatively on the project. The first report is dedicated to the
memory of Brendan Hickey, the National Study Manager for Ireland, whose death in
1995 held up the analysis of the Irish data and prevented its inclusion in the final
report. Ireland's results were published in the second report (OECD, 1997). However,
France withdrew its data from the initial report and has not rejoined the group of
countries reporting IALS data.

The French withdrawal, which came despite French
participation in the design, development and testing of the
study, was motivated by concerns about the comparability of
the IALS results about to be published, especially that they
tended to underestimate the true literacy skills of the adult
population of France relative to the populations of other
countries participating in the survey. (OECD, 2000, p123)

The French withdrawal from the reporting process has made it rather difficult to test
the hypothesis that French cultural elements could compensate for the inherent
deficits associated with a deep orthography. The French had raised a number of
methodological criticisms of the conduct of the survey when the results were known
and the outcome had conflicted with the French expectations. It was impossible to
ignore the criticisms of one of the founder members of the IALS group, and
eventually the European Commission funded research to investigate the validity of the
objections (Carey, 2000). The results of this study confirmed the poor performance of
France that was apparent from the initial study (Figure 4.2). On the prose scale the
distribution of French performance is almost the opposite of the Swedish distribution,
with the largest group performing at the lowest literacy level. This pattern is repeated
for the document and quantitative literacy scales. It could be argued from this that the
results are predictable from the data which are emerging from comparisons of
languages with differing orthographic transparencies: deeper languages inhibit the
development of the basic foundation level of literacy and this has a cumulative effect
resulting in poor adult performance on more complex tasks. However, this appears not
to be the case because the original IALS report included Switzerland as two separate
groups, German and French speaking. The French Swiss results (see Figure 4.3)
show that deep languages do not automatically translate into adult deficits, and this is
encouraging because it demonstrates that societies can compensate for inadequacies
in their written languages. Sweden, with a complex syllable structure and a
moderately deep orthography, also achieves this. The reasons for the poor French
results clearly require further exploration, but it does appear that it is not a
methodological artefact and it may be related to a combination of deep orthography
and inappropriate teaching methodologies, together with other sociological factors.

The French results also illustrate a basic tenet of this thesis: in research our *a priori* expectations may not be met by the data we collect, and this provides science with a great opportunity. The challenge, especially in academic life, is to understand why these expectations are not met.

	France	UK	Sweden
Level 1	41	22	8
Level 2	34	30	20
Level 3	22	31	40
Level 4/5	3	17	32
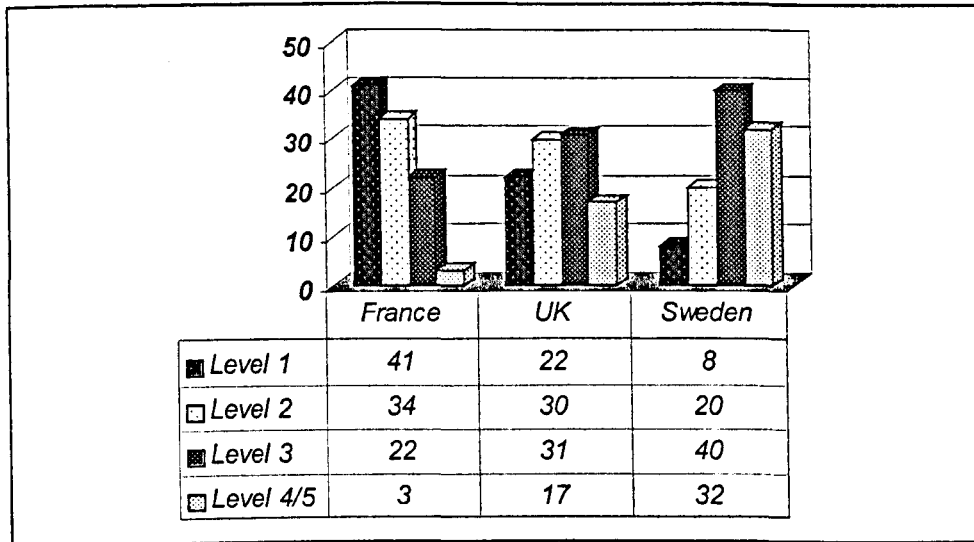
Figure 4.2 Prose literacy by country (France, UK and Sweden), based on Carey, 2000.

	France	Swiss (French)	UK	Sweden
Level 1	41	18	22	8
Level 2	34	34	30	20
Level 3	22	39	31	40
Level 4/5	3	10	17	32
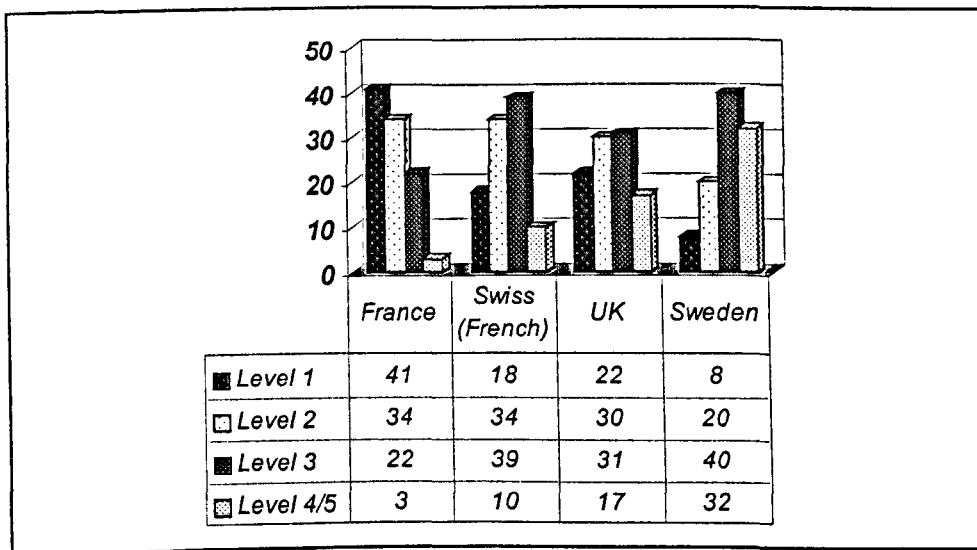
Figure 4.3 Prose literacy by country (France, Switzerland [French speaking], UK and Sweden), based on Carey, 2000 and IALS, 1995.

Conclusion

Clark (1983) compared media to trucks delivering food and this was later challenged by Kozma (1994) who called for the acknowledgement that new media may offer 'tools for thought'. Clark (1994) responded by adopting a medical analogy and concluded that:

> ...aside from the identification of necessary methods for learners and tasks, it is important to derive media that are capable of delivering the method at the least expensive rate and in the speediest fashion. Media influence cost or speed (efficiency) of learning but methods are causal in learning. (Clark, 1994, p 26)

If this is the case, then teaching and learning programmes that change the way children think, by providing problem-solving environments and simulations, may move from the real world (such as the 'cognitive acceleration' programmes developed by Adey and Shayer, 1994) to the digital world (the ThinkerTools environment described by Kozma, 1994), and in doing so may actually decrease the time that is required for learning or schemata development. Here we may begin to observe a new phenomenon which will form the basis for future method-medium comparisons: learning acceleration, stemming from the synergism of medium and method. Such digital environments when combined with the speed and reliability of new distribution media may critically affect the diffusion of knowledge, skills and expertise throughout society in much the same way that the printing press was the vehicle for the revolution in Western thought during the Renaissance. For example, if deep orthographies impede the progress of pupils when learning to read, and the method of remediation consists of 'daily half-hour lessons taught by a teacher trained to diagnose and support children's problem-solving approach to reading,' (Hobsbaum et al, 1996), an approach that is consistent with the constructivist principles of Vygotsky's theory on the acquisition of cultural tools, then a medium such as the computer which can duplicate the performance of a skilled teacher may provide an effective approach to accelerating the acquisition of literacy skills. This development of the computer as tutor has not, as yet, achieved maturity, but there are promising developments in this field, as suggested in the following extract (Spencer, 1999):

> We must remember that the development of sophisticated computer-based systems is in its infancy, but even so... it is now beginning to equal the effectiveness of individual tutoring methods. It is not surprising that the computer begins to excel when it is used to provide simulations. It is inevitable that we shall see the computer competently simulating human one-to-one exchanges, passing what is sometimes known as the Turing test.
>
> The recent chess victory of the Deep Blue computer over world champion Kasparov has fulfilled the expectations that Shannon and Weaver (1949) expressed in their theory of information, published just 50 years ago. They tentatively suggested that information theory could lead to intelligent systems tackling the supreme game of chess. Their work

stemmed out of the advances made in communication systems during the Second World War, and the impact of new computer methods for breaking codes. It was also related to the concept of feedback in machines, based on work with rockets and missiles, proposed by the father of cybernetics, Norbert Weiner.

The greatest impact of Deep Blue's achievement is likely to be in the field of education. The key to education is the assimilation of information to create new mental schemes, which enable us to look at the world anew, to go beyond the obvious. And how do we do this, how does this mental construction come about? By entering into a dialogue with the world, the physical world and the social world, as suggested by Piaget (1971) and Vygotsky (1962). One way in which we extend the capabilities of the child is to have her enter into a dialogue with an expert who can answer questions, provide hints, set expectations. The expert has traditionally been the teacher, but there has been a search for mechanical means to do the same job, just as there were searches for mechanical devices to play chess...

But, much more is to come, and it will come soon. Already it has been demonstrated that artificial intelligence (AI) systems can teach, or rather tutor, as effectively as human tutors in advanced courses (Lajoie and Derry, 1993). Intelligent tutor systems are now being used in college level maths courses, helping students gain understanding of the complexities of geometry proofs. These computer-based tutors, using diagnostic modelling procedures, have been found to be as capable as their human counterparts at identifying and correcting student misunderstandings, even in the complex, advanced field of avionics.

Artificial intelligence systems actually learn from experience and develop powerful rules and strategies, which when applied may be even more effective than those used by the human experts whose skills have been tapped by the computer. Such tutoring systems will get to know their pupils, will have extracted rules and strategies for optimal methods of teaching, and pupils will really have the very best of personal tutors, a dream that has been sought for many years. Deep Blue's victory shows that the potential for sophisticated dialogues with an understanding, motivating, superior intelligence are achievable. There can be no doubt that just as the filing cabinet size of the first teaching computers was reduced, within a few years, to that of the watch on my wrist, so computers more powerful than Deep Blue will shrink to sit on a child's desk, early in the next millennium.

And teachers, what will become of them? Teachers will always be needed, because of the human touch; their role will undoubtedly change, it may even become more rewarding.

Finally, I must admit that I was wrong in my choice of title for this paper. Teachers do use educational technology and they always have: the written word, on a blackboard or in a book, represents technology which is so embedded in teaching that we hardly acknowledge its presence. The new technologies, such as artificial intelligence tutors, are just emerging. Soon they, too, will become ubiquitous, and so totally embedded within the educational context that they will become transparent, in much the same way that written communication is hardly noticed as an embedded technology today. (Spencer, 1999c, p9-10)

5 REFERENCES

Ackers, G. W and Oosthoek, J. F. (1972) 'The evaluation of an audio-tape mediated course-I', *British Journal of Educational Technology*, 3, 136-147.

Adey, P. and Shayer, M. (1994) Really Raising Standards. London: Routledge.

Bartolic-Zlomislic, S. and Bates, A. W. (1999) Assessing the Costs and Benefits of Telelearning: Case Study from the University of British Columbia. (http://research.cstudies.ubc.ca/nce/EDST565.pdf)

Bartolic-Zlomislic, S. and Brett, C. (1999) Assessing the Costs and Benefits of Telelearning: Case Study from the Ontario Institute for Studies in Education. (http://research.cstudies.ubc.ca/nce/OISEcbreport.pdf)

Bates, A. (1995) Technology, Open Learning and Distance Education. London: Routledge.

Block, J. H. (1971) Mastery Learning. London: Holt Rinehart & Winston.

Bloom, B. S. (1968/70) 'Learning for mastery', (UCLA-CSEIP) *The Evaluation Comment*, 1(2), 1968. Reprinted in Block, J. H. (Ed) Mastery Learning. New York: Holt, Rinehart and Winston, 47-63.

British Educational Communications and Technology Agency (1998) The UK ILS Evaluations. Coventry: Becta.

Brown, H. E. (1928) 'Motion picture or film-slide', *Science and Maths*, 28, 517-526.

Campeau, P. L. (1966) Selective review of the results of research on the use of audio visual media to teach adults. In Briggs, L. J., Campeau, P. L., Gagné, R. M and May, M. A. (1966) Instructional Media: a procedure for the design of multi media instruction and suggestions for future research. Monograph 2. American Institutes for Research: Pittsburgh.

Carey, S. (2000) Measuring Adult Literacy: the International Adult Literacy Survey (IALS) in the European Context. London: Office for National Statistics.

Carney, E. (1994) A Survey of English Spelling. London: Routledge.

Carpenter, C. R. (1953) 'A theoretical orientation for instructional film research', *Audio Visual Communication Review*, 1, 38-52.

Carpenter, H.A. (1937) 'Science in the Rochester school of the air', *Science Education*, 21, 77-81.

Chu, G. C. and Schramm, W. (1968) Learning from Television: what the research says. Institute for Communication Research: Stanford University.

Clark, C. C. (1932) Sound Motion Pictures as an Aid in Classroom Teaching. Doctoral Thesis: New York University.

Clark, R. E. (1983) 'Reconsidering research on learning from media', *Review of Educational Research*, 53(4), 445-459.

Clark, R. E. (1987) Which technology for what purpose? Paper presented at an invited symposium for the annual meeting of the Association for Educational Communication and Technology, Atlanta, GA.

Clark, R. E. (1994) 'Media will never influence learning', *Educational Technology Research and Development*, 42(2), 21-29.

Cohen, J. (1977) Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press.

Cox, S. (1976) 'Colour in learning from film and tv: a survey of the research with some indications for future investigations', *Educational Broadcasting International*, June, 85-96.

Davies, A. and Ritchie, D. (1996) THRASS: Teacher's Manual. London: Harper Collins.

Dwyer, F. M. (1967) 'Adapting visual illustrations for effective learning', *Harvard Educational Review*, 37, 259-263.

Forsythe, R. D. (1970) Instructional Radio: a position paper. Stanford University. ERIC ED 044 933

Freeman, F. N. (1924) Visual Education. Chicago: University of Chicago Press.

Frith, U. (1985) Beneath the surface of developmental dyslexia. In K.E. Patterson, J.C. Marshall & M. Coleath (Eds.), Surface Dyslexia (301-330). London: Erlbaum.

Frith, U., Wimmer, H. & Landerl, K. (1998) 'Differences in phonological recoding in German- and English-speaking children', *Scientific Studies of Reading*, 2(1), 31-54.

Greenhill, L. P. (1959) Closed Circuit Television for Teaching in Colleges and Universities. Pennsylvania: Pennsylvania University (mimeographed).

Hall, W. E. and Cushing, J. R. (1947) 'The relative value of three methods of presenting learning material', *Journal of Psychology*, 24, 57-64.

Heron, W. T. and Zeibarth, E. W. (1946) 'A preliminary experimental comparison of radio and classroom lectures', *Speech Monographs*, 13, 54-57.

Hoban, C. F. and Van Ormer, E. B. (1950) Instructional Film Research 1918-1950. Technical Report No. SDC-269-7-19 Port Washington, New York: Special Devices.

Hobsbaum, A., Peters, S. and Sylva, K. (1996) 'Scaffolding in Reading Recovery', *Oxford Review of Education*, 22(1), 17-35.

Hofland, K. and Johansson, S. (1982) Word frequencies in British and American English. Bergen: The Norwegian Computing Centre for the Humanities.

Inglis, A. (1999) 'Is online delivery less costly than print and is it meaningful to ask?', *Distance Education*, 20(2), 220-239.

Jung, I. And Leem, J.H. (2000) 'Comparing cost-effectiveness of Web-based instruction and televised distance education'. Manuscript under review, quoted in Jung, I. and Rha, I. (2000) 'Effectiveness and cost-effectiveness of on-line education: a review of the literature', *Educational Technology*, 40(4), 57-60.

Jung, I. and Rha, I. (2000) 'Effectiveness and cost-effectiveness of on-line education: a review of the literature', *Educational Technology*, 40(4), 57-60.

Kanner, J. H. and Rosenstein, A. J. (1960) 'Television in army training: colour versus black and white', *Audio Visual Communication Review*, 8, 243-252.

Kearsley, G. (2000) Online Education: learning and teaching in cyberspace. Belmont, Ca: Wadsworth.

Keller, F. S. (1968) 'Goodbye, Teacher', *Journal of Applied Behaviour Analysis*, 1, 79-87.

Khalali, A. and Shashaani, L. (1994) 'The effectiveness of computer applications: a meta-analysis', *Journal of Research on Computing in Education*, 27(1), 48-61.

Kozma, R. B. (1994) 'Will media influence learning? Reframing the debate', *Educational Technology Research and Development*, 42(2), 7-19.

Kulik, C. C., Kulik, J. A. and Bangert-Drowns, R. L. (1986) Effects of Testing for Mastery on Student Learning. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1986.

Kulik, J. A. (1985) 'The importance of outcome studies: a reply to Clark', *Educational Communications and Technology Journal*, 34(1), 381-386.

Kulik, J. A. (1994) Meta-analytic Studies of Findings on Computer-based Instruction. In E. L. Baker & H. F. O'Neil (Eds) Technology Assessments in Education and Training (9-33). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kulik, J. A. and Kulik, C. -L. C. (1987) 'Review of recent literature on computer-based instruction', *Contemporary Educational Psychology*, 12, 222-230.

Kulik, J. A. and Kulik, C. -L. C. (1991) 'Effectiveness of computer-based instruction: an updated analysis', *Computers in Human Behavior*, 7, 75-94.

Kulik, J. A., Kulik, C. -L. C. and Cohen, P. A. (1980) 'Effectiveness of Computer-based college teaching: a meta-analysis of findings', *Review of Educational Research*, 50(4), 525-545.

Kulik, J. A., Bangert, R. L. and Williams, G. W. (1983) 'Effects of computer based teaching on secondary school students: a meta-analysis', *Journal of Educational psychology*, 75, 19-26.

Lajoie, S. P. and Derry, S. J. (1993) Computers as Cognitive Tools. Lawrence Erlbaum Associates: Hillsdale, New Jersey.

Landerl, K., Wimmer, F. & Frith, U. (1997) 'The impact of orthographic consistency on dyslexia: A German-English comparison', *Cognition*, 63, 315-334.

Levie, W. H. and Lentz, R. (1982) 'Effects of text illustrations: a review of research', *Educational Communication and Technology Journal*, 30, 195-232.

Lockee, B. B., Burton, J. K. and Lawrence, H. C. (1999) 'No comparison: distance education finds a new use for 'no significant difference'', *Educational Technology Research and Development*, 47(3), 33-42.

Lumsdaine, A. A. and May, M. A. (1965) 'Mass communication and educational media', *Annual Review of Psychology*, 16, 475-534

Mason, J. M., Anderson, R. C., Omura, A., Uchida, N., and Imai, M. (1989) 'Learning to read in Japan', *Journal of Curriculum Studies*, 21, 389-407.

Mathur, J. C. and Neurath, P. (1959) An Indian Experiment in Farm Radio Forums. UNESCO: Paris.

May, M. A. and Lumsdaine, A. A. (1958) Learning from Films. Yale University Press: New Haven.

McClusky, F. D. (1924) Comparison of different methods of visual instruction. In F. N. Freeman (Ed) Visual Education. Chicago: University of Chicago Press, 83-166.

McGuinness, C. and McGuinness, G. (1998) Reading Reflex. London: Penguin.

McGuinness, C., McGuinness, D. and McGuinness, G. (1996) 'Phono-Graphix: A new method for remediation reading difficulties', *Annals of Dyslexia*, 46, 73-96.

