Conducting Process-Product Studies: Some Considerations

Mark G. Borg.

Introduction

One of the major factors which determines the validity of research findings is undoubtedly the effectiveness of the decision-making process in setting up a design and a general methodology which are both rigorous and compatible with the aims of the study. Indeed, all the decisions made, be they major or minor ones, and related problems, vary from one study to the next depending largely on the nature and purpose of the exercise. The great majority of decisions required in designing a study are made in its early stages. It is usually the case, however, that in the course of a study other decisions would have to be taken. Since most decisions are interrelated, a change in one would precipitate a change in, or a reconsideration of, at least a second decision.

Vis-a-vis the above, it is the purpose of this paper to discuss some of the aspects and issues which should be considered in a research study of the relationship between teaching and attainment. Although most of the following arguments and considerations would be valid for such a study at the primary or secondary level of education, these would be more true at the former level.

The variables

In very broad terms, a study of this type would essentially entail an investigation into the effects, the products, of teaching. It would be concerned with the extent to which pupil performance in previously acquired cognitive abilities and skills improve as a result of their classroom experiences; whether or not the type of teaching enabled them to acquire new skills. In the literature, such investigations are referred to as "process-product" studies (e.g. Eggleston et al., 1976).

One of the first major decisions that the investigator has to make relates to the definition of the variables. What will "teaching", the main independent variable, mean in the context of the study? One can define and differentiate between teaching methods either on criteria as established by other studies in the field (e.g. Hilsum & Cane, 1971; Bennet, 1976; Galton et al., 1980), or by establishing, empirically or otherwise, the study's own criteria. The choice of the first option would seem to depend on whether or not these criteria would still be relevant in terms of age-range and the level of education as decided upon in the proposed study.

Will "attainment", the dependent variable, be taken to mean accomplishment in school subjects as measured by standardised or non-standardised attainment tests, teacher-based assessment, or a combination of any of these? The school subjects on which attainment measures will be taken have to be determined at the outset bearing in mind the commitments that participants (especially teachers) will have to make; and where applicable the availability of a research team, funds and resources.

There are several extraneous variables which might affect attainment. The age, social class, and the initial cognitive abilities of the pupils; the teacher's teaching experience, and the class and school environments are merely some of these variables. If the study is concerned with the casual aspects of the relationship between teaching and attainment, then it is imperative that systematic control is exercised on all such variables. Failing to do so would mean that any possible conclusions would be correlational in nature; like, for instance, the incidence of improvement in pupil performance and teaching style.

The Research Design

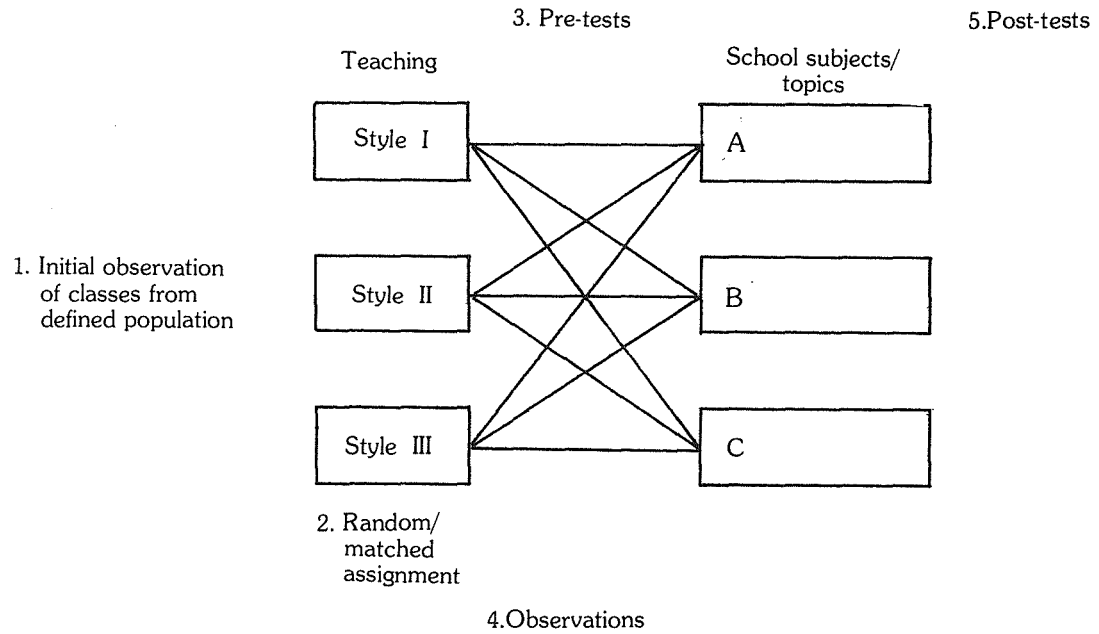
The study would lend itself to two main types of research designs: the experimental design and the longitudinal/survey design.