McGuinness, D. (1998) Why Children Can't Read: And What We Can Do About It. London: Penguin.

McIssac, M. S. and Gunawardena C. N. (1996) Distance Education. In D. H. Jonassen (Ed) Handbook of Research for Educational Communications and Technology. New York: Simon & Schuster Macmillan.

Menne, J. W., Hannum, T. E., Klingensmith, J. E. and Nord, D. (1969) 'Use of taped lectures to replace class attendance', *Audio Visual Communication Review*, 17, 42-46.

National Council for Educational Technology (1994) Integrated Learning Systems: a report of the pilot evaluation of ILS in the UK. Coventry: NCET.

National Council for Educational Technology (1996) Integrated Learning Systems: a report of Phase II of the pilot evaluation of ILS in the UK. Coventry: NCET.

Niemeic, R., Samson, G., Weinstein, T., and Walberg, J. (1987) 'The effects of computer based instruction in elementary schools: a qualitative synthesis', *Journal of Research on Computing in Education*, 20(2), 85-103.

Oney, B. and Goldman, S.R. (1984) 'Decoding and comprehension skills in Turkish and English: effects of regularity of grapheme-phoneme correspondences', *Journal of Educational Psychology*, 76(4), 557-568.

Organisation for Economic Co-operation and Development and Human Resources Development Canada (1997) Literacy Skills for the Knowledge Society: further results from the International Adult Literacy Survey. Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development and Human Resources Development Canada (2000) Literacy in the Information Age: Final Report of the International Adult Literacy Survey. Paris: Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development and Statistics Canada (1995) Literacy, Economy and Society. Paris: Organisation for Economic Co-operation and Development.

Paivio, A. (1971) Imagery and Verbal Processes. New York: Holt, Rinehart and Winston.

Paulesu, E., Demonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S.F., Cossu, G., Habib, M., Frith, C.D., & Frith, U. (2001) 'Dyslexia: Cultural Diversity and Biological Unity', *Science*, 291, 2165-2167.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C.D., & Frith, U. (2000) 'A cultural effect on brain function', *Nature Neuroscience*, 3(1), 91-96.

Popham, J. W. (1961) 'Tape recorded lectures in the college classroom', Audio Visual Communication Review, 9(2),109-118.

Postlethwait, S. N., Novak, J. and Murray, H. T. (1972) The Audio-Tutorial Approach To Learning. Minneapolis: Burgess Publishing Co.

Publication Manual of the American Psychological Association, 4th Edition (1994) Washington, DC: American Psychological Association.

Reid, C. J. and MacLennan, D. W. (1967) Research in Instructional TV and Film. US Government Printing Office: Bureau of Research: Washington.

Reid, D. (1989) Word for word. London: LDA.

Rowntree,D. (1982) Educational Technology in Curriculum Development (2nd Edition). London: Harper and Row.

Russell, T. L. (1999) No Significant Difference Phenomenon. North Carolina State University, Raleigh, NC.

Saettler, P. (1990) The Evolution of American Educational Technology. Englewood: Libraries Unlimited.

Schonnel, F. J. and Schonnel, F. E. (1950) Diagnostic and Attainment Testing. Edinburgh: Oliver and Boyd.

School Curriculum & Assessment Authority (1997) Standards at Key Stage 1 English & Mathematics. Report on the 1996 National Curriculum Assessments for 7-year-olds. London: QCA.

School Curriculum & Assessment Authority (1997) Standards at Key Stage 2 English, Mathematics & Science. Report on the 1996 National Curriculum Assessments for 11-year-olds. London: QCA.

Seymour, P. H. K. (2001) How do children learn to read? Is English more difficult than other languages? Paper presented to the BA Festival of Science, Glasgow (personal communication).

Shannon, C. E. and Weaver,W. (1949) The Mathematical Theory of Communication. Urbana: University of Illinois Press.

Share, D. L. (1995) 'Phonological recoding and self-teaching: sine qua non of reading acquisition', *Cognition*, 55, 151-218.

Singer, C., Holmyard, E. J. & Hall, A. R. (1954) A History of Technology, Volume 1, Part 1, Chapter 4: Speech and Language. Oxford: Oxford University Press.

Skinner, B. F. (1954) 'The science of learning and the art of teaching', *Harvard Educational Review*, 24(2), 86-97.

Spencer, K. A. (1977) 'An evaluation of the cost-effectiveness of audiotaped and videotaped self-study media presentations', *British Journal of Educational Technology*, 8(1), 71-77.

Spencer, K. A. (1981a) 'An evaluation of the cost-effectiveness of film and tape-slide instructional presentations for the mentally handicapped', *Programmed Learning and Educational Technology*, 18(1), 23-9.

Spencer, K. A. (1981b) Alternative media: a review of the instructional effectiveness, *Educational Broadcasting International*, 14(2), 90-93.

Spencer, K. A. (1981c) 'AV communication: small is beautiful', *Audiovisual Librarian*, 7(1), 22-28.

Spencer, K. A. (1988) The Psychology of Educational Technology and Instructional Media, London: Routledge.

Spencer, K. A. (1990a) 'HyperCard: teaching technology for successful learning', *Journal of Audiovisual Media in Medicine*, 13, 25-30.

Spencer, K. A. (1990b) 'HyperHeart: does animated illustration contribute to mastery learning?', *British Journal of Educational Technology*, 21(3), 227-8.

Spencer, K. A. (1991) 'Modes, media and methods: the search for educational effectiveness', *British Journal of Educational Technology*, 22(1), 12-22.

Spencer, K. A. (1992) Information technology and human information processing: a critical perspective, Proceedings of the European Conference about Information Technology in Education, Barcelona, November, 512-522.

Spencer, K. A. (1996) 'Recovering reading using computer mastery programmes', *British Journal of Educational Technology*, 27(3), 191-203.

Spencer, K. A. (1998) 'Predictive models of spelling behaviour in 7- and 11-year-olds', *Journal of the Simplified Spelling Society*, 24(2), 23-27.

Spencer, K. A. (1999a) 'Predicting spelling behaviour in 7 to 11 year-olds', *Reading*, July, 72-77.

Spencer, K. A. (1999b) 'Predicting word-spelling difficulty in 7- to 11-year-olds', *Journal of Research in Reading*, 22(3) 283-292.

Spencer, K. A. (1999c) Educational Technology - An Unstoppable Force: a selective review of research into the effectiveness of educational media, *Educational Technology and Society*, 2(4), 23-34.

Spencer, K. A. (2000) 'Is English a dyslexic language? Orthographic transparency and multilingual teaching', *Dyslexia*, 6, 152-162.

Spencer, K. A. (2001, in press[a]) 'Differential effects of orthographic transparency on dyslexia', *Dyslexia*, 7(4).

Spencer, K. A. (2002, in press[b]) 'English spelling and its contribution to literacy problems: word difficulty for common English words', *Reading*.

Spencer, K. A. and Clarke, A. (1981) 'A BBC television series for the mentally handicapped', *Educational Broadcasting International*, 14(3), 110-113.

Stickell, D. W. (1963) A Critical Review of the Methodology and Results of Research Comparing Televised and Face-to-Face Instruction. Doctoral dissertation: Pennsylvania State University.

Thompson, B. (1996) 'AERA editorial policies regarding statistical significance testing: Three suggested reforms', *Educational Researcher*, 25(2), 26-30.

Thorstad, G. (1991) 'The effect of orthography on the acquisition of literacy skill', *British Journal of Psychology*, 82, 527-537.

Travers, R. M. W. (1964) 'The transmission of information to human receivers', *Audio Visual Communication Review*, 12, 373-385.

Twyford, L. C., Davis, J. S. and Seitz, C. P. (1954) Evaluation of the film: military police support in emergencies (Riot Control). Technical Report no.sdc 269-7-52. Instructional Film Program. Naval Training Devices Centre; Office of Naval Research: Port Washington.

Vandermeer, A. W. (1954) 'Colour vs black and white in instructional films', *Audio Visual Communication Review*, 2, 121-134.

Vernon, P. E. (1946) 'An experiment on the value of film and filmstrips in the instruction of adults', *British Journal of Educational Psychology*, 16, 149-162.

Whalen, T. and Wright, D. (1999) 'Methodology for cost-benefit analysis of web-based telelearning: case study of the Bell Online Institute', *American Journal of Distance Education*, 13(1), 23-44.

Winn, W. and Snyder, D. (1996) Cognitive Perspectives in Psycholgy. In D. H. Jonassen (Ed) Handbook of Research for Educational Communications and Technology. New York: Simon & Schuster Macmillan.

APPENDIX 1: Published work 1977-2002(in press)

Educational Technology: towards an understanding of effective technologies, with particular reference to literacy.

Published work 1977 - 2002 (items not submitted for the degree of Doctor of Philosophy in **BOLD** type):

1. Spencer, K.A. (1977) An evaluation of the cost-effectiveness of audiotaped and videotaped self-study media presentations, British Journal of Educational Technology, 8(1), 71-77.

2. **Spencer, K.A. (1980) 'Audio-visual instruction for the mentally handicapped', Journal of the British Institute of Mental Handicap, 8(1), 10-13.**

3. Spencer, K.A. (1981a) An evaluation of the cost-effectiveness of film and tape-slide instructional presentations for the mentally handicapped, Programmed Learning and Educational Technology, 18(1), 23-9.

4. Spencer, K.A. and Clarke, A. (1981b) A BBC television series for the mentally handicapped, Educational Broadcasting International, 14(3), 110-113.

5. Spencer, K.A. (1981c) Alternative media: a review of the instructional effectiveness, Educational Broadcasting International, 14(2), 90-93.

6. Spencer, K.A. (1981d) AV communication: small is beautiful, Audiovisual Librarian, 7(1), 22-28.

7. Spencer, K.A. (1988) The Psychology of Educational Technology and Instructional Media, pp198, London: Routledge.

8. Spencer, K.A. (1990a) HyperCard: teaching technology for successful learning, Journal of Audiovisual Media in Medicine, 13, 25-30.

9. Spencer, K.A. (1990b) HyperHeart: does animated illustration contribute to mastery learning?, British Journal of Educational Technology, 21(3), 227-8.

10. Spencer, K.A. (1991) Modes, media and methods: the search for educational effectiveness, British Journal of Educational Technology, 22(1), 12-22.

11. **Simatos, A. & Spencer, K.A. (1992a) Children and Media, p.158, Liverpool: Manutius Press.**

12. Spencer, K.A. (1992b) Information technology and human information processing: a critical perspective, Proceedings of the European Conference about Information Technology in Education, Barcelona, November, 512-522.

13. **Spencer, K.A. (1996a) Media & Technology in Education, p. 173, Liverpool: Manutius Press.**

14. Spencer, K.A. (1996b) Recovering reading using computer mastery programmes, British Journal of Educational Technology, 27(3), 191-203.

15. Spencer, K.A. (1998) Predictive models of spelling behaviour in 7- and 11-year-olds, Journal of the Simplified Spelling Society, 24(2), 23-27.

16. Spencer, K.A. (1999a) Predicting spelling behaviour in 7 to 11 year-olds, Reading, July, 72-77.

17. Spencer, K.A. (1999b) Predicting word-spelling difficulty in 7- to 11-year-olds, Journal of Research in Reading, 22(3) 283-292.

18. Spencer, K.A. (1999c) English: the culpable language, Curriculum, 20(1), 7-12.

19. Spencer, K.A. (1999d) On becoming an expert, virtually. Paper presented at CAL99 "Virtuality in Education" Conference, 28-31 March, London, UK. Received "Outstanding Contribution Award".

20. Spencer, K.A. (1999e) Orthographic transparency in a computerised world. Paper presented at the Eleventh International Writing and Computers Conference, Cambridge, UK.

21. Spencer, K.A. (1999f) Educational Technology - An Unstoppable Force: a selective review of research into the effectiveness of educational media, Educational Technology and Society, 2(4), 23-34.

22. Spencer, K.A. (2000a) Is English a dyslexic language? Orthographic transparency and multilingual teaching, Dyslexia (Journal of the British Dyslexia Association), 6, 152-162.

23. Spencer, K.A. and Xing, M. (2000b) Creating communicative language environments via software. Paper presented at 17th Annual CALICO Symposium – CALICO 2000, Core Technologies: Impact on the Future, 30 May-3 June, Tucson, Arizona.

24. Spencer, K. A. and Xing, M. (2001) Differential effects of orthographic transparency on dyslexia. Paper presented at 5th BDA International Conference – Dyslexia: At the Dawn of the New Century, 18-21 April, York, UK.

25. Spencer, K. A. (2001, in press) 'Differential effects of orthographic transparency on dyslexia: word reading difficulty for common English words', Dyslexia, 7(4).

26. Spencer, K. A. (2002, in press) 'English spelling and its contribution to literacy problems: word difficulty for common English words', Reading.

APPENDIX 2 Supplementary materials: 1999-2002 (in press)

[1-7 on accompanying CDROM; 2, 3, 6 & 7 in hardcopy form]

1. Spencer, K.A. (1999) On becoming an expert, virtually. Paper presented at CAL99 Conference – Virtuality in Education, 28-31 March, London, UK. Received "Outstanding Contribution Award". [1.CAL99.pps]

2. Spencer, K.A. (1999) Educational Technology - An Unstoppable Force: a selective review of research into the effectiveness of educational media, Educational Technology and Society, 2(4), 23-34. [2.Edtech-soc1999.pdf]

3. Spencer, K.A. (1999) English: the culpable language, Curriculum, 20(1), 7-12. [3.Curriculum1999.doc]

4. Spencer, K.A. and Xing, M. (2000) Creating communicative language environments via software. Paper presented at 17th Annual CALICO Symposium – CALICO 2000, Core Technologies: Impact on the Future, 30 May-3 June, Tucson, Arizona. [4.Calico2000.pps]

5. Spencer, K. A. and Xing, M. (2001) Differential effects of orthographic transparency on dyslexia. Paper presented at 5th BDA International Conference – Dyslexia: At the Dawn of the New Century, 18-21 April, York, UK. [5.BDA2001.pps]

6. Spencer, K. A. (2001, in press) 'Differential effects of orthographic transparency on dyslexia: word reading difficulty for common English words', *Dyslexia*, 7(4). [6.Dyslexia2001.doc]

7. Spencer, K. A. (2002, in press) 'English spelling and its contribution to literacy problems: word difficulty for common English words', *Reading*. [7.Reading2002.doc]

2. Spencer, K.A. (1999) Educational Technology - An Unstoppable Force: a selective review of research into the effectiveness of educational media, Educational Technology and Society, 2(4), 23-34. [2.Edtech-soc1999.pdf]

Educational Technology – An Unstoppable Force: a selective review of research into the effectiveness of educational media.

Ken Spencer
Lecturer in Media Studies
Centre for Educational Studies
University of Hull, Hull
United Kingdom
k.a.spencer@educ.hull.ac.uk

ABSTRACT

Technologies are often adopted because they make processes more efficient. Some technologies, such as writing and printing, have been so successfully embedded in education that we are hardly aware of them: these are the ubiquitous technologies that have formed the very foundations of education for centuries. Other more recent technologies seem not to have been taken up by education. The reasons why some technologies are adopted, while others are not, are explored by considering educational technology research in an historical perspective. It is concluded from this research review that those technologies which appear to make education more efficient and productive, such as computer-based systems which incorporate mastery learning strategies, are those that will be adopted, becoming embedded and ubiquitous in the educational systems of the near future.

Keywords

Education, Technology, Methodology, Research, Audio-visual, Computers, Mastery learning, Future

Introduction

Let me start by, rapidly, reviewing the use of technology in education over the past few thousand years. First, there was language, a very powerful tool that allowed accumulated knowledge to be passed from one generation to the next, speeding up the evolutionary process. Then, after thousands of years, came written language, about 5,000 years ago, allowing thoughts and ideas to be transmitted across almost limitless vistas of time. Again, the process of human evolution was speeded up:

> "The invention of writing was highly significant for the development not only of language, but of society, and favoured the progress of commerce. It confirmed the power of the priests through the trained scribes, and even more the might and prestige of the ruler" (Singer et al., 1954).

The next major step forward was the invention of moveable type. Printing allowed a dramatic increase in the rate and volume of information distributed, and drove the European Renaissance. As many books were printed in the fifty years after Gutenburg's invention as had been produced by the scribes of Europe during the previous thousand years. And after such a revolutionary educational invention, things did not change a great deal for 500 years. Teachers taught a variety of subjects using basically the same tools: books and writing materials. Indeed, The Visible World in Pictures (Orbis Sensualium Pictus), produced by Comenius in 1658, was still widely used in schools the 19th century.

So, to a certain extent, teachers have always used technology, but the available technology was rather limited. With new technologies being introduced, in the 19th century, to boost production and aid in the distribution of the products of industrialisation, there was also a demand for improvements in education to meet the need for skilled workers and clerical staff. This was met by the monitorial training system introduced by Lancaster, based on Bell's ideas: it was a very efficient, but labour intensive, system, capitalising on human resources rather than the introduction of machines. Indeed, while Lancaster was introducing his system of schooling Charles Babbage and Ada Lovelace, were laying the foundations for the future revolution in educational technology, with their design for the precursor of the modern programmable computer, the analytic engine, which ran programmes from punched cards.

The twentieth century has been obsessed with the idea that the new communications technologies, such as film, radio and television, should make a significant impact on education, and that the introduction of such audio and visual aids should raise the level of achievement of pupils. The main early theory governing this approach, the realism theory, was summarised by Carpenter (1953):

"Sign similarity hypothesis: that films whose signals, signs, and symbols have high degrees of similarity ("iconicity") to the objects and situations which they represent will be more effective for most instructional purposes than films whose signals, signs and symbols have low degrees of "iconicity"... Signs (or symbols) vary in "iconicity" to the degree to which they are similar to the things or situations signified. Thus, for example, sound motion pictures have potential capacities for high degrees of "iconicity" in representing objects in motion as well as reproducing authentic sound" (p. 41).

Unfortunately, most of the research, which compared the new media, such as film and television, with traditional ways of teaching, found that there were no significant differences in student performance.

Early Film Research

With the advent of sound films for educational purposes, in the early 1930s, it was natural for researchers to consider the relative effectiveness of lectures and filmed recordings of such lectures. Hoban and Van Ormer's report, subtitled *Rapid Mass Learning*, summarised the research available during the period 1918 to 1950.

The earliest research investigating comparisons of sound films and lecture demonstrations was conducted by Clark in 1932. Three sound films, *Radioactive Substances*, *Liquid Air*, and *Characteristics of Sound*, were found to be as effective as the lecture demonstrations given by regular class instructors, in tests designed to measure thinking and reasoning ability.

A later study by Hall and Cushing in 1947 investigated the difference between a sound film presentation and a lecture with enlarged illustrations, dealing with 3 science topics. No differences were found.

In an experiment conducted by Vernon (1946) information test scores of seamen who witnessed two showings of a film demonstrating two methods of taking depth soundings, a total of 50 minutes of instructions, were only 6 per cent below those of groups which had usual instruction lasting 3 hours. Vernon concluded that an hour's film appears to be as effective as three hours weak oral instruction.

Hoban and Van Ormer summarised their review of the research, in which films are compared with demonstrations or lectures, and conclude that films reduce instruction time and are often equivalent to good instructors. They hasten to add that this should not be interpreted as meaning that films can eliminate the need for instructors but rather that the effectiveness of instructors of average or below-average ability can be improved (and instructional time can be saved). Also, they suggest that films can be projected on large screens, increasing the size of the viewing group but with no loss in instructional effectiveness. Finally, they conclude that films, used alone, can offset a shortage of instructors.

> Films Reduce Instruction Time:
>
> The conclusion which continually recurs in these studies is that films reduce instruction time with little or no sacrifice of instructional results. In some of the experiments in film presentation, from one-half to two-thirds of the instructional time was saved by the use of films in place of lecture or demonstration. There thus appears to be considerable support for the Navy slogan "More learning in less time", although this may not always be true for all films.

> Films are Often Equivalent to Good Instructors:
>
> A second conclusion that is recurringly supported by the research data is that, in communicating facts and demonstrating concepts, films (or filmstrips) are about equivalent, and sometimes better than superior instructors using the best non-filmic materials at their disposal (Hoban, C.F. and Van Ormer, E., 1950).

From the results of the first thirty years of research comparing the comparatively new educational media there is strong confirmation of their effectiveness when compared with traditional methods. Indeed, Hoban and Van Ormer suggest that by using film recordings of above average teachers there can be some compensation for poorer quality personnel.

Peggie Campeau's stringent review of the literature concerning audio-visual media, in 1966, cites nine studies conducted at university, senior high school, junior high school and elementary school levels in which no significant differences in achievement were found when students were taught by either motion pictures or conventional instruction. This is confirmed by Greenhill in his introduction to a volume of abstracts on film and tv research by MacLennan & Reid (1964).

Early Television Research

Just as research was indicating quite conclusively that there is no disadvantage in studying from filmed courses, a new technological innovation was entering the mass media market place. Television, although invented in the late 1920s, was only beginning to make headway in the early 1950s, and it was natural for researchers to turn their attention to a medium which offered all the potentialities of film, at a possible lower cost, for, as Lumsdaine and May concluded in 1965, film and tv can be considered substantially identical media for many purposes.

The first research effort used television as a means of expanding the total audience for a given lecture via closed circuit television, in which the transmission was carried from the lecture room, by cable, to several locations, enabling a single teacher to communicate with many hundreds of students. In fact, at one stage, audiences of 7,000 were taught via closed circuit tv in New York University. In Greenhill's report of the Pennsylvania experiment he acknowledges that the major use of closed circuit tv was for the presentation of regular classroom instruction to students located in multiple classrooms, as a means of coping with mounting enrolments. Greenhill also argues that the standard of instruction would also be raised by:

(1) extending the influence of its best professors to large numbers of students and
(2) by making it possible for these professors to present demonstrations and other teaching materials that it would be impossible or impractical to use under normal classroom conditions.

This new medium, which is substantially the same as motion film, in that it presents moving pictures with sound accompaniments, should produce substantially the same results when compared with traditional teaching. The shift towards tv, which was viewed as a panacea for all the educational ills during the mid 1950s and early 1960s, resulted in a proliferation of research studies. Stickell, in 1963, reviewed 250 comparisons of educational television and conventional face-to-face instruction from 31 research reports. Overall, 75% of the studies showed no difference, with equal percentages favouring tv or face-to-face instruction. Chu and Schramm found that, by 1968, of 421 separate comparisons taken from 207 published reports, 308 showed no difference, 63 showed tv to be superior and 50 found conventional instruction superior.

The research reviewed was so wide-ranging that they concluded that tv can be used efficiently to teach any subject matter where one-way communication will contribute to learning.

These results lead to the inevitable conclusion that course enrolment can be greatly expanded by the use of educational television and that student performance will not suffer. However, the second point referred to by Carpenter, concerning the possible improvements in instruction, does not seem to be tenable. If this is the case, what are the advantages of using film or television? The answer is: when there is no difference in performance, the most obvious measure then to come under scrutiny is the cost.

An early cost analysis is available in Greenhill's final report on tv teaching at Pennsylvania State University, during the period 1956-57. Comparisons were made between actual costs of televised instruction and the costs that would have been incurred in courses had they been taught in the usual way. The analysis showed a total saving, in favour of the tv service, of $40,000, which represented more than the total cost of running the service. However, it was necessary to have at least 200 students per course before any savings were made. The system required large recurrent finance to keep it in operation and, if the system was not fully operational for most of the academic year, it rapidly became less cost-effective.

Radio and Audio Recordings

It is generally recognised that for teaching via radio or audio recordings essential graphic information, usually in the form of printed materials, must be provided in order to fully exploit the potential of the medium. This means that comparisons of radio and traditional teaching are often comparisons of audio plus print with traditional teaching.

Beginning in the 1920s instructional radio was widely used in the United States and Britain, but with the advent of television its use dwindled in the US, although it has continued to be widely employed by the BBC schools service. Developing countries, however, are making increasing use of radio, its principal attraction lying in its low cost when compared with television. It is also an effective instructional medium, as much of the research confirms.

Carpenter, in 1937, prepared 15-minute radio lessons on science for pupils ranging from fourth grade to senior high school level. The results of the end-of-term examinations indicated that pupils taught by radio did as well as, or better than, those taught by conventional methods. Attitude reports from pupils showed a high degree of interest in radio lessons. Heron & Ziebarth (1946) found that the radio was as effective as the face-to-face instruction in college psychology courses.

Another example of the effectiveness of radio as a teaching instrument, this time in a developing country, is reported by Mathur & Neurath (1959). A total of 145 villages in Bombay state, averaging about 850 people per village were chosen as the experimental group and were provided with radio sets. A similar number of villages without radio sets served as the control group. Twenty special farm programmes were broadcast twice a week for 30 minutes. Comparison of test results, both before and after the broadcast programmes, found a significant increase in knowledge in the radio villages, but only negligible increases in the non-radio groups. This and many other reports indicate the efficacy of instructional radio. Forsythe's (1970) review concluded:

> "Research clearly indicates that radio is effective in instruction. Experimental studies comparing radio teaching with other means or media have found radio as effective as the so-called "conventional methods". Even though radio has been criticised for being only an audio medium, studies have shown that visual elements in learning are not uniformly important. In many educational situations visuals may be more harmful than helpful. Also, the efficiency of combined audio and visual media has been challenged by studies which show that multi-channel communications may not be inherently more effective than single channel presentations."

These conclusions can be extended to include the equivalent recorded form of instruction, such as discs or tapes. Popham, in 1961, divided an introductory graduate course into two sections. In one he taught in a lecture-discussion format; in the other, he played a tape-recorded version of the lecture and then led a brief discussion period. The two groups were carefully matched on scholastic aptitude and two achievement pre-tests. Following instruction several post-tests were administered and it was found that there were no differences between the groups.

By 1969, the rapid expansion of cassette tapes was making the medium ideally suited to individualised learning and it was this aspect which was of interest to Menne, who used it in an introductory psychology course at Iowa State University. The lectures were recorded on tape and notes were taken from the blackboard material used by the instructor during the presentation of his lectures. The blackboard notes were then assembled to form a booklet. Each member of the experimental tape group was issued with a tape recorder, a complete set of lecture tapes, a booklet containing the transcribed blackboard material and a schedule of the lecture topics to be given to the lecture group. The audio-taped group was self-paced, though they were required to take 3 objective tests during the course. Information was available concerning student performance on several measures, which enabled a covariance analysis to be applied to the results for the regular exams, class points and a final grade. There were no significant differences.

Ackers and Oosthoek (1972) reports a similar experimental course. The subjects in the taped group again had individual access to recorders and tapes on the subject of micro-economics, and were able to follow the course at their own speed, within certain broad limits. Ackers does not elaborate on the broad limits and we can consider the taped course to be substantially student-paced. Both groups had ample opportunity to participate in test problems, which formed an integral feature of the instruction, and were encouraged to take part in fortnightly group discussions. The performance was assessed in a June examination which, according to the authors, called for the sub-categories 'Application' and 'Analysis' of the category 'Comprehension' from Bloom's Taxonomy of Objectives. The results indicated a slight advantage in favour of the tape group.

General Conclusion from early media research

Much of the early research was concerned with the new mass media, and it was clear from this research that these new approaches produced results similar to more traditional methods of teaching. Traditional approaches had a limited number of technologies embedded in them, such as books, and writing and drawing materials; the new approaches incorporated the teachers themselves, whose performance had previously been evanescent. The new media removed the need to have teachers actually present in the classroom, because they can produce a facsimile of the teacher. This represents a significant change, because prior to the introduction of the new media, if a teacher was not present, instruction could only be given through the medium of print, which need a decoding skill, the ability to read, to be present in the student. The new media literally spoke directly to the students, and did so as effectively as if the teacher was actually present. This lies as the heart of the embedding process: the new media could actually replace teachers, although researchers usually denied that this was a possible outcome, and their results were euphemistically disguised as demonstrating that the new media could provide a means to compensate for a lack of teachers. Nonetheless, teachers were well-able to perceive the threat from the new media.

Media Attribute Research

Early research into the effectiveness of media considered the effectiveness of the new medium compared to traditional approaches. In addition to this, groups of media which differed along one dimension only, were compared in order to determine the effectiveness of a particular attribute, such as visual motion or pictorial colour.

Motion

One of the earliest accounts of the experimental investigation of media effectiveness is Freeman's (1924) Visual Education, published 20 years after the first public demonstration of moving pictures by the Lumiere brothers. The results are not as scientifically valid as later research, but they do confirm later findings.

McClusky's (1924) experiment represents one of the first comparisons between film and a lecture illustrated with slides. He used a film on the life history of the Monarch butterfly and compared this with two lecture conditions: a slide lecture using eight slides, each illustrating a step in the life cycle; an oral presentation illustrated with two pictures and two blackboard sketches. Each was presented to 20 pupils in grades 6-8 in two schools and lasted 12 minutes. The results failed to show any difference between the methods.

Brown (1928) found similar results when comparing films and filmstrips for teaching factual information about the physiology of seeing to high school students. In the filmstrip group discussion was free and questions were asked both by the teacher and the students. A multiple choice test indicated a superior performance for the filmstrip group and Brown concluded that this was because of the greater exchange of comment within the teacher-paced filmstrip group.

A more satisfactory approach to determining the effectiveness of the visual motion attribute was undertaken by Twyford (1954). The topic investigated was methods of riot control, under the title Military Police Support in Emergencies, and introduced the problems of training soldiers to cope with such complex situations as restless, disturbed city populations, agitated groups, mobs and rioting crowds. The film had a Hollywood budget and expensive crowd scenes organised in an American city. The question of simpler and less expensive production methods was raised and Twyford was charged with determining the effectiveness of other methods. Twyford's group suggested an alternative approach using stock film or newsreel coverage of riots, and even the use of still pictures if motion film was not available. The project eventually compared the Hollywood style film with two filmograph versions, which were similar to sound filmstrips. In all three versions the soundtrack was identical. One filmograph was based on the motion film and consisted of individual still frames taken from the original. The second filmograph was made up of stock still pictures of riots taken from news libraries, or simple diagrammatic representations of troop movements. The groups of recruits were tested using a 42 question test with 10% of the questions using pictures. The full motion version scored 4% more than the filmographs, which were equally effective. The difference in performance was educationally insignificant, but the difference in cost was very substantial.

This is one of the earliest well-controlled experiments that shows that motion aids such as film and television will not automatically improve student performance when compared with simpler aids such as filmstrips. The essence of this argument is that film and tv can teach many different groups and subjects about as effectively as traditional methods, but so can simpler aids such as sound filmstrips or sound tapes with booklets, and these simpler aids cost less to produce.

Colour and Pictorial Quality

It is also worth, at this point, considering the experimental evidence concerning comparisons between colour illustrations and their counterparts in monochrome, and the effects of changing pictorial quality.

Vander Meer's (1954) work has been described as demonstrating a rigorous methodology which sets the standard for similar studies. The first experiment involved 500 students, 14 and 15 years old. One half of the students saw colour versions of 5 films, whilst the other half simultaneously saw black and white prints made from the original colour materials. The films were commercially produced titles including: Maps are fun; How man made days; Rivers of the Pacific Slope; Snakes, and Sulphur & its compounds. Two types of tests of perceptual and conceptual learning were developed for each film: non-verbal and verbal. The results for the verbal tests indicate that in only one case was there a statistically significant result in favour of the colour film version. The non-verbal test results reverse the statistically significant results for the verbal tests, with two of the three films favouring the b & w version. However, the differences do not persist and the delayed recall test indicates no difference between the two versions.

The main conclusion reached by Vander Meer was:

> "The use of colour in instructional films which may superficially seem to 'call for colour' does not appear to be justified in terms of greater learning on the part of those who view the films. If colour is to be used effectively in films there must be careful preproduction consideration of the probable psychological impact of specific uses of colour upon the learner."

A similar project was undertaken at Yale University (May and Lumsdaine, 1958). In the "Learning from Films" report the effects of pictorial quality and colour is considered, especially the importance of factors generally regarded as entering into the degree of polish or quality of the pictorial component of teaching films, both factors being related to the cost of producing and printing films.

The Yale team produced a colour film 'Seasons', which dealt with the causes of seasonal change, and was to be used to investigate the efficacy of colour instructional films. During the production phase a story board was produced as a guide for the eventual production of animated and live colour footage. The story board consisted of very crude b & w pencil sketches for each scene. In order to aid in visualising the content of the final film a so-called pencil test running reel was made by photographing these sketches on motion picture film in the planned sequence.

Before the final film became available a silent print of the pencil test version was shown to a sixth grade class, with a staff member reading the commentary. The post test scores were later compared with the performance of a similar class who viewed the full colour version and the result was that the learning from the crude, jerky b & w version was substantially as great as from the full colour version. These surprising results were at first only accepted as being very tentative, because the groups compared were not selected to assure equivalence or provide any valid measure of error. A second experiment utilised 4 classes of fifth grade pupils, and Lumsdaine concluded that the difference between the two mean scores was so small that it is interpreted as the result of mere chance fluctuations — in other words they are not statistically significant.

Kanner and Rosenstein in 1960 evaluated the need for colour rather than b & w instructional television in the US Army. The report indicates that reliable colour television equipment was available but that the cost was higher than monochrome equipment although costs were expected to fall with technological developments as indeed they have done.

The experimental study beamed the eleven lessons from a mobile colour television facility into two classrooms, one viewed the lessons on colour receivers while the other group viewed them on monochrome receivers. Immediately following a lesson the subjects were tested using multiple-choice questions. Every effort was made to incorporate colour items into the tests. A total of 368 trainees took part in the experiment and pairs of subjects

were matched on electronics aptitude or general technical scores and then were randomly assigned to one of the two experimental conditions.

Ten out of 11 comparisons show no significant differences and the single statistically significant result is considered to be unimportant in view of the overall picture and small differences in test performance. The overall mean scores are remarkably similar, bearing out the results of May and Lumsdaine.

Stephen Cox produced a survey of the research into the effects of colour in learning from film and television in 1976, and from the results of twenty or so studies concluded that overall there is no marked difference in learning from colour or black and white film or television.

Dwyer took this research forward in an attempt to improve visualised instruction. He planned and carried out his programme of systematic evaluation of the effects of a variety of pictorial types over an extended period of time, involving 100 separate studies with a total population of 23,000 students.

The first piece of research in this programme is recorded in Dwyer's 1967 research report: Adapting Visual Illustrations for Effective Learning, published in the Harvard Educational Review. He describes an experiment which compares four different audio-visual presentations of the same material. The commentary is the same for each presentation but the pictorial image is different.

Initially, only black and white illustrations were considered. The four conditions used in the experiment were:

1. Oral/verbal presentation. The students in this group saw no accompanying illustrations, but the name of the parts of the heart mentioned were projected on the screen.
2. Abstract linear representations, later called simple line drawing: the line drawings used in this presentation were similar to instructional drawings used in many current science books.
3. Detailed shaded drawing presentation: the illustrations were more complete than the simple line drawings, and they represent the heart and its related parts more realistically rather than merely identifying and locating them as in the simple line drawings.
4. Realistic photographic presentation: photographs of the heart were used.

Dwyer also designed four individual criterial measures, which were administered in the following order: drawing, identification, terminology, comprehension tests. After the presentation of the instructional materials each student was permitted to take as much time as he required to complete one criterial measure before proceeding to the next. The important question, a re-statement of the 'Sign Similarity Hypothesis', was: do students learn more if illustrations are more realistic?

The results of this study indicated that when students viewed their respective instructional presentations for equal amounts of time, the simple line drawing presentation was significantly more effective in facilitating achievement than was the oral presentation without visuals on the drawing, identification, and total criterial tests. The oral presentation without visuals of the heart was found to be as effective as each of the visually complemented treatments on both the terminology and comprehension tests. Dwyer also concluded that, contrary to previously stated theories of visual communication (e.g. Carpenter, 1953), the more realistic illustrations were found to be the least effective in complementing oral instruction.

The purpose of this and other experiments was to test the hypothesis that an increase in realistic detail in visual illustrations increases the probability that learning will occur. Eventually nine slide sequences, possessing differing degrees of realistic detail and colour, were produced so that variations in visual stimuli could be assessed in terms of their ability to facility student achievement on five criterion measures. The results indicated that increasing the amount of realistic detail in visual illustrations does not necessarily lead to greater learning.

Why are there no differences?

There are no differences because much of what is happening in mediated instruction hardly differs from what is happening in the classroom. And the classroom is a most inefficient device for education. The reason colour, motion, or pictorial quality adds little to understanding a range of topics, is that the human information processing system has processing limits and best deals with information that has been simplified. This actually matches better with the types of tests which are administered to measure learning, such as those used by Dwyer: simple line drawings are best for instruction when assessment uses simple line drawings. Travers (1964)

successfully linked this to the emerging discipline which was seeking to apply information theory to psychology. He demonstrated that much of the information that is attended to by the sense organs is actually filtered out before it reaches the higher levels of cognitive processing. In many cases, as exemplified by Dwyer's line drawings, simplification makes the world more comprehensible because it places less demands on the processing system: it is, by its very nature, partly processed, the extraneous information having been stripped out.

There are no differences because the information passed to the student by the teacher, the television or radio programme, book or picture, is not usually sufficiently well-adapted to the student's needs. The information is often too much, in quantity or speed of delivery, and the student perceives only a fraction of it, and understands even less.

Can there be any differences?

There must be an unequivocal "yes" to this question. Students who are average in a class can be turned into above-average "A"-grade students, but not by merely changing the medium of instruction. Clark (1983) concluded from the research on learning from media that:

> "Consistent evidence is found for the generalisation that there are no learning benefits to be gained from employing any specific medium to deliver instruction."

"Media in education" is often misconstrued as being what educational technology is all about. In fact, it is only one aspect of educational technology and should be more properly termed: technology IN education. It does not have a good record for showing significant educational gains when introduced into the classroom. However, there is a technology OF education that has been much more successful in raising levels of achievement. Its roots are in the behavioural sciences, and can be traced back to the work of Skinner and Bloom; it is associated with mastery learning methodologies. A review of the relative effect sizes of different media and methods (Spencer, 1991) shows that the Learning for Mastery (LFM) approaches advocated by Bloom (1968) and Keller's (1968) Personalised System of Instruction (PSI) produced educationally significant results when compared to the media approaches.

The Personalised System of Instruction is a totally student-paced system, in which students can take course units at any time, and then receive immediate feedback on their performance on tests taken directly after each unit of study. The feedback is provided by proctors, who are paid students who have already completed the course of study. This approach is very similar to that advocated by Lancaster in the monitorial system. Students can only move on from one unit to the next when they have achieved a high level of proficiency in the work being studied: a criterion level of at least 80% correct is usual, and Kulik (1986) has shown that raising that criterion level to >90% produces a much larger effect size of 0.8, compared to 0.4 for 70-80%.

Learning for Mastery, Bloom's approach, is teacher-paced and more suitable for use in the classroom. It requires constant formative evaluation of each pupil's performance, with support for failing students being provided by the use of audio-visual aids and other suitable remedial materials, and extra sessions with pupils and adults providing remedial help. Bloom is adamant that the tests must be diagnostic:

> "For students who fail to master a given unit, the tests should pinpoint their particular learning difficulties — that is, the specific questions answered incorrectly and thus the particular ideas, skills, and processes which need additional work. We have found that students respond best to diagnostic results when the diagnosis is accompanied by very specific prescription of particular alternative instructional materials and processes they can use to overcome their learning difficulties" (Bloom, 1968).