The experimental design

This design (see Fig. 1) enables the researcher to investigate cause-and-effect relationships between several styles of teaching and pupil attainment in any number of school subjects (using a factorial design; Lewis, 1968). It would necessarily involve the initial observation of classes from a defined population; from which a sample is drawn randomly or by matching subjects. Pupils are then assigned in a systematic way to the different teaching styles. They are tested before and after the experimental teaching period and mean gain test scores are compared to determine which teaching style caused the greatest improvement in performance.

Apart from problems of a methodological nature (e.g. the extensive control of extraneous variables), the main objection to this design is an

Figure 1: An Experimental Design



ethical one arising from exposing groups of children to experimental conditions which may make them underachieve. Also, the design presupposes that, all other things being equal, the effect of teaching styles is constant for all pupils irrespective of such factors as motivation and temperament.

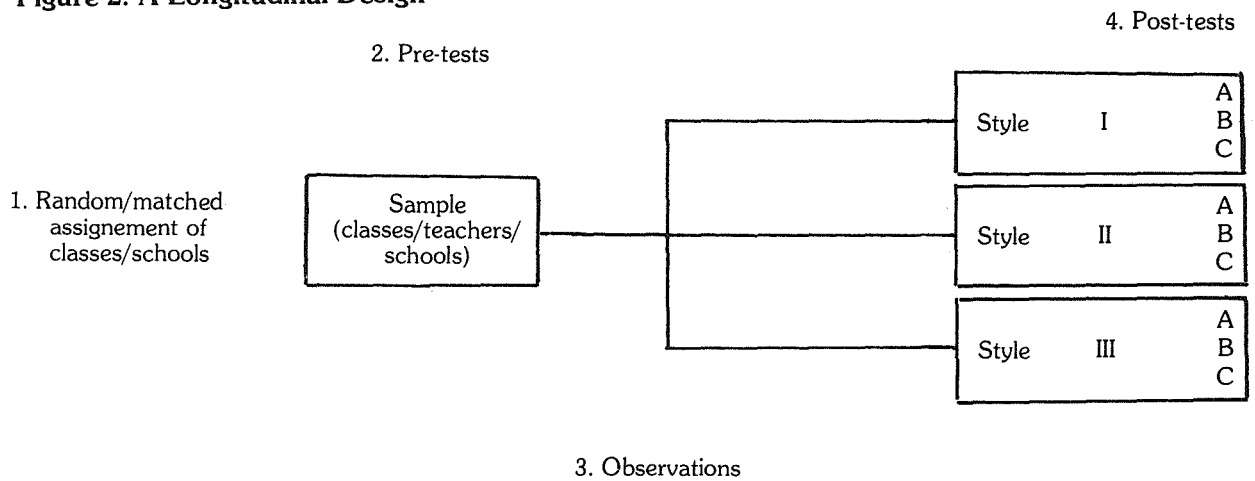
Most of these problems can be overcome by using a quasi-experimental research design (cf. Bennett, 1976). However, since in this situation the pupils and conditions are taken as they occur in the sample, a cause-and-effect relationship cannot be established (Vasta, 1979).

The longitudinal design

As is evident from studies which are based on this approach (e.g. the ORACLE studies: Galton et al, 1980; Galton & Simon, 1980), it involves the systematic collection of information in its ordinary settings. Depending on the rigorourness of sampling procedures, the information can be used to establish a causal or non-causal relationship. In the ORACLE studies, for instance, the sample was matched for several variables and the definition of teaching styles depended entirely on analysis of the interaction between teachers and pupils.

Each of these designs has its own particular problems. Both of them, however, require that the

Figure 2: A Longitudinal Design



pupils be exposed to the styles of teaching (in ordinary or experimental settings) for a substantial length of time (Carroll, 1963).

The Researcher's decision on which design is most appropriate for the study will determine several aspects of the general methodology, the most outstanding of which are: the sampling procedures; the methods of data collection; and the statistical analysis of the data. These will be discussed in turn.

The Sample

Sampling unit

This may be any of the following: pupil teacher, class, school, or catchment area. In view of the nature of the study it seems that the ideal sampling unit would be the class/teacher or the school. An important decision in this regard is which classes/schools would make up the population to which the study is addressed. Special classes/schools or children falling outside an established age range can be excluded; in which case the sampling unit is restricted to normal classes/schools or children of a certain age. Such restrictions could, of course, affect the representatives of the selected sample.