It is interesting to compare these mastery approaches with Postlethwait's (1972) Audio Tutorial (AT) method. Postlethwait's system, based on the provision of audio-taped lectures, has many of the characteristics of LFM and PSI: it is student-paced within a set time for particular units of work (often 1 week), with students visiting the learning centre whenever they need to study; there is continual formative assessment, usually weekly, with students completing a written and oral test; the course is broken down into small units; and students are provided with a set of learning objectives for each unit. However, there is no mastery requirement and students can move from one unit to the next obtaining very low grades. When compared with traditional approaches, this system shows only a small effect size, and the reason seems to be that no matter how flexible a system is, no matter how carefully instructional materials are prepared for students, if the students do not invest sufficient mental effort in

the learning process they will fail to master all the material presented. If later units require full understanding of earlier materials, then those students who have previously achieved only partial understanding will inexorably develop a cumulative learning deficit, and the gap between top and bottom students will widen.

A Synthesis for the Future?

We must begin to accept that what takes place in the classroom can be replaced by a whole host of alternative media, without a deterioration in pupil performance. Research has demonstrated this using different media, teaching subjects and ages of students. If we wish to improve levels of performance, we must look to the new methodologies. Those systems that incorporate mastery learning strategies seem to offer the most hope for such improvements. Bloom and Keller have demonstrated mastery systems based on traditional media, essentially human resources, backed up with written materials, but this mastery methodology can also be automated and applied to computer-based systems. Spencer (1996) demonstrated such an effective system for teaching reading and spelling to pupils who had failed to gain literacy skills. Integrated learning systems, which manage student progress by constantly assessing performance and indicating suitable learning materials, are also following this path. At this stage results are variable, but encouraging:

> "In the first phase of the evaluation pupils using SuccessMaker made learning gains in numeracy above that of equivalent control groups (an effect size of 0.4 which was equated to progress of 20 months over a six-month period" (P.14, NCET, 1996).

	Effect Size	
Programmed learning (Secondary)	0.08	*
Individualized instruction	0.10	*
Ability groups	0.10	*
Visual-based instruction	0.15	**
Programmed learning (Higher)	0.24	**
Audio-Tutorial	0.25	***
Assigned homework	0.28	***
Computer Assisted Instruction	0.29	***
Tutoring	0.40	****
Integrated Learning Systems	0.40	****
Computer Simulations	0.49	*****
Personalised System of Instruction	0.58	******
Learning for Mastery	0.80	********

* represents the size of effect in tenths of a standard deviation

[An effect size of 1.0 improves an average pupil's performance such that they achieve results previously associated with the most able 10% of the class; an ES of 0.25 and less is considered educationally insignificant]

Table 1. Summary of Effect Sizes

We must remember that the development of sophisticated computer-based systems is in its infancy, but even so, as Table 1 illustrates, it is now beginning to equal the effectiveness of individual tutoring methods. It is not surprising that the computer begins to excel when it is used to provide simulations. It is inevitable that we shall see the computer competently simulating human one-to-one exchanges, passing what is sometimes known as the Turing test.

The recent chess victory of the Deep Blue computer over world champion Kasparov has fulfilled the expectations that Shannon and Weaver (1949) expressed in their theory of information, published just 50 years ago. They tentatively suggested that information theory could lead to intelligent systems tackling the supreme game of chess. Their work stemmed out of the advances made in communication systems during the Second World War, and the impact of new computer methods for breaking codes. It was also related to the concept of feedback in machines, based on work with rockets and missiles, proposed by the father of cybernetics, Norbert Weiner.

The greatest impact of Deep Blue's achievement is likely to be in the field of education. The key to education is the assimilation of information to create new mental schemes, which enable us to look at the world anew, to go beyond the obvious. And how do we do this, how does this mental construction come about? By entering into a dialogue with the world, the physical world and the social world, as suggested by Piaget (1971) and Vygotsky (1962). One way in which we extend the capabilities of the child is to have her enter into a dialogue with an expert who can answer questions, provide hints, set expectations. The expert has traditionally been the teacher, but there has been a search for mechanical means to do the same job, just as there were searches for mechanical devices to play chess.

Skinner was the most effective early proselytiser of mechanical methods using feedback mechanisms, demonstrating his teaching machines in 1954. He argued that a country that could mass-produce washing machines and cars could surely develop a machine for providing sufficient feedback to students to enable them to reach high levels of performance, especially in basic numeracy and literacy. Programmed learning was only moderately successful. Later computerised methods, using essentially the same limited psychology, have shown themselves to be at least as effective as teachers, and in some cases even more effective.

However, the newer integrated learning systems, combining mastery strategies with the ability to provide rapid feedback and make decisions about suitable remedial materials, represent the true state of Educational Technology. But, much more is to come, and it will come soon. Already it has been demonstrated that artificial intelligence (AI) systems can teach, or rather tutor, as effectively as human tutors in advanced courses (Lajoie and Derry, 1993). Intelligent tutor systems are now being used in college level maths courses, helping students gain understanding of the complexities of geometry proofs. These computer-based tutors, using diagnostic modelling procedures, have been found to be as capable as their human counterparts at identifying and correcting student misunderstandings, even in the complex, advanced field of avionics.

Artificial intelligence systems actually learn from experience and develop powerful rules and strategies, which when applied may be even more effective than those used by the human experts whose skills have been tapped by the computer. Such tutoring systems will get to know their pupils, will have extracted rules and strategies for optimal methods of teaching, and pupils will really have the very best of personal tutors, a dream that has been sought for many years. Deep Blue's victory shows that the potential for sophisticated dialogues with an understanding, motivating, superior intelligence are achievable. There can be no doubt that just as the filing cabinet size of the first teaching computers was reduced, within a few years, to that of the watch on my wrist, so computers more powerful than Deep Blue will shrink to sit on a child's desk, early in the next millennium.

And teachers, what will become of them? Teachers will always be needed, because of the human touch. Their role will undoubtedly change; it may even become more rewarding. And, of course, teachers do use educational technology. They always have done: the written word, on a blackboard or in a book, represents technology which is so embedded in teaching that we hardly acknowledge its presence. The new technologies, such as artificial intelligence tutors, are just emerging. Soon they, too, will become ubiquitous, and so totally embedded within the educational context that they will become transparent, in much the same way that written communication is hardly noticed as an embedded technology today.

References

Ackers, G. W. & Oosthoek, J. F. (1972). The evaluation of an audio-tape mediated course-I. *British Journal of Educational Technology*, 3, 136-147.

Bloom, B. S. (1968/70). Learning for mastery. *(UCLA-CSEIP) The Evaluation Comment*, 1 (2), Reprinted in Block, J. H. (Ed.) *Mastery Learning*, New York: Holt, Rinehart and Winston, 47-63.

Brown, H. E. (1928). Motion picture or film-slide. *Science and Maths*, 28, 517-526.

Campeau, P. L. (1966). Selective review of the results of research on the use of audio visual media to teach adults. In Briggs, L.J., Campeau, P.L., Gagné, R.M and May, M.A. (Eds.) *Instructional media: a procedure for the design of multi media instruction and suggestions for future research*, Monograph 2, American Institutes for Research: Pittsburgh, 99-176.

Carpenter, C. R. (1953). A theoretical orientation for instructional film research. *Audio Visual Communication Review*, 1, 38-52.

Carpenter, H. A. (1937). Science in the Rochester school of the air. *Science Education*, 21, 77-81.

Chu, G. C. and Schramm, W. (1968). *Learning from television: what the research says*, Institute for Communication Research: Stanford University.

Clark, C. C. (1932). *Sound Motion Pictures as an Aid in Classroom Teaching*, Doctoral Thesis: New York University.

Clark, R. E. (1983). Reconsidering the research on learning from media. *Review of Educational Research*, 53 (4), 445-459.

Cox, S. (1976). Colour in learning from film and tv: a survey of the research with some indications for future investigations. *Educational Broadcasting International*, June, 85-96.

Dwyer, F. M. (1967). Adapting visual illustrations for effective learning. *Harvard Educational Review*, 37, 259-263.

Forsythe, R. D. (1970). *Instructional Radio: a position paper*, Stanford University. ERIC ED 044 933.

Freeman, F. N. (1924). *Visual Education*, Chicago: University of Chicago Press.

Greenhill, L. P. (1959). *Closed Circuit Television for Teaching in Colleges and Universities*, Pennsylvania: Pennsylvania University (mimeographed).

Hall, W. E. and Cushing, J. R. (1947). The relative value of three methods of presenting learning material. *Journal of Psychology*, 24, 57-64.

Heron, W. T. and Zeibarth, E. W. (1946). A preliminary experimental comparison of radio and classroom lectures. *Speech Monographs*, 13, 54-57.

Hoban, C. F. and Van Ormer, E. B. (1950). Instructional Film Research 1918-1950, *Technical Report No. SDC-269-7-19*, Port Washington, New York: Special Devices.

Kanner, J. H. and Rosenstein, A. J. (1960). Television in army training: colour versus black and white. *Audio Visual Communication Review*, 8, 243-252.

Keller, F. S. (1968). Goodbye, Teacher. *Journal of Applied Behaviour Analysis*, 1, 79-87.

Kulik, C. C., Kulik, J. A. & Bangert-Drowns, R. L. (1986). Effects of Testing for Mastery on Student Learning. *Paper presented at the annual meeting of the American Educational Research Association*, San Francisco, April 16-20, 1986.

Lajoie, S. P. and Derry, S. J. (1993). *Computers as Cognitive Tools*, Lawrence Erlbaum Associates: Hillsdale, New Jersey.

Lumsdaine, A. A. and May, M. A. (1965). Mass communication and educational media. *Annual Review of Psychology*, 16, 475-534

Mathur, J. C. and Neurath, P. (1959). *An Indian Experiment in Farm Radio Forums*, UNESCO: Paris.

May, M. A. and Lumsdaine, A. A. (1958). *Learning from films*, Yale University Press: New Haven.

McClusky, F. D. (1924). Comparison of different methods of visual instruction. In F. N. Freeman (Ed.) *Visual education*, Chicago: University of Chicago Press, 83-166.

Menne, J. W., Hannum, T. E., Klingensmith, J. E. and Nord, D. (1969). Use of taped lectures to replace class attendance. *Audio Visual Communication Review*, 17, 42-46.

NCET (1996). *Integrated Learning Systems: A report of phase II of the pilot evaluation of ILS in the UK*, London: NCET.

Piaget, J. (1971). The Science of Education and the Psychology of the Child, London: Longmans.

Popham, J. W. (1961). Tape recorded lectures in the college classroom. Audio Visual Communication Review, 9 (2),109-118.

Postlethwait, S. N., Novak, J. and Murray, H. T. (1972). The Audio-Tutorial Approach To Learning, Minneapolis: Burgess Publishing Co.

Reid, C. J. and MacLennan, D. W. (1967). Research in Instructional TV and Film, US Government Printing Office: Bureau of Research: Washington.

Shannon, C. E. and Weaver, W. (1949). The Mathematical Theory of Communication, Urbana: University of Illinois Press.

Singer, C., Holmyard, E. J. & Hall, A. R. (1954). A History of Technology, Volume 1, Part 1, Chapter 4: Speech and Language, Oxford: Oxford University Press.

Skinner, B. F. (1954). The scienc of learning and the art of teaching. Harvard Educational Review, 24 (2), 86-97.

Spencer, K. A. (1991). Modes, media and methods: the search for educational effectiveness. British Journal of Educational Technology, 22 (1), 12-22.

Spencer, K. A. (1996). Recovering Reading Using Computer Mastery Programmes. British Journal of Educational Technology, 27 (3), 191-203.

Stickell, D. W. (1963). A Critical Review of the Methodology and Results of Research Comparing Televised and Face-to-Face Instruction, Doctoral dissertation: Pennsylvania State University.

Travers, R. M. W. (1964). The transmission of information to human receivers. Audio Visual Communication Review, 12, 373-385.

Twyford, L. C., Davis, J. S. and Seitz, C. P. (1954). Evaluation of the film: military police support in emergencies (Riot Control). Technical Report no.sdc 269-7-52, Instructional Film Program. Naval Training Devices Centre; Office of Naval Research: Port Wa

Vandermeer, A. W. (1954). Colour vs black and white in instructional films. Audio Visual Communication Review, 2, 121-134.

Vernon, P. E. (1946). An experiment on the value of film and filmstrips in the instruction of adults. British Journal of Educational Psychology, 16, 149-162.

Vygotsky, L. S. (1962). *Thought and Language*, London: John Wiley.

3.   Spencer, K.A. (1999) English: the culpable language, Curriculum, 20(1), 7-12.
[3.Curriculum1999.doc]

English: the culpable language

Evidence from recent international surveys indicates that literacy in the UK is at an unacceptably low level (DfEE, 1999). Why is this the case? Is it due to poor teaching methods, inherent personal deficits in our children, or could it be the fault of the English language, and in particular the complexity of the written form of English?

Some languages have a simple one-to-one mapping of sounds and letters: a language with 40 sounds would have 40 symbols, and no divergence from the way in which each sound is represented. Turkish is a good example of this approach, as is Finnish. Other languages, such as German, are not as rigorous as this, but they often have strict rules governing the form of representation.

It follows from this that some languages will be easier to read and write than others, because the actual form of the written representation, and especially the number of alternatives for each sound, affects accuracy and speed of decoding. Writing is a key technology; it was one of the earliest technologies and had a profound effect on the development of civilization. Like all technologies, the exact way in which it is implemented will make it more or less efficient. It can be hypothesized that simple mapping will make for a more efficient system than complex mapping; and it will follow from this that reading and writing languages such as Turkish and German will be easier than English.

Oney and Goldman (1984) confirmed that this was the case with Turkish, which was radically modernized in the 1920s. They compared first and third grade pupils in American and Turkish schools. Turkish pupils were faster and more accurate at reading words than their American counterparts. There was little difference between Turkish first and third grade pupils: it was almost as though, having learnt the skill, they were set up for life. Turkish first-grade children actually outperformed American third-grade children on most words, and especially on three syllable words. Turkish has an almost perfect grapheme-to-phoneme correspondence, and this orthographic transparency is clearly a more efficient writing technology than the system used in English.

Figure 1:   Comparison of Reading Performance for 3 and 9 Letter Words for American Third-Grade and Turkish First-Grade (based on Oney, 1984).
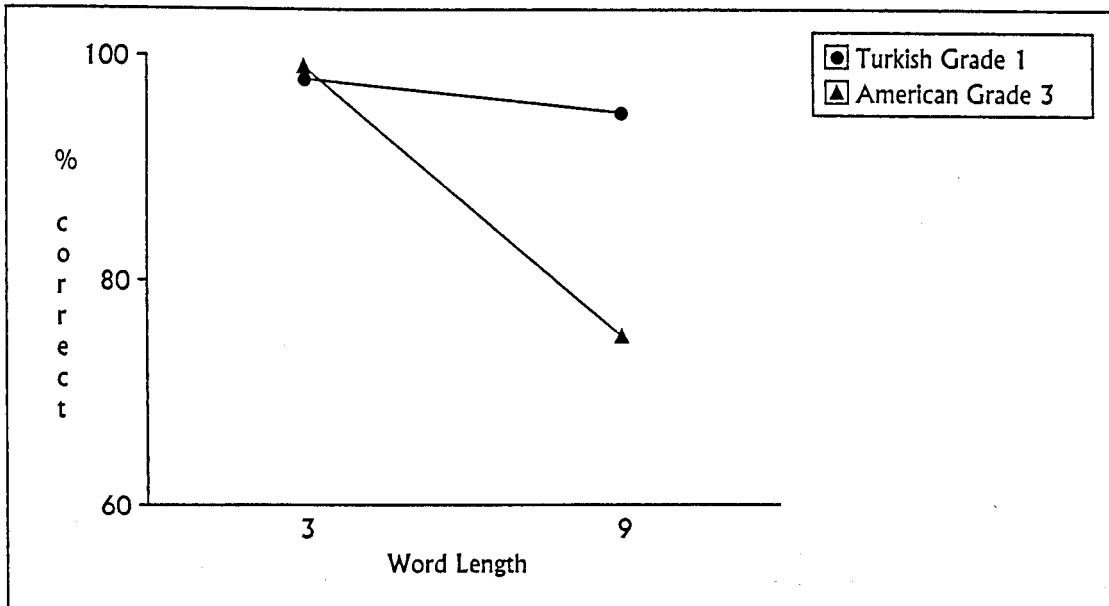
Landerl, Wimmer and Frith (1997) concluded that if letter-sound relations are consistent, then even a child with dyslexia could learn to map print onto speech with little or no delay in reading acquisition. In their study, English dyslexic children made many more errors than the German dyslexic children. The German pupils actually read the words in the most difficult category, three-syllable non-words, more accurately than the English children read the real one-syllable words, the easiest category. Performance was related to speed of reading as well as accuracy: the English children also took much longer to decode the words. These differences were also found in the study's reading-age equivalent pupils, and later studies have confirmed this to be the case with a variety of age groups (Frith et al, 1998). Even after their prolonged period of acquisition of reading skills, 12-year-old English-speaking children were still error-prone, although they read at the same speed as German children. It can be argued that systematic phonics teaching, as used in German schools, accounts for the large differences. However, as Frith et al indicate, orthographic consistency and phonic teaching methods are linked: a consistent orthography lends itself to phonic methods. The degree of orthographic transparency seems to be a more likely explanation for the differences in performance, especially in view of the confirmatory work of Oney et al (1984), and others. Indeed, there is now a body of evidence that demonstrates that young readers of Italian, Spanish, Turkish, Greek and German, which all have greater orthographic transparency than English, have fewer problems decoding written words than English-speaking children.

Figure 2: Comparison of Reading Performance for Low Frequency Words for German and English Dyslexic and Reading Age Controls (based on Lenderl et al, 1997)
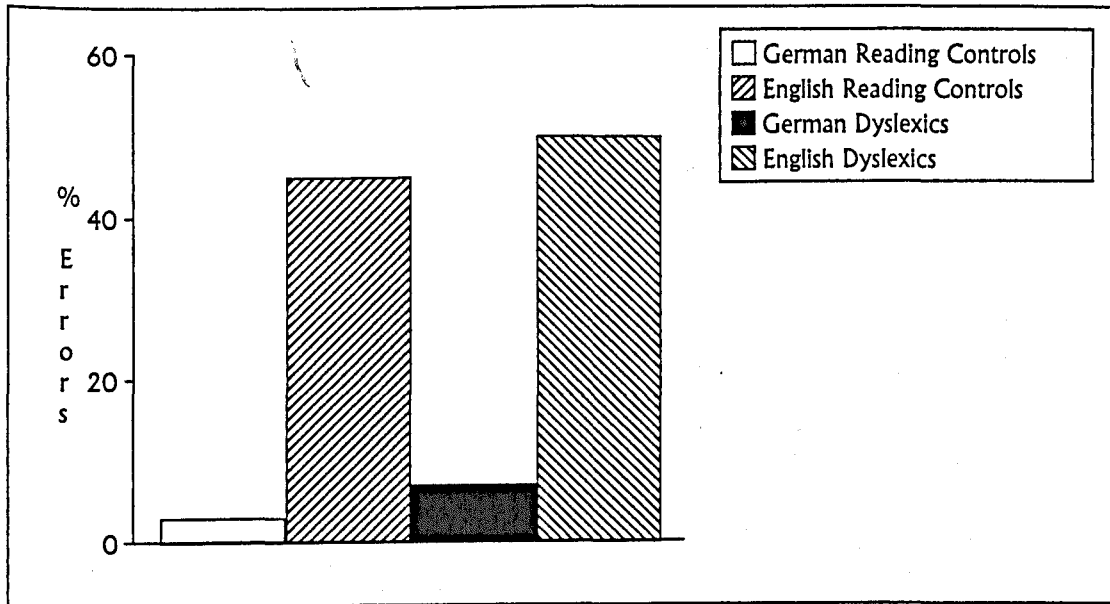
Some authors, such as McGuinness (1997), have suggested that dyslexia may not really exist, and that with a deep understanding of the complexities of the English language all children can become fluent readers. This is an extreme position. It is more likely that English-speaking children, who fail to acquire reading skills, fall into two distinct categories. There are those who would succeed in languages, other than English, that have greater orthographic consistency, and those who would still have problems even with perfect transparency. The methods advocated by McGuinness, which place considerable emphasis on a highly structured approach to reading, are effective with both groups. McGuinness acknowledges the different probabilities associated with the various written forms of the 44 English phonemes, and suggests methods which will give pupils confidence when learning to read.

Working from a similar perspective, Spencer (1998, 1999) has demonstrated that there are word factors, independent of individual pupil factors, which contribute to poor spelling of English words. Data from national SATs indicate three factors that are associated with ease of spelling: the frequency of the word in the English language; the number of letters in the word; and the "phoneticity" of the word.

Frequency of the word
The frequency of occurrence influences the spelling and reading of words because the more common a word, the more likely it is that the particular form will be internalized by the learner. The Lancaster-Oslo-Bergen corpus (Hofland and Johansson, 1982) provides an ordered list of the most common 7,000 words. The total number of occurrences of a word within the entire corpus (1,000,000 words) is also given, and this absolute frequency was used as a factor in the analysis.

3

Length of the word

This is a simple measure and was included because in the initial stages of spelling (and reading) 7 year-olds are still developing short term memory strategies, and any lapses in memory are likely to manifest themselves with longer words. Longer words also give more opportunities for errors.

Phoneticity of the word

Children learn at an early age that a variety of representations can be used for the same sound and, as SCAA recognized, the difficulty is less knowing the patterns than knowing which pattern to use in each individual word. In order to establish the range of representations of the phonemes that make up the English language, and the frequency of each representation, the 3,500 most common words from the LOB Corpus (Hofland and Johansson, 1982) were analyzed.

The phonic representation of each of the 3,500 words was determined from the Oxford English Dictionary (Second Edition, CD-ROM version) enabling the alphabetic representation of each phoneme to be determined for each word. With each phoneme coded, tables showing the various forms of representation for each phoneme were extracted. The average number of representations per phoneme is 5.95. Of course, some phonemes have relatively few forms of representation, while others have many more. Knowing the different representations of each phoneme allows the frequency of each representation for that phoneme to be calculated. The relative frequencies for the hard "o" sound are given below:

o	romance	50%
o+e	note	21%
ow	snow	16%
oa	boat	05%
ou	soul	03%
oe	toe	01%

Knowing the frequency of each form of representation for each phoneme allows an average phonetic value (average phoneticity) to be calculated for each word. In addition, the most unusual phonemic representations within each word can be identified. The most infrequent form of representation for each word is used to provide a value for the "tricky" phoneme in a word, and words can be arranged according to the value of this "tricky" phoneme.

Having identified the main word factors that contribute to spelling performance it is possible to predict the likely outcome for groups of words (Spencer, 1999). The most common 150 words in the English language were classified according to their average phoneticity and placed in one of three groups: high (>60%), medium (20-59%) and low (<20%). It was predicted that the words that were more phonetic would be spelt correctly by a larger number of pupils. This has proved to be the case. All year 4 pupils from a primary school, which performs at the average SATs level for English and Maths, were given the most common 150 words (extracted from the LOB corpus) to spell, at the rate of 20 words per day. Figure 3 shows the average performance associated with the three groups of words. As predicted, the highly phonetic group of words is correctly spelt by significantly more pupils (87%) than those words containing obscure representations of phonemes (65%). The mean values for average phoneticity and "tricky" phonemes are also shown for each of the three word groups.

Figure 3:   Percentage Correct Scores, Average Phonetic Word Values, and Average "Tricky" Phoneme Values for High, Medium and Low Phoneticity Word Groups.
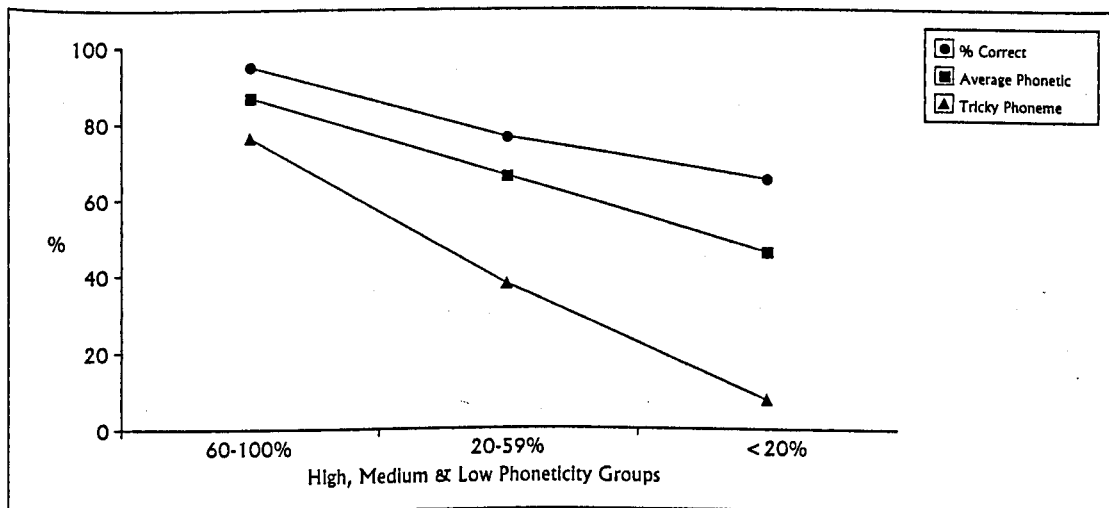
For statistical purposes the raw percentage scores, shown in Figure 3, are converted into logarithmic probability (logprob) values and these are shown in Figure 4. Essentially these are the same values as shown in Figure 3, but drawn on a different scale, which is quite convenient for illustrating the effect of extraneous letters in these words. The data shows that the three groups of words, which differ markedly in the average phoneticity of their words, have the same average number of phonemes (3 phonemes), but varying numbers of letters. Those words spelt correctly by the largest number of pupils have the same number of phonemes and letters: they show one-to-one mapping. As the average number of letters per phoneme increases, the words are spelt correctly by fewer pupils: performance deteriorates as one-to-many mapping increases. The differences observed in

Figure 4 are, statistically, highly significant (p = 0.000) for the measure of spelling difficulty (logprob values) and the number of letters. However, there is no significant difference (NS) in the number of phonemes in each group.

Figure 4:   Standardized Percentage Correct Scores (Logprob), Average Number of Phonemes, and Average Number of Letters for High, Medium and Low Phoneticity Word Groups.
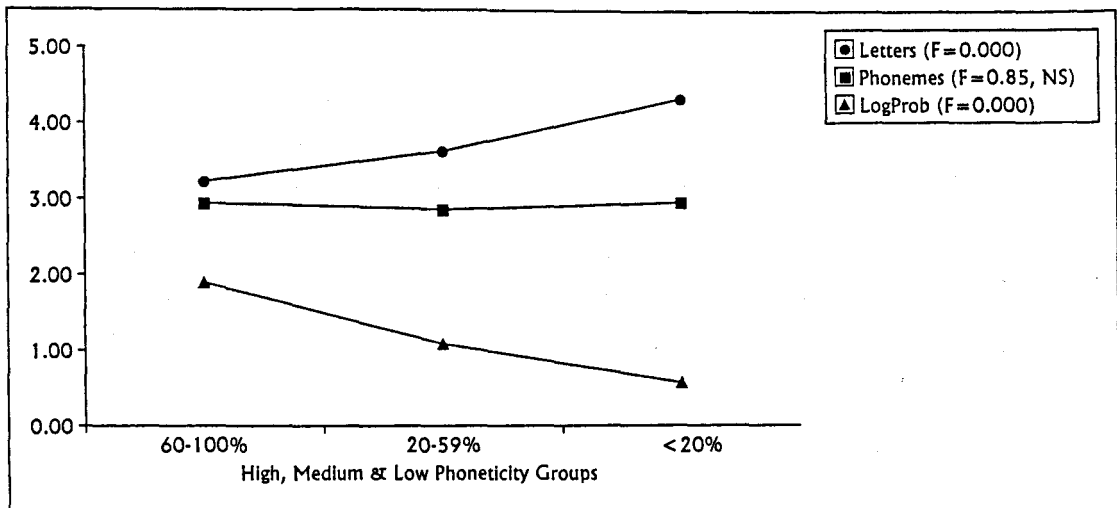
English: the dyslexic language

There is an accumulation of evidence that many English children have problems reading and writing because of the illogical system of orthography used for English. Much of the blame has been attributed to Samuel Johnson, who produced his Dictionary of the English Language in 1755. This is a little unfair because Johnson did acknowledge that his work was capable of many improvements and that the orthography which he recommended was controvertible, and the etymology uncertain and perhaps even erroneous. Unfortunately, there has been little change since that time.

The predictive model, developed from the SATs data and experimental studies, indicates which English words are easy or difficult to read and write. It comes as no surprise that the nine-year-old pupils in the study found words with one-to-one mapping (can, did, him) easy. Whereas, words with silent letters and obscure representations (know, should, though) are spelt correctly by less than 35% of pupils.

Compared to languages that are more consistent, English demands that children spend far more time learning the idiosyncrasies of the language. This is wasted time. A simplified, regular system of orthography enables children to master reading and writing within a year. English-speaking children could also be set up for life if their written language was

6

simplified, thereby removing the excessive memory burden associated with unusual spellings.

English is now the most important international language and this strengthens the case for revising and simplifying written English. Many European countries have already demonstrated that reform is possible, recognizing the fact that writing is a technology that requires regular updating. All technologies can vary in their effectiveness and inefficient ones are usually discarded in favor of later, more efficient, developments. This is not only associated with industrial technologies. The Roman system of numbering hindered the development of mathematics, until Arabic numerals, which incorporated the Indian idea of zero, introduced the concept of decimal fractions.

The writing technology used for the English language is failing to fulfil the demands of the modern world. Unless steps are taken to update this technology, English-speaking children will always be at a disadvantage when compared with those whose written languages are simpler. If we continue to use a written form that is difficult, we must accept the consequence that a huge proportion of our children will always fail.

References

Department for Education and Employment (1999) The Moser Report: Improving Literacy and Numeracy. London: DfEE.

Frith, U., Wimmer, H. and Landerl, K. (1998) 'Differences in phonological recoding in German- and English-speaking children', Scientific Studies of Reading, 2(1), pp.31-54.

Hofland, K. and Johansson, S. (1982) Word Frequencies in British and American English. Bergen: The Norwegian Computing Centre for the Humanities.

Landerl, K., Wimmer, H. and Frith, U. (1997) 'The impact of orthographic consistency on dyslexia: a German-English comparison', Cognition, 63, pp.315-334.

McGuinness, D. (1997) Why Children Can't Read: and what we can do about it. London: Penguin.

Oney, B. and Goldman, S.R. (1984) 'Recoding and comprehension skills in Turkish and English: effects of the regularity of grapheme-phoneme correspondences', Journal of Educational Psychology, 76, pp.447-568.

Spencer, Ken (1998) 'Predictive Models of Spelling Behaviour in 7- & 11-year-olds', Journal of the Simplified Spelling Society, 24(2), pp.23-28.

Spencer, Ken (1999) 'Predicting Spelling Behaviour in 7- & 11-year-olds', Reading, 33(2), in press.

6.   Spencer, K. A. (2001, in press) 'Differential effects of orthographic transparency on dyslexia: word reading difficulty for common English words', *Dyslexia*, 7(4). [6.Dyslexia2001.doc]

Differential Effects of Orthographic Transparency on Dyslexia:

Word reading difficulty for common English words

Ken Spencer

Institute for Learning

The University of Hull

UK

Address for correspondence:

Ken Spencer, IfL, The University of Hull, Cottingham Road, HULL HU6 7RX

Phone: +44 1482 465954 Fax: +44 1482 466133 Email: k.a.spencer@ifl.hull.ac.uk

Abstract

Orthographic transparency is increasingly being recognised as an important factor in determining the manifestation of dyslexic tendencies in individuals. Recent evidence has shown that normal English-speaking children have reading deficits in the range associated with same age dyslexic German-speaking children for less frequently used words, and English orthography has been identified as a contributing factor. Spencer (1999) has proposed a predictive model for English children's reading and spelling deficits, based on orthographic features. Two main factors have been identified: consistency of sound representation and inclusion of redundant letters in English words. Using new data for reading, a unified model will be discussed which incorporates measures of depth of orthographic transparency and relates these to recent positron emission tomography (PET) research. Implications for the relative frequency and severity of dyslexia in both deep and shallow orthographies is considered in the light of such studies.

Introduction

In a recent paper, Spencer (1999a) suggested that the individual characteristics of words could be used to predict the level of difficulty for spelling and reading. Most research investigating reading and spelling difficulties in children and adults has tended to look at the human characteristics to account for deficits. Another approach is to consider language and writing as technologies, and investigate the efficiency of written language in much the same way that the efficiency of other technologies is evaluated. Writing is so ubiquitous and embedded in daily life that it is often not recognised as being a technology and English is such a dominant international language that it is often forgotten that its written form differs from most other alphabetic languages. It has even been suggested that English is a dyslexic language (Spencer, 2000). This perspective views deep orthographies such as English and French as defective technologies that make extra demands on brain functioning (Paulesu et al, 2000, 2001) and incur additional costs, which in the case of written language will mean higher rates of illiteracy, and more energy and time spent in learning, especially for dyslexic students.

Spencer (1999a) suggested that this approach to analysing the effects of written English may have relevance for the teaching of the most common 150 words used in the English language, which represent about half of those we encounter on a daily basis (Hofland et al, 1982) and 60% of the words most frequently used in writing by 8 year-olds (Reid, 1989). This report presents the results of reading tests using these frequent words with 6 and 7 year-olds, and confirms predictions concerning the effects of word transparency on reading difficulties.

Methodology of the present study

Spelling difficulty for English words has been shown to be associated with several characteristics of the words: the number of phonemes and graphemes, the relative frequency of occurrence of the word, and a measure of the frequency of graphemic representational forms for a particular phoneme. This study uses regression methods to predict reading difficulty of words from the same factors. For each word the number of phonemes and graphemes were obtained from the New Oxford English Dictionary (2000), and the frequency of usage was obtained from several sources (Carroll et al, 1971; Hofland and Johansson, 1982; and Reid, 1989). The frequency of graphemic forms was also obtained from a variety of sources (Hanna et al, 1966; Berndt et al, 1987; Carney, 1994; Spencer, 1999b). The reading difficulty measure for each word was based on the number of children with normally developing reading skills who could read a particular word. The regression modelling was conducted using the word factors obtained from the

various sources to predict the level of reading difficulty of a word as measured by the number of pupils in a class who could read the word.

150 common words

The 150 words investigated in this study are shown in Table 1. Word frequency was originally obtained from the Lancaster/Oslo-Bergen (LOB) Corpus (Hofland and Johansson, 1982) of one million words, but frequencies from other corpuses were also obtained (Carroll, 1971; Reid, 1989) and used in the regression modelling.

In previous studies word length measured by the number of letters was found to be a significant factor in predicting spelling difficulty, but this rather underestimates the complexity of this component. In a perfect orthography there would be one letter for each sound and the number of sounds would exactly match the number of letters. This is not the case with English, although some languages do approach this degree of correspondence. English words usually have more letters than sounds, and it is this discrepancy (the phonetic difference) which is one of the main factors linked to spelling difficulty, although the number of sounds in a letter is also a component. In this reading study both the word length in terms of the number of phonemes (NPh) and the phonetic difference (PhD) are considered as individual factors, with NPh + PhD equalling the word length in graphemes (NGr).

The nature of English orthography means that individual sounds, of which there are 44, are associated with varying numbers of letter combinations. Previous research (Spencer, 1998, 1999b, & 2000) has demonstrated that the relative frequency of representations for a given sound is related to spelling difficulty for a word: the more frequently a grapheme is associated with a sound, the easier it is to spell.

In this and previous studies, one measure of the degree of orthographic transparency (phoneticity) for a sound or word is based on an analysis of the most common 7,000 English words. For each phoneme the frequency of each form of representation was calculated and expressed as a percentage of all instances of the phoneme. For example, the "hard-o" sound is usually represented by "o" [50%] or "o + magic e" [21%], but occasionally by "oa" [5%] or "oe" [1%]. When looking at word composition the phoneme which has the lowest phoneticity value has been identified as making the major contribution to spelling difficulty (more so than the average value for a word). This element in a word is the least transparent phoneme (LTP), and small LTP values reflect few instances of a grapheme representing the particular sound and tend to make a word difficult to spell. For the reading multiple regression analysis, these values are labelled LTP (S).

The phoneticity values can also be obtained from other analyses of the English language. Hanna et al (1966) undertook an extensive analysis of the English language with a view to using the phoneme-grapheme correspondences as cues to spelling improvement. Values from this study (LTP (H)) have also been entered in the regression analysis.

Berndt and her colleagues (1987) used the Hanna study as the basis for deriving probabilities for grapheme-to-phoneme correspondences in English. Berndt's team were motivated by the fact that the Hanna study was concerned with spelling rather than reading, commenting that "the probability that a particular grapheme is pronounced as a particular phoneme is not directly retrievable from information concerning the frequency with which a particular phoneme is spelled as a particular grapheme." The values derived from this study are labelled LTP (B) in the correlation analysis.

Carney (1994) provides two values for graphemes associated with each English phoneme. Lexical frequencies have been calculated in a similar way to Spencer (1999b) and Hanna et al (1966), but his text frequency takes account of varying frequencies of words in written texts (see Carney, 1994, for a full discussion). Two values are given for the LTP factors in the correlation analysis: LTP (CL) for the lexical values, and LTP (CT) for text values.

Data collection and analysis

The reading data were collected, during March 2001, for all pupils in years 1 and 2 in an urban Hull school primary school, which performs at average national levels in English, maths and science. Details of the ages, reading quotients and reading scores for each year group are shown in Table 2. Reading quotients (obtained from the schools records) were based on the administration of several standardised reading tests that are widely used in UK schools.

The 150 words were randomly assigned to five lists of 30 words, which were administered in one session. Children were asked to read each word in the list and there was no time limit for the test. The pupil responses were marked as incorrect if the pupil could not read the word, there were no partial scores recorded. Each of the 150 words received a score out of 28 (the number of pupils attempting to read each word) and this score was converted to a probability value for use in the regression analysis. The following formula was used:

$$\text{probability value (PV)} = \text{Log}_{10}(\text{probability right/probability wrong})$$

Percentage correct scores for each word in the present study were converted to probability values, using the above formula.

Results of the reading test

The 150 words used in this study are from a very different frequency range when compared with the previous studies spelling studies (Spencer, 1998, 1999a, 1999b, 2000). For example, only 8% of words from earlier studies fell within the frequency range of the most common 150 words (588 to 67,727 occurrences per million in the LOB corpus). However, the same underlying factors also controlled the spelling difficulty for these common words (Spencer, 2002, in press). These factors are also found to predict a large proportion of the variance in reading.

The analysis of the data was undertaken with the Statistical Package for the Social Sciences (SPSS, version 9). To determine how the various factors contributed to reading difficulty, regression methods were applied. These use the presence of an association between two variables to predict the values of one from those of another. The regression analysis of the data predicted the reading behaviour of pupils on the basis of four factors: frequency, number of phonemes in a word, the difference between the number of phonemes and graphemes (PhD), and phoneticity (LTP value). Tables 3a-d show the correlations between the reading probability values for Years 1 and 2 and the various factors, with individual word values obtained for frequency and phoneticity (LTP) from the previously cited sources. For the LTP values it is interesting to note that the set specifically designed to relate to reading as opposed to spelling, derived from Berndt et al's data, shows the lowest correlation, and the highest correlation is obtained from the spelling study from which Berndt's data is calculated (Hanna et al, 1966).

When all the values derived from the 5 LTP sources and 4 frequency sources, as well as the NPh and PhD values, are entered in a step-wise multiple regression, slightly different models emerge accounting for 58% (Y1) and 52% (Y2) of the variability in reading. The PhD factor dominates in Y1 and LTP(H) in Y2, although all models also include significant values for the number of phonemes and frequency (FREQ R8). The regression coefficients for LTP(H), PhD, NPh and FREQ(R8) are given in Table 4.

The relative influence of the word characteristics on reading difficulty can be seen in Figures 1a-c, with the 150 words divided into 2 sub-groups. The "easy" words are those read by all the children in Y2 (N=31); and a similar number (N=27) of "difficult" words at the bottom of the scale, read by less than 51% (mean 40%; SD 10%) of the Y2 children.

Figure 1a illustrates how the mean scores for the 2 groups of words fall from 71% (easy) to 14% (difficult) in Y1 and from 100% (easy) to 40% (difficult) in Y2; Figure 1b illustrates how the mean value for the strong LTP phoneticity factor also falls, from 55% to 12%, as fewer pupils read the words correctly. The means for the weaker word frequency factor also fall as reading performance drops.

A similar perspective is presented in Figure 1c. As words become more difficult to read they tend to have more phonemes, but the small change reflects the weakness of the factor. The strong factor is the phonetic difference (PhD). The easy words tend to show almost one-to-one mapping, with few additional letters (PhD=0); the difficult words move away from the alphabetic principle and tend to have two additional letters (PhD=2; i.e. a three phoneme word would be represented by five letters, as the graphemic length illustrates).

The results in an international perspective

The data from this and previous studies suggest that the make-up of English words, across a wide frequency range, contribute to spelling and reading difficulties for primary school pupils. Clearly, the suggestion is that English spelling hinders children's progress in spelling and reading, and the corollary is that reading and writing in other languages that are more regular will not be inhibited to the same extent. There is an accumulating body of research data which supports this thesis (Turkish: Oney and Goldman,1984; Italian: Cossu et al, 1995; German: Landerl et al, 1997).

Paulesu et al (2000) confirmed the effects that irregular orthographies have on speed of processing. When controlling for reaction time, articulation speed, naming speed and verbal fluency, Italian students were faster at both word and non-word reading, even when non-words were derived from English. The PET scans indicated that the pronunciation of a stimulus in English involves access to more regions of the brain than Italian and that this requires both time and resources. They conclude that 'reading in Italian can proceed more efficiently because of the consistent mapping between individual letter sounds and whole-word sound'.

A further PET scan study (Paulesu, 2001) demonstrates the biological unity of dyslexia across deep (English and French) and shallow (Italian) orthographies. What is particularly striking is that the dyslexic group showed significantly less activity than the normal group in a large area of the left hemisphere (middle, inferior and superior temporal gyri) that included the area previously identified as being associated with reading in a deep orthography (English). In addition, the reduced area of left hemisphere activation in the dyslexic groups still shows activation in the left planum temporale (at the temperoparietal junction), shown to be more active in the shallow orthography with normal Italian subjects. In other words, dyslexics show deficiencies in areas associated with deep language processing, but retain functionality in areas associated with shallow languages. Figure 1d illustrates how the characteristics of English words can be mapped on to the PET scan data. The "easy" words, which are more consistent in their spelling, require only those areas associated with shallow language processing

and should present few problems even for dyslexic children. The "difficult" words, however, require activation of areas not found to be functioning in the dyslexic subjects.

The results of this and previous studies also suggest that English must be seen as an exceptional variation on the alphabetic language theme. Sampson (1985) describes the English script as 'a compromise between the phonographic and logographic principles - somewhat akin, in fact, to Japanese script...' and this perspective indicates why there is a tension between methods for teaching spelling and reading of English. Phonic methods are clearly suited to transparent orthographies such as Turkish, Italian and German, but a language that also exhibits logographic tendencies will require more than a purely phonetic approach. Transparent orthographies are very efficient because they do not make heavy demands on memory and require a much more limited activation of brain regions, making them more accessible to dyslexic children; deeper orthographies being more memory dependent and requiring greater activation of the brain may actually prevent dyslexic children from achieving reading fluency.

References

Berndt, R.T., Reggia, J.A. and Mitchum, C.C. (1987) 'Empirically derived probabilities for grapheme-to-phoneme correspondences in English', Behaviour Research Methods, Instruments, & Computers, 19(1), 1-9.

Carney, E. (1994) A Survey of English Spelling. London: Routledge.

Carroll, J.B., Davies, P. and Richman, B. (1971) Word Frequency Book. New York: American Heritage Publishing Co.

Cossu, G., Gugliotta, M. and Marshall, J.C. (1995) 'Acquisition of reading and written spelling in a transparent orthography: two non parallel processes', Reading and Writing: An Interdisciplinary Journal, 7: 9-22.

Hanna, P.R., Hanna, J.S., Hodges, R.E. and Rudforf, E.H. (1966) Phoneme-grapheme correspondences as cues to spelling improvement. Washington, DC: U.S. Department of Health, Education, and Welfare.

Hofland, K. & Johansson, S, (1982) Word Frequencies in British and American English. Bergen: The Norwegian Computing Centre for the Humanities.

Landerl, K., Wimmer, F. & Frith, U. (1997) 'The impact of orthographic consistency on dyslexia: A German-English comparison', Cognition 63: 315-334.

Oney, B. & Goldman, S.R. (1984) 'Decoding and comprehension skills in Turkish and English: effects of regularity of grapheme-phoneme correspondences', Journal of Educational Psychology 76(4): 557-568.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C.D., & Frith, U. (2000) 'A cultural effect on brain function', Nature Neuroscience 3(1): 91-96.

Paulesu, E., Demonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S.F., Cossu, G., Habib, M., Frith, C.D., & Frith, U. (2001) 'Dyslexia: Cultural Diversity and Biological Unity', Science, 291, 2165-2167.

Reid, D. (1989) Word for word: the top 2,000 words used by 7 and 8-year-olds. Learning Development Aids: Wisbech.

Sampson, J. (1985) Writing systems, Stanford University Press: Stanford

Spencer, K.A. (1998). 'Predictive models of spelling behaviour in 7- and 11-year-olds', Journal of the Simplified Spelling Society 24(2): 23-27.

Spencer, K.A. (1999a) 'Predicting spelling behaviour in 7 to 11 year-olds', Reading 33(2): 72-77.

Spencer, K.A. (1999b) 'Predicting word-spelling difficulty in 7- to 11-year-olds', Journal of Research in Reading 22(3): 283-292.

Spencer, K.A. (2000). 'Is English a dyslexic language? Orthographic transparency and multilingual teaching', Dyslexia, 6: 152-162.

Spencer, K.A. (2002, in press) 'English Spelling and its contribution to illiteracy: word difficulty for common English words', Reading 36(1).

Table 1: 150 common English words in order of frequency according to the Lancaster/Oslo-Bergen corpus (Hofland & Johansson, 1982).

	FREQ (LOB)		FREQ (LOB)		FREQ (LOB)
the	68315	out	2035	little	895
of	35716	about	1895	down	885
and	27856	what	1872	did	884
to	26760	up	1860	how	873
in	21108	some	1851	last	870
that	11188	my	1813	between	867
is	10978	only	1813	your	853
was	10499	them	1699	just	836
it	10010	can	1675	still	823
for	9299	into	1657	work	819
he	8776	time	1654	see	814
as	7337	than	1646	know	798
with	7197	could	1614	make	796
be	7186	me	1554	might	781
on	7027	two	1549	because	777
his	6266	then	1546	through	773
at	6043	other	1533	same	768
by	5796	its	1516	own	751
had	5391	these	1501	long	750
this	5287	now	1489	get	735
not	5142	do	1478	here	717
but	4956	may	1447	go	714
from	4686	any	1416	great	702
have	4592	made	1294	three	697
are	4544	first	1287	never	686
which	4467	should	1276	life	684
her	4030	over	1264	year	677
she	3912	very	1229	come	676
or	3781	our	1228	both	675
you	3590	like	1205	old	670
they	3579	new	1181	another	668
an	3467	must	1129	without	665
were	3400	such	1125	again	663
there	3180	after	1119	us	657
been	3116	man	1072	each	654
one	3088	much	1071	used	648
all	2940	before	1061	came	645
we	2926	most	1059	under	645
their	2808	where	1033	take	639
has	2802	many	1029	say	635
would	2682	well	1009	found	624
when	2540	even	999	though	623
if	2479	also	994	men	616
so	2413	being	961	thought	611
no	2393	those	957	day	600
will	2269	people	953	right	599
him	2258	way	941	went	594
who	2200	back	934	world	594
more	2183	too	923	while	590
said	2074	good	900	course	588

Table 2: Mean values & S.D.s for age, reading quotient and reading score for 150 common words for years Y1 and Y2

	Y1	Y2
Mean Age	5.53	6.42
SD	0.32	0.30
Mean Reading Score*	100.18	102.13
SD	11.02	12.21
Mean % Reading Score for 150 common words	44.88	76.28
SD	22.87	22.66
N	28	28

* Youngs PR

Table 3a: Correlations for reading difficulty scores (Y1, Y2) and least transparent phoneme values (see text for sources)

Correlations

.		Y1	Y2	LTP (S)	LTP (H)	LTP (B)	LTP (CLF)	LTP (CTF)
Y1	Pearson Correlation	1.000	.811**	.520**	.602**	.236**	.414**	.412**
	Sig. (2-tailed)	.	.000	.000	.000	.004	.000	.000
	N	150	150	150	135	147	142	142
Y2	Pearson Correlation	.811**	1.000	.458**	.595**	.207*	.394**	.362**
	Sig. (2-tailed)	.000	.	.000	.000	.012	.000	.000
	N	150	150	150	135	147	142	142
LTP (S)	Pearson Correlation	.520**	.458**	1.000	.900**	.355**	.833**	.848**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000
	N	150	150	150	135	147	142	142
LTP (H)	Pearson Correlation	.602**	.595**	.900**	1.000	.286**	.827**	.805**
	Sig. (2-tailed)	.000	.000	.000	.	.001	.000	.000
	N	135	135	135	135	135	132	132
LTP (B)	Pearson Correlation	.236**	.207*	.355**	.286**	1.000	.298**	.320**
	Sig. (2-tailed)	.004	.012	.000	.001	.	.000	.000
	N	147	147	147	135	147	141	141
LTP (CLF)	Pearson Correlation	.414**	.394**	.833**	.827**	.298**	1.000	.965**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000
	N	142	142	142	132	141	142	142
LTP (CTF)	Pearson Correlation	.412**	.362**	.848**	.805**	.320**	.965**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.
	N	142	142	142	132	141	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Y1	Year 1 reading difficulty score (probability value) for words
Y2	Year 2 reading difficulty score (probability value) for words
LTP (S)	Least Transparent Phoneme, based on Spencer, 1999b
LTP (H)	Least Transparent Phoneme, based on Hanna et al, 1966
LTP (B)	Least Transparent Phoneme, based on Berndt et al, 1987
LTP (CLF)	Least Transparent Phoneme, lexical frequency, based on Carney, 1994
LTP (CTF)	Least Transparent Phoneme, text frequency, based on Carney, 1994

Table 3b: Correlations for reading difficulty scores (Y1, Y2) and word frequency values (see text for sources)

Correlations

		Y1	Y2	FREQ (LOB)	FREQ (R7)	FREQ (R8)	FREQ (C)
Y1	Pearson Correlation	1.000	.811**	.313**	.382**	.395**	.304**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
	N	150	150	150	150	150	150
Y2	Pearson Correlation	.811**	1.000	.315**	.421**	.426**	.309**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000
	N	150	150	150	150	150	150
FREQ (LOB)	Pearson Correlation	.313**	.315**	1.000	.817**	.829**	.986**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000
	N	150	150	150	150	150	150
FREQ (R6)	Pearson Correlation	.382**	.421**	.817**	1.000	.994**	.827**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000
	N	150	150	150	150	150	150
FREQ (R8)	Pearson Correlation	.395**	.426**	.829**	.994**	1.000	.838**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000
	N	150	150	150	150	150	150
FREQ (C)	Pearson Correlation	.304**	.309**	.986**	.827**	.838**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.
	N	150	150	150	150	150	150

. Correlation is significant at the 0.01 level (2-tailed).

Y1   Year 1 reading difficulty score (probability value) for words
Y2   Year 2 reading difficulty score (probability value) for words

FREQ (LOB)   Word Frequency, based on Hofland & Johansson, 1982
FREQ (R6)   Word Frequency for 6 year-olds, based on Reid, 1989
FREQ (R8)   Word Frequency for 8 year-olds, based on Reid, 1989
FREQ (C)   Word Frequency, based on Carroll et al, 1971

Table 3c: Correlations for reading difficulty scores (Y1, Y2), number of graphemes (NGr), number of phonemes (NPh) and phonetic difference (PhD)

Correlations

		Y1	Y2	NGr	NPh	PhD
Y1	Pearson Correlation	1.000	.811**	-.669**	-.343**	-.588**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	N	150	150	150	150	150
Y2	Pearson Correlation	.811**	1.000	-.579**	-.271**	-.533**
	Sig. (2-tailed)	.000	.	.000	.001	.000
	N	150	150	150	150	150
NGr	Pearson Correlation	-.669**	-.579**	1.000	.673**	.732**
	Sig. (2-tailed)	.000	.000	.	.000	.000
	N	150	150	150	150	150
NPh	Pearson Correlation	-.343**	-.271**	.673**	1.000	-.010
	Sig. (2-tailed)	.000	.001	.000	.	.900
	N	150	150	150	150	150
PhD	Pearson Correlation	-.588**	-.533**	.732**	-.010	1.000
	Sig. (2-tailed)	.000	.000	.000	.900	.
	N	150	150	150	150	150

**. Correlation is significant at the 0.01 level (2-tailed).

Y1	Year 1 reading difficulty score (probability value) for words
Y2	Year 2 reading difficulty score (probability value) for words
NGr	Number of graphemes in word, based on New Oxford English Dictionary, 2000
NPh	Number of phonemes in word, based on New Oxford English Dictionary, 2000
PhD	Phonetic difference (NGr - NPh)

Table 3d: Correlations for reading difficulty scores (Y1, Y2) and factors in regression analysis: phonetic difference (PhD), number of phonemes (NPh), least transparent phoneme (LTP [H]) and word frequency (FREQ [R8])

Correlations

		Y1	Y2	PhD	NPh	LTP (H)	FREQ (R8)
Y1	Pearson Correlation	1.000	.811**	-.588**	-.343**	.539**	.395**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000
	N	150	150	150	150	150	150
Y2	Pearson Correlation	.811**	1.000	-.533**	-.271**	.545**	.426**
	Sig. (2-tailed)	.000	.	.000	.001	.000	.000
	N	150	150	150	150	150	150
PhD	Pearson Correlation	-.588**	-.533**	1.000	-.010	-.541**	-.217**
	Sig. (2-tailed)	.000	.000	.	.900	.000	.008
	N	150	150	150	150	150	150
NPh	Pearson Correlation	-.343**	-.271**	-.010	1.000	-.031	-.204*
	Sig. (2-tailed)	.000	.001	.900	.	.710	.012
	N	150	150	150	150	150	150
LTP (H)	Pearson Correlation	.539**	.545**	-.541**	-.031	1.000	.168*
	Sig. (2-tailed)	.000	.000	.000	.710	.	.040
	N	150	150	150	150	150	150
FREQ (R8)	Pearson Correlation	.395**	.426**	-.217**	-.204*	.168*	1.000
	Sig. (2-tailed)	.000	.000	.008	.012	.040	.
	N	150	150	150	150	150	150

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Y1	Year 1 reading difficulty score (probability value) for words
Y2	Year 2 reading difficulty score (probability value) for words
PhD	Phonetic difference (NGr - NPh)
NPh	Number of phonemes in word, based on New Oxford English Dictionary, 2000
LTP (H)	Least Transparent Phoneme, based on Hanna et al, 1966
FREQ (R8)	Word Frequency for 8 year-olds, based on Reid, 1989

Table 4: Regression analyses on word reading difficulty

	Standardized Beta weight	t	p	Unique variance explained
Year 1 (multiple R = 0.773)				
Least Transparent Phoneme (LTP[H])	0.32	4.64	<0.001	0.09
Phonetic Difference (PhD)	-0.41	-6.03	<0.001	0.12
Number of Phonemes (NPh)	-0.25	-4.3	<0.001	0.06
Frequency (FREQ[R8])	0.19	3.18	0.002	0.04
Year 2 (multiple R = 0.737)				
Least Transparent Phoneme (LTP[H])	0.37	5.08	<0.001	0.11
Phonetic Difference (PhD)	-0.29	-3.99	<0.001	0.06
Number of Phonemes (NPh)	-0.19	-3.12	0.002	0.03
Frequency (FREQ[R8])	0.26	4.24	<0.001	0.06

Figure 1a: Reading difficulty mean scores (Y1, Y2) for easy and difficult words
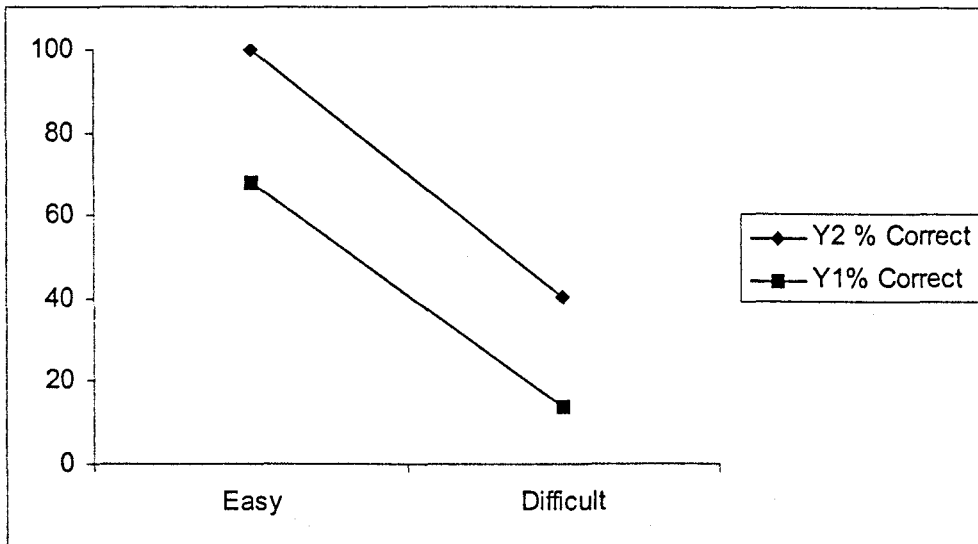
Figure 1b: Least transparent phoneme (LTP [H]) and frequency (FREQ [R8]) for easy and difficult words
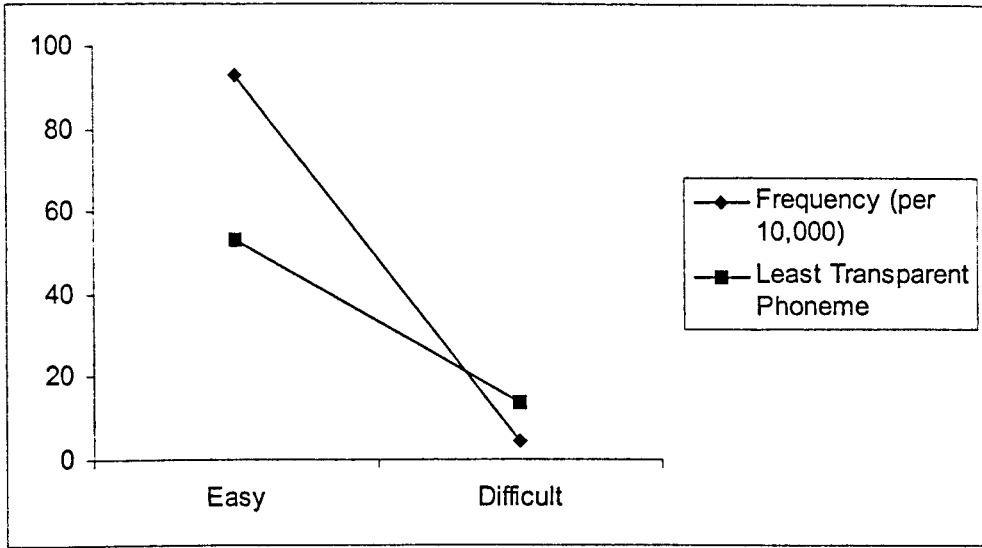
Figure 1c: Number of graphemes (NGr), phonetic difference (PhD)and number of phonemes (NPh) for easy and difficult words
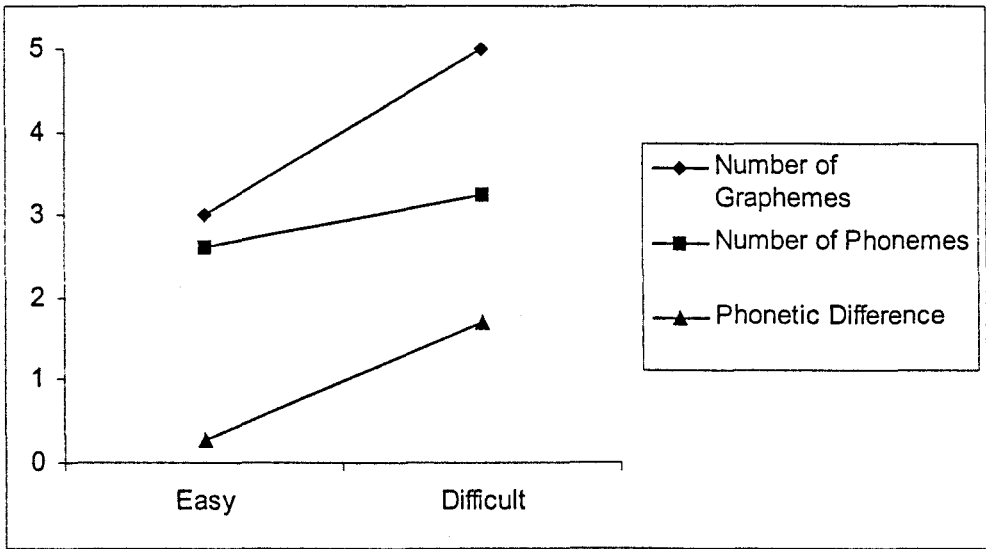
Figure 1d: Easy and difficult word characteristics mapped to reading centres
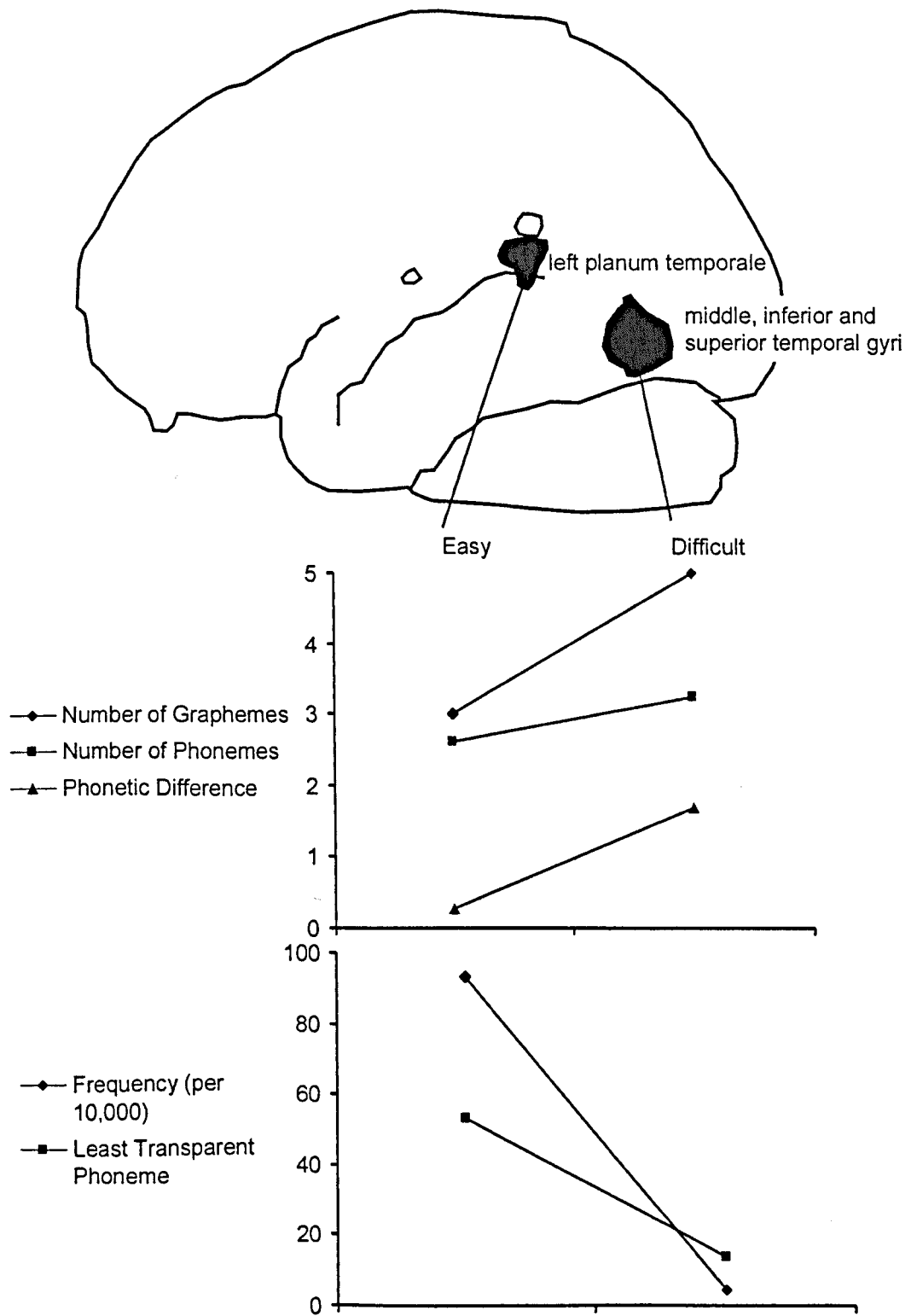
7.   Spencer, K. A. (2002, in press) 'English spelling and its contribution to literacy problems: word difficulty for common English words', *Reading*. [7.Reading2002.doc]

English spelling and its contribution to literacy problems:

Word difficulty for common English words

Ken Spencer

Institute for Learning

The University of Hull

UK

Address for correspondence:

Ken Spencer, IfL, The University of Hull, Cottingham Road, HULL HU6 7RX

Phone: +44 1482 465954 Fax: +44 1482 466133 Email: k.a.spencer@ifl.hull.ac.uk

Abstract

Writing like all technologies can be implemented in different ways with varying efficiencies. It is often not recognised that the written form of English is at variance with a majority of other alphabetic scripts and is increasingly being seen as a contributing factor to illiteracy in the English-speaking world.

The predictive spelling and reading model developed at Hull University indicates that the contributing factors making English words difficult for children to spell are: word frequency, length and phoneticity.

This paper, using the most frequent 150 words in the English language, confirms results from earlier studies and clearly links word characteristics to literacy failure. The impact of obscure forms of phoneme representation is shown to be particularly noticeable with less able children. This unnecessary burden placed on young children is not present in most European languages which have more regular written forms than English. The implications from this and previous studies are related to the teaching of English literacy skills.

Introduction

In a recent paper (Spencer, 1999a) I suggested that the individual characteristics of words could be used to predict the level of difficulty for spelling and, perhaps, reading. Most research investigating reading and spelling difficulties in children and adults has tended to look at the human characteristics to account for deficits. My approach was to turn the tables on the English language, and to use the words as my subjects. This approach relates to language and writing as technologies, and considers the efficiency of written language in much the same way that we would evaluate the efficiency of other technologies. Language is a powerful tool and its power is greatly extended by technologies which fix it into a permanent written form. However, writing is so ubiquitous and embedded in daily life that it is often not recognised as being a technology. Further, English has become such a dominant international language that it is often forgotten that its written form differs from most other alphabetic languages. It has even been suggested that English is a dyslexic language (Spencer, 2000). This perspective views written English as a defective technology. We have come to recognize the consequences of employing defective technologies: their use will incur additional costs, which in the case of written language will mean higher rates of illiteracy, and more energy and time spent in learning.

The 150 words investigated in this study represent 60% of the words used by 7 year-olds in their writing (Reid, 1989). Tables 1(a-d) give descriptive information about each word. Word frequency was obtained from the Lancaster/Oslo-Bergen Corpus (Hofland and Johansson, 1982) of one million words. In previous studies word length, measured by the number of letters, was found to be a significant factor in predicting spelling difficulty, but this rather underestimates the complexity of this component. In a perfect orthography there would be one letter for each sound and the number of sounds would exactly match the number of letters. This is not the case with English, although some languages do approach this degree of correspondence. English words usually have more letters than sounds, and it is this discrepancy (the phonetic difference) which is one of the main factors linked to spelling difficulty. The nature of English orthography means that individual sounds, of which there are 44, are associated with varying numbers of letter combinations, and it has been suggested that the ratio of sounds to letter combinations is 1:28 for the English language! I have demonstrated in previous research (Spencer, 1998, 1999b & 2000) that the relative frequency of representations for a given sound is related to spelling difficulty for a word: the more frequently a grapheme is associated with a sound, the easier it is to spell.

In this, and previous, studies the phoneticity, for a sound or word, is based on the most common 7,000 English words, which were analysed in terms of the different phonemes (20,000) and their associated forms of representation. For each phoneme the frequency of each form of representation was calculated

and expressed as a percentage of all instances of the phoneme. For example, the "hard-o" sound is usually represented by "o" [50%] or "o + magic e" [21%], but occasionally by "oa" [5%] or "oe" [1%]. When looking at word composition the phoneme which has the lowest phoneticity value has been identified as making a major contribution to spelling difficulty. This weakest element in a word is termed the "tricky" phoneme value, and small values reflect few instances of a grapheme, or grapheme configuration, representing the particular sound. Low "tricky" values tend to make a word difficult to spell.

Data collection

The spelling data were collected, during June 1999, for all pupils in an urban Hull primary school, which performs at average national levels in English, maths and science. The collected data of spelling performance spanned five year groupings (Y2-Y6; ages 7 to 11 years), for a total of 306 pupils. Details of the ages, reading quotients and spelling scores for each year group are shown in Table 2. Reading quotients (obtained from the schools records) were based on the administration of several standardised reading tests that are widely used in UK schools.

The 150 words were randomly assigned to five lists of 30 words, which were administered on five consecutive days. Class teachers administered the spelling tests, giving the word followed by an example sentence, and a further repetition of the word. There was no time limit for the test. Pupils wrote their answers on forms with word numbers clearly indicated. The marking for the test demanded complete mastery: pupil responses were marked as incorrect if the spelling was not perfect.

Results of the whole-school test

The 150 words used in this study are from a very different frequency range when compared with the previous studies. For example, only 8% of words from earlier studies fell within the frequency range of the most common 150 words (588 to 67,727 occurrences per million). However, when analysed, the same underlying factors controlled the spelling difficulty for these words.

The analysis of the data was undertaken with the Statistical Package for the Social Sciences (SPSS, version 9). To determine how the various factors contributed to spelling difficulty, regression methods were applied. These use the presence of an association between two variables to predict the values of one from those of another. The regression analysis of the data predicted the spelling behaviour of pupils on the basis of four factors: frequency, number of phonemes in a word, the difference between the number of phonemes and graphemes (phonetic difference), and phoneticity ("tricky" phoneme value).

All these factors are highly significant across the five years (Y2 to Y6), except for word frequency in Y4, and number of phonemes in Y6. Between 40% and 60% of the spelling variability was explained by the four word factors .

The inter-correlations between the spelling scores for the 5 year groups were highly significant [mean 0.8], and on this basis the entire data set was combined and a further analysis conducted. The four factors explained 63% of the variability in the spelling of the 150 words throughout the school. The influence the word characteristics have on spelling is not equal: two were strong factors [phonetic difference and "tricky" phoneme value]; frequency and number of phonemes had 50% less influence on spelling difficulty.

The relative influence of the word characteristics on spelling difficulty can be seen in Figures 1a & 1b. Here the 150 words are divided into 3 groups on the basis of the number of children in the school (Y2-6) spelling the words correctly. The easy words (Table 1a) range from 90% to 100% correct spelling; moderate words from 80% to 89% (Table 1b); and difficult words from 34% to 80% (Table 1c).

Figure 1a illustrates how the mean scores, across the 5 school years, for the 3 groups of words, fall from 96% (easy) to 66% (difficult); the mean values for the strong "tricky" phoneme factor also fall, from 55% to 12%, as fewer pupils spell the words correctly. The means for the weaker word frequency factor also fall as spelling performance drops.

A similar perspective is presented in Figure 1b. As words become more difficult they tend to have more phonemes (from 2.6 to 3.2), but this small change reflects the weakness of the factor; the strong factor is the phonetic difference (PhD). The easy words tend to show one-to-one mapping, with few additional letters (PhD=0); the difficult words move away from the alphabetic principle and tend to have two additional letters (PhD=2; i.e. a three phoneme word would be represented by five letters, as the graphemic length illustrates).

The results also give an indication of the damage that the strong factors, which reflect the degree to which English words deviate from the alphabetic principle, do to less able children. Figure 2 presents scores for the least able children (25%) in each year. For this group of children there is a sharp contrast between the easy words, with 65% of children spelling them correctly, and the difficult words, with only 10% of the least able pupils spelling them correctly. Although this is not a longitudinal study it does illustrate how much extra time may be needed for spelling proficiency to be raised for the more difficult words. For example, after two further years of schooling the least able 9 year-olds (Y4) have only reached the same level of proficiency on the moderate words that the 7 year-olds had attained on the easy words. Further, after 3 extra years, the least able 10 year-olds are still having problems with the difficult words,

and have still not reached the same level of proficiency that the 7 year-olds have reached for the easy words.

Spelling factors and reading difficulty

If the same underlying processes are involved in spelling and reading, the factors which control spelling difficulty should also influence reading difficulty. Preliminary results (Spencer and Xing, 2001) indicate that this is the case. Reading data for a class of 28 Y2 children in a similar Hull school were obtained and analysed with the same factors that were used in the spelling regression analysis. The results indicate that these four factors are all highly significant, but explain less of the variability in reading difficulty in the Y2 children (Y2 reading 42%; Y2 spelling 62%); however, another significant factor, the number of phonemes with phoneticity values less than 25% in a word, raises the power of the reading model (with 5 factors) to account for 50% of the variability. As with the spelling data the number of phonemes in a word and the frequency of the word were weak factors.

The results in an international perspective

The data from this and previous studies suggest that the make-up of English words, across a wide frequency range, contribute to spelling and reading difficulties for primary school pupils. Clearly, the suggestion is that English spelling hinders children's progress in spelling and reading, and the corollary is that reading and writing in other languages that are more regular will not be inhibited to the same extent.

When writing technologies are upgraded the effect on literacy can be quite profound. Turkish was written for a thousand years using the Arabic script, until the Latin alphabet was introduced in 1928. The national reform, which was completed in just 2 years, achieved nearly perfect one-to-one mapping of sounds and written symbols. After the reform it was claimed that children and adults learned to read and write fluently within a few months. Oney and Goldman (1984) confirmed the great advantage this form of orthography afforded over English. Working with American and Turkish first and third grade children they found that in reading tests Turkish first grade children out-performed their third grade American counterparts, and the difference was most apparent in longer words. They concluded that the regularity in letter-sound correspondence in the Turkish language seems to facilitate the acquisition of decoding skills and 'this is reflected in both speed and accuracy'.

Landerl et al (1997) found that reading times for English dyslexic children were double those for German dyslexic children, for three-syllable words and non-words. English children made three times as many errors when reading such words. They concluded that consistent orthographies, such as German,

pose much less of a problem for readers than inconsistent orthographies such as English, and that the differences in reading difficulties between English and German dyslexic children are due to differences in orthographic consistency. Indeed, the results demonstrated that for low frequency words German dyslexic children made fewer errors than normal English children of the same age.

Italian, another consistent orthography, also facilitates reading: the phoneme/grapheme ratio for English is 1:28, and compares unfavourably with 1:1.3 for Italian. Cossu et al (1995) found that young Italian readers achieved 92% accuracy on word reading tests after only 6 months schooling, whereas learning to read in English takes much longer. Paulesu et al (2000) confirmed the effects that irregular orthographies have on speed of processing. When controlling for reaction time, articulation speed, naming speed and verbal fluency, Italian students were faster at both word and non-word reading, even when non-words were derived from English. They found, using PET scans, that the pronunciation of a stimulus in English involves access to more regions of the brain than Italian and that this requires both time and resources. They conclude that 'reading in Italian can proceed more efficiently because of the consistent mapping between individual letter sounds and whole-word sound'.

Figure 3 illustrates the relative orthographic transparencies of Italian, German and English. The data is based on the 150 common words used in the spelling study, and equivalent words in Italian and German. The "tricky" value for words in the three languages, a key orthographic factor in spelling and reading, was calculated in relation to the spelling variability in the 150 words. Italian words are very consistent, with 80% in the easy category ("tricky" value 60-100%). This indicates why Italian children have so little difficulty in reading. German is not as consistent as Italian, but still has 60% of its words in this category, whereas for English only 30% of words can be classed in this way.

Discussion

The results of this research have implications for the teaching of literacy skills. They also confirm the probable outcome of employing a deficient technology, and actually raise the question: what is reading (and spelling)?

Deficient technologies waste time and energy, and that is precisely what is achieved in the English-speaking world. Figure 3 indicates that most pupils in the lowest ability group of year 2 (7 year-olds) can spell common words which conform to the alphabetic principle, but require an additional 3 years of schooling to reach the same level for the most difficult common words, which tend to deviate from the alphabetic principle. It appears that the consequence of using a very inconsistent orthography will

inevitably be very high rates of illiteracy, as we find for English-speaking countries in international comparisons (OECD, 1995).

There is undoubted resistance to the idea of spelling reform in the UK, and a cynical point of view suggests that this reluctance reflects a desire to prevent a large sector of society from achieving its full potential. There was great concern in the UK during the 19th century that schooling might foment revolution, and that literacy was a very powerful political tool. An orthography which was not transparent would afford two advantages in a society divided by class: it would make the process of learning to read and write more difficult because it would necessitate code-breaking skills and a knowledge of Latin and Greek; and it would reinforce a feeling of inadequacy and inferiority in individuals who had not the resources and time to master what should be an easy skill. The English script incorporates a social etiquette dimension to a much greater degree than most other alphabetic scripts, a device that was used to great effect in Richardson's 18th century novel 'Clarissa' (1751) in which the lower classes apply the rules to English words (especially doubling consonants) and are shown up for doing so. The aristocrats were word perfect and obviously familiar with the version of English which formed the basis for Johnson's dictionary (1755) which fixed English spelling for 250 years.

So, what is reading? Paulesu and his colleagues titled their paper "A cultural effect on brain function", indicating that what an Italian child does when looking at a page of written information is very different to what an English child is doing. The Italian child is reading; the English child is deciphering. Reading is easy and can be learned in a couple of months; deciphering a complex code requires much more time and energy.

When researchers study reading and writing, the word factors which control word difficulty must be taken into consideration. For example, controlling for word frequency in comparisons between English and German or Italian is not sufficient because English words may be infrequent but easy if they adhere to the alphabetic principle and vice versa (e.g. 'ban' and 'through'). In fact, perhaps research purporting to investigate reading or spelling should only be conducted in Italian or other transparent orthographies: deciphering and coding skills would be the exclusive preserve of English.

The results of this and previous studies also suggest that English must be seen as an exceptional variation on the alphabetic language theme. Sampson (1985) describes the English script as 'a compromise between the phonographic and logographic principles - somewhat akin, in fact, to Japanese script...' and this perspective indicates why there is a tension between methods for teaching spelling and reading of English. Phonic methods are clearly suited to transparent orthographies such as Turkish, Italian and German, but a language that also exhibits logographic tendencies will require more than a purely

phonetic approach. Transparent orthographies are very efficient because they do not make heavy demands on memory; logographic scripts, such as Chinese, are much more memory dependent. The challenge for educators is to find a middle way, between the extremes of purely phonic or look-and-say methods, for teaching English literacy skills. Dombey (1999) reflects this when suggesting that 'a balanced approach to reading recognises and encourages a close relationship between grapho-phonic learning, whole word learning and the experience of texts.'

The essential problem for early readers of English is the lack of self-teaching opportunities that push words into the lexical store which aids automaticity in reading (Share, 1995). In essence, most regular orthographies require a shorter period of rule learning than English and empower their pupils, enabling them to decode new words, even non-words, from an early age (Thorstad, 1991). This ability affords the opportunity for words to be assimilated quickly into the lexical system (Coltheart et al, 1993) leading to faster, expert reading, so essential for reading for meaning. Children working with transparent orthographies will be performing this action, with new words, as they apply the relatively straight-forward recoding rules. This means that in a shallow language as a child encounters a new word, which is not sufficiently embedded for the lexical system to evoke an immediate response, the easy decoding afforded by the orthography enables the link between text and word to build up a stronger representation, making it more likely to evoke an immediate response when next encountered. Such is not the case for English pupils. Even for the most common words in the language, which represent 50% of what we read, most have a difficult, low probability, grapheme. Because these words cannot be decoded by the pupil's application of simple rules they must be learned by association. For this to occur an external agency (such as a teacher, parent or paired-reader) must decode the word for the child, and this slows the rate at which children build up their lexical store. Without an external source being present to recode difficult words many children will be prevented from making progress if a text contains too many unfamiliar items. This is not the case for a child working with a shallow language: the meaning of unfamiliar items is simply unlocked by applying the rules that are usually learnt in a couple of months.

However, if self-teaching is not readily available because of the form of orthographic encoding, how can the lexical store be enlarged for children who are failing? The answer seems to be intensive one-to-one teaching. This appears to be the common feature of most methods which claim success at remediation, such as Clay's (1979) reading recovery, or the reading reflex method of McGuinness (1998). Torgesen (2001) has recently presented an interesting perspective on this when comparing two instructional programmes, both involving explicit instruction in phonemic decoding skills, stimulation of phonemic awareness, building a sight word vocabulary of high frequency words, and applications of these

skills to reading and understanding text. The interventions lasted eight weeks, and consisted of two 50 minute one-to-one sessions per day, five days per week. However, their instructional strategies were very different. One programme worked intensively to build strong phonemic awareness and a large amount of time (85%) was devoted to teaching children to accurately identify the number, order, and identity of sounds in words. In contrast the other method spent most of the time on reading connected text (50%), with the teacher providing careful error correction and discussion to help children generalise effective decoding strategies. Both methods afforded opportunities for guided practice of new skills, with very intensive instruction which provided systematic cueing of appropriate strategies in reading words or text, and explicit instruction in phonemic decoding. When rate of reading growth was compared both instructional strategies showed very similar benefits, removing about 40% of the children from special education in the year immediately following the 8 week intervention. These very different methods produced substantially the same results, and the probable causal factor was the intensive sessions with an expert decoder. A rough estimate suggests that one teacher could help 6 failing pupils per term.

This seems to be the price that has to be paid for our out-moded writing technology. The exact method of teaching may not be as important as the costly provision of individual or very small group tuition. If children fail to develop adequate decoding strategies early in their school life then intensive one-to-one teaching may be the only effective way of raising their literacy skills to the level at which they can begin to self-teach.

References

Clay, M. (1979) The early detection of reading difficulties : a diagnostic survey with recovery procedures. London : Heinemann Educational,

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993) 'Models of reading aloud: dual route and parallel-distributed-processing approaches', Psychological Review, 100, 589-608.

Cossu, G., Gugliotta, M. and Marshall, J.C. (1995) 'Acquisition of reading and written spelling in a transparent orthography: two non parallel processes', Reading and Writing: An Interdisciplinary Journal, 7: 9-22.

Dombey, H. (1999) 'Towards a balanced approach to phonics teaching', Reading 33(2): 52-58

Hofland, K. & Johansson, S. (1982) Word frequencies in British and American English. Bergen: The Norwegian Computing Centre for the Humanities.

Landerl, K., Wimmer, F. & Frith, U. (1997) 'The impact of orthographic consistency on dyslexia: A German-English comparison', Cognition 63: 315-334.

McGuinness, C. and McGuinness, G. (1998) Reading Reflex: the foolproof phono-graphix method for teaching your child to read. London: Penguin Books.

Oney, B. & Goldman, S.R. (1984) 'Decoding and comprehension skills in Turkish and English: effects of regularity of grapheme-phoneme correspondences', Journal of Educational Psychology 76(4): 557-568.

Organization for Economic Co-operation and Development and Statistics Canada (1995) Literacy, Economy and Society, Results of the International Adult Literacy Survey. Paris: OECD.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C.D., & Frith, U. (2000) 'A cultural effect on brain function', Nature Neuroscience 3(1): 91-96.

Reid, D. (1989) Word for word: the top 2,000 words used by 7 and 8-year-olds. Wisbech: Learning Development Aids.

Richardson, S. (1751) Clarissa. London. Full text version at: http://lion.chadwyck.co.uk/

Sampson, J. (1985) Writing systems. Stanford: Stanford University Press.

Share, D.L. (1995) 'Phonological recoding and self-teaching: sine qua non of reading acquisition', Cognition, 55: 151-218

Spencer, K.A. (1998). 'Predictive models of spelling behaviour in 7- and 11-year-olds', Journal of the Simplified Spelling Society 24(2): 23-27.

Spencer, K.A. (1999a) 'Predicting spelling behaviour in 7 to 11 year-olds', Reading 33(2): 72-77.

Spencer, K.A. (1999b) 'Predicting word-spelling difficulty in 7- to 11-year-olds', Journal of Research in Reading 22(3): 283-292.

Spencer, K.A. (2000) 'Is English a dyslexic language? Orthographic transparency and multilingual teaching', Dyslexia, 6: 152-162.

Spencer, K.A. & Xing, M. (2001) Differential Effects of Orthographic Transparency on Dyslexia. Paper presented at the 5th BDA International Conference, Dyslexia: At the Dawn of the New Century, 18-21 April 2001, University of York, UK.

Thorstad, G. (1991) 'The effect of orthography on the acquisition of literacy skills', British Journal of Psychology, 82: 527-537.

Torgesen, J.K. (2001) Theory and Practice of intervention: comparing outcomes from prevention and remediation studies, in A. Fawcett (Ed.) Dyslexia: theory and good practice, London: Whurr Publishers.

Table 1a: Easy Words

Word	Tricky Phoneme Value	Graphemic Length	Phonemic Length	Phonetic Difference	Frequency per million	% Correct Score
in	71	2	2	0	21108	100.0
is	69	2	2	0	10978	100.0
it	71	2	2	0	10010	99.6
man	95	3	3	0	1072	99.6
and	92	3	3	0	27856	99.6
the	17	3	2	1	68315	99.6
get	90	3	3	0	735	99.2
on	93	2	2	0	7027	98.7
he	21	2	2	0	8776	98.7
up	68	2	2	0	1860	98.7
can	70	3	3	0	1675	98.3
into	17	4	4	0	1657	98.3
not	93	3	3	0	5142	98.2
but	68	3	3	0	4956	97.9
at	96	2	2	0	6043	97.8
go	54	2	2	0	714	97.5
me	21	2	2	0	1554	97.0
an	97	2	2	0	3467	96.6
his	69	3	3	0	6266	96.6
we	21	2	2	0	2926	96.6
day	8	3	2	1	600	96.6
had	92	3	3	0	5391	96.2
men	90	3	3	0	616	96.2
if	71	2	2	0	2479	96.2
one	1	3	3	0	3088	95.7
will	10	4	3	1	2269	95.6
him	71	3	3	0	2258	95.3
as	69	2	2	0	7337	94.8
you	3	3	2	1	3590	94.4
from	84	4	4	0	4686	94.4
that	96	4	3	1	11188	94.0
no	54	2	2	0	2393	94.0
did	71	3	3	0	884	94.0
us	68	2	2	0	657	94.0
she	21	3	2	1	3912	93.7
old	54	3	3	0	670	93.7
my	10	2	2	0	1813	93.6
was	5	3	3	0	10499	93.6
has	69	3	3	0	2802	93.5
see	10	3	2	1	814	93.4
last	40	4	4	0	870	93.4
well	10	4	3	1	1009	92.9
good	70	4	3	1	900	92.9
be	21	2	2	0	7186	92.8
have	32	4	3	1	4592	92.8
its	63	3	3	0	1516	92.8
then	90	4	3	1	1546	92.4
do	8	2	2	0	1478	92.4
them	90	4	3	1	1699	91.5
this	63	4	3	1	5287	90.8
Mean	**55**	**3**	**3**	**0**	**5723**	**95.7**

Table 1b: Moderate Words

Word	Tricky Phoneme Value	Graphemic Length	Phonemic Length	Phonetic Difference	Frequency per million	% Correct Score
how	26	3	2	1	873	89.9
now	26	3	2	1	1489	89.9
three	10	5	3	2	697	89.9
her	45	3	2	1	4030	89.7
out	74	3	2	1	2035	89.7
came	48	4	3	1	645	89.4
all	6	3	2	1	2940	89.3
so	54	2	2	0	2413	89.3
went	64	4	4	0	594	89.3
long	53	4	3	1	750	89.1
said	1	4	3	1	2074	88.9
for	49	3	2	1	9299	88.6
may	8	3	2	1	1447	88.5
like	13	4	3	1	1205	88.4
come	0	4	3	1	676	88.0
with	64	4	3	1	7197	87.3
time	33	4	3	1	1654	87.1
are	1	3	1	2	4544	86.9
when	12	4	3	1	2540	86.9
make	13	4	3	1	796	86.7
two	1	3	2	1	1549	86.6
or	49	2	1	1	3781	86.5
back	5	4	3	1	934	86.3
down	26	4	3	1	885	86.3
to	8	2	2	0	26760	86.3
some	0	4	3	1	1851	85.9
they	1	4	2	2	3579	85.9
new	3	3	3	0	1181	85.0
than	97	4	3	1	1646	85.0
way	8	3	2	1	941	84.9
more	7	4	2	2	2183	84.5
what	5	4	3	1	1872	84.5
under	19	5	4	1	645	84.2
even	1	4	3	1	999	84.2
same	48	4	3	1	768	84.2
much	55	4	3	1	1071	84.1
made	48	4	3	1	1294	84.0
after	19	5	4	1	1119	83.7
very	64	4	4	0	1229	83.7
year	9	4	2	2	677	83.7
over	19	4	3	1	1264	83.6
most	54	4	4	0	1059	83.5
without	64	7	5	2	665	83.1
your	5	4	2	2	853	82.4
must	63	4	4	0	1129	82.4
just	25	4	4	0	836	81.9
take	13	4	3	1	639	81.9
say	8	3	2	1	635	81.5
life	33	4	3	1	684	80.2
Mean	28	4	3	1	2258	86.0

Table 1c: Difficult Words

Word	Tricky Phoneme Value	Graphemic Length	Phonemic Length	Phonetic Difference	Frequency per million	% Correct Score
by	10	2	2	0	5796	80.1
never	19	5	4	1	686	79.8
about	25	5	4	1	1895	79.5
found	74	5	4	1	624	78.5
of	1	2	2	0	35716	78.0
work	8	4	3	1	819	76.9
still	10	5	4	1	823	76.8
both	54	4	3	1	675	76.5
little	4	6	4	2	895	76.4
our	1	3	2	1	1228	76.4
here	6	4	2	2	717	75.5
such	55	4	3	1	1125	74.7
could	4	5	3	2	1614	74.6
each	12	4	2	2	654	73.8
first	14	5	4	1	1287	73.4
own	17	3	2	1	751	72.5
been	10	4	3	1	3116	72.3
people	0	6	4	2	953	72.2
being	9	5	4	1	961	71.7
right	9	5	3	2	599	71.3
who	1	3	2	1	2200	70.9
only	54	4	4	0	1813	70.9
because	0	7	5	2	777	70.8
would	4	5	3	2	2682	70.4
world	8	5	4	1	594	69.4
other	19	5	3	2	1533	69.2
these	9	5	3	2	1501	69.1
there	19	5	2	3	3180	68.8
might	9	5	3	2	781	66.8
also	6	4	4	0	994	66.7
again	1	5	4	1	663	65.2
before	7	6	4	2	1061	65.0
another	19	7	5	2	668	63.1
too	12	3	2	1	923	62.8
between	10	7	6	1	867	62.2
any	1	3	3	0	1416	62.0
many	1	4	4	0	1029	61.9
should	4	6	3	3	1276	61.4
used	7	4	4	0	648	61.4
where	12	5	2	3	1033	60.1
great	2	5	4	1	702	59.7
know	1	4	2	2	798	59.7
were	1	4	2	2	3400	57.9
those	21	5	3	2	957	57.5
which	12	5	3	2	4467	56.2
while	12	5	3	2	590	54.6
their	1	5	2	3	2808	50.2
through	1	7	3	4	773	47.9
thought	3	7	3	4	611	40.8
though	1	6	2	4	623	35.4
course	4	6	3	3	588	33.8
Mean	**12**	**5**	**3**	**2**	**1942**	**66.1**

Table 1d: Mean values for word groups

Word Group	Tricky Phoneme Value	Graphemic Length	Phonemic Length	Phonetic Difference	Frequency per million	% Correct Score
Easy	54.8	2.9	2.6	0.3	5723.3	95.7
Moderate	27.7	3.8	2.8	1.0	2257.7	86.0
Difficult	11.5	4.8	3.2	1.6	2019.9	66.1

Table 2: Mean values & S.D.s for age, reading quotient and spelling score for years Y2 to Y6

	Y2	Y3	Y4	Y5	Y6
· Mean Age	6.93	7.93	8.86	9.46	10.83
SD	0.32	0.30	0.33	0.94	0.30
Mean Reading Score*	100.08	102.38	95.23	96.26	96.66
SD	11.82	12.19	11.28	11.40	11.08
Mean % Score	64.02	77.60	86.07	88.91	92.87
SD	25.08	18.81	14.36	10.99	8.10
N	51	60	59	66	70

* Y2-3: Youngs PR; Y4-6 Cloze

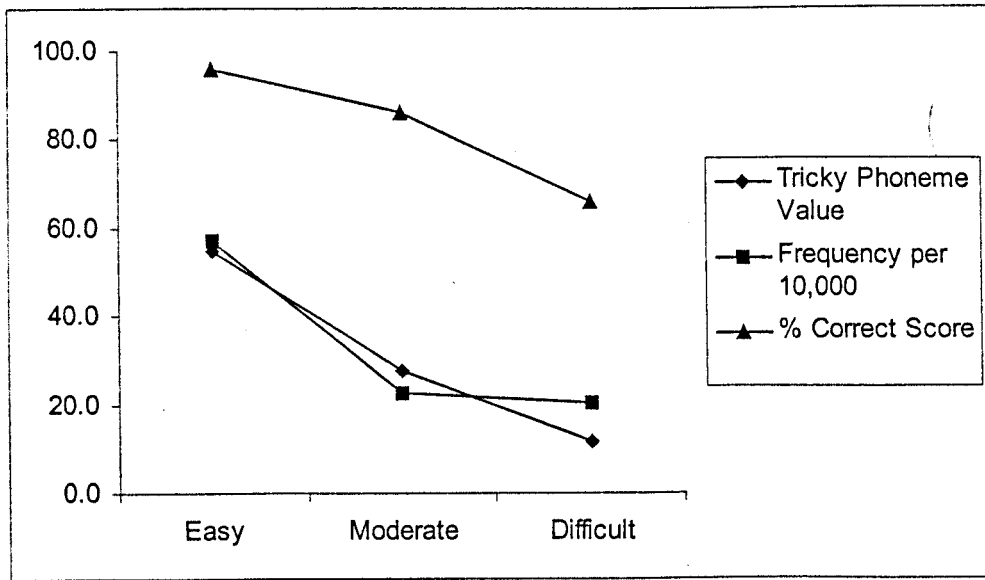**Figure 1a: Mean values (tricky phoneme, frequency and score) for word groups**

Figure 1b: Mean values (word length and phonetic difference) for word groups
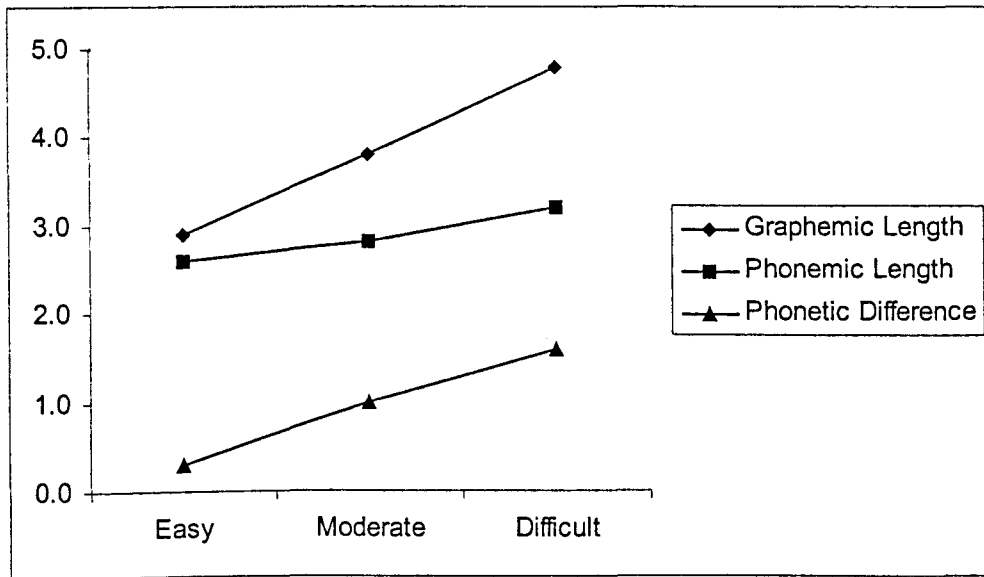
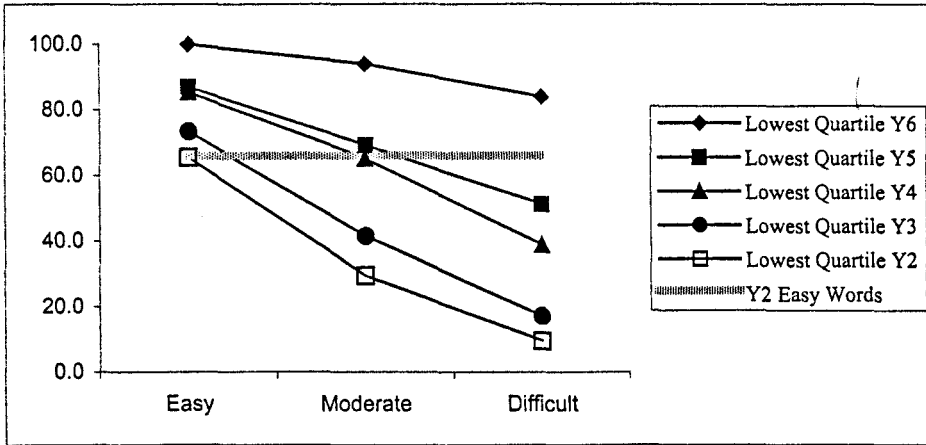Figure 2: Lowest quartile scores (Y2-Y6) for word groups

Figure 3: Mean "tricky" values for 150 words in English, German and Italian (in groups arranged from most [1] to least [10] phonetic)