Sample size

This is a relatively minor decision. What is really crucial is not so much the actual size in terms of large numbers as much as having a large enough sample as to be representative of the population. It is indeed imperative that the sample should have representative characteristics in terms, for instance, of the social and home background of the pupils (e.g. Stallings, 1976), or on how teachers compare (in terms of age, sex, teaching experience, experience with the present age-group, etc.) with the national population.

In educational research, it is often the case that the size of the sample is dictated by the availability of subjects (Evans, 1968). Some schools or teachers, for instance, may refuse to participate. Then there is always the possibility of pupils and teachers absenting themselves from school (the absence rate, incidentally, can indicate attitudes towards teaching and schooling), or teachers who agree to participate but leave the sample at some stage of the study (especially so in schools where the teaching staff is unstable). One can, of course, decide on selecting a large sample for the purpose of containing such non-random sample attrition. But because sample loss is "non-random" it becomes increasingly difficult to ensure that the representative of the sample in terms of some is not adversely affected. When deciding upon the size of the sample, moreover, the researcher must bear in mind what types of comparisons of variables are envisaged so as to avoid the situation where not enough subjects (or data) are available; or if they were available, they are not representative of the same in terms of some variable (e.g. attainment), as was the case in the Lancaster studies (Bennet, 1976; pages 105-106).

Sampling procedures

Of all the aspects of the general methodology, sampling procedure is perhaps the most crucial mainly because it determines whether or not the sample is representative ofthe population; whether or not the eventual findings could be generalised.

If the sample can be considerably large, then the ideal sampling model would be stratified or dimentional sampling (Cohen & Manion, 1980). But as is often the case in educational research carried out in the school, the ideal is rarely possible. Therefore, it seems, that the matching model is the most practicable of the two. Pupils can be matched on several variables like sex, age, initial cognitive abilities, intellectual ability, and social class. Teachers, on sex, qualifications, teaching experience, their attitude towards the subject/s taught, and their teaching objectives. Schools can be matched on certain features like catchment area (urban - rural), and its socio-economic level, organizational and curricular patterns (as in the ORACLE studies), and goals, Since the school building can influence the method of teaching then the researcher should probably have to include this as one of the criteria for selecting schools. A further decision is whether or not to include all the classes in these schools or to randomly select a number of classes from each school. Irrespective of the level of education (primary, secondary, or both), the selection of classes would determine which teachers would be expected to participate. If the study is carried at the secondary level, however, then one has to take account of the possibility of "spill-over" from one teacher/subject to the next. A "good" science teacher, for instance, can influence pupil performance in say mathematics. Ideally, of course, related subjects should be included as part of the variables so as to avoid biased data. Moreover, apart from the problem of teacher reluctance to take part in the study, the teachers (within a catchment area/school) could be non-randomly distributed like, for instance, having highly efficient teachers employed in particular schools. The random selection of teachers by class should, however, average this out provided of course that these schools (or some of them) form part of the sample. The researcher must also be wary of the common characteristics of teachers who agree or decline to participate. Volunteer teachers, for instance, are likely to be "self-confident", relatively open teachers, almost all of whom may be superior to a non-volunteer sample on an unknown number of unidentified dimensions" (Berliner, 1976; page 10)

Data Collection

The study would essentially involve the collection of data on the styles/methods of classroom teaching and on the scholastic performance of the pupils.

Teaching styles

This data can be collected through two major procedures: (a) systematic classroom observation, and (b) the structured interview and/or self-administered questionnaire.

(a) Systematic observation

An important decision in this regard is whether the observation should focus on describing the actual types of teaching or to include also a description of what happens in the classroom (i.e. describing processes). Several writers have criticised the use of observation instruments which disregard contextual data like the subject taught, the use of teaching aids and available facilities, seating arrangements, and set-work (Hamilton & Delamont, 1974), and argued in favour of an adequate description of this context (Rosenshine, 1970; Rosenshine & Furst, 1973; Dunkin & Biddle, 1974). The categories making up the schedule, as Bennett & McNamara (1979) point out, "are usually derived in three ways, either from theory or previous research, by modification of existing schedules, or from prior observations in the classroom" (page 35). Although it is sometimes possible that such an instrument which suffices the purpose of the study already exists - a schedule whose reliability and validity are established - most researchers, as McIntrye (1980) maintains, attempt to develop their own schedules. This, of course, together with the piloting of the instrument in the classroom setting, would involve a great deal of time and effort in terms of deriving meaningful categories either by liasoning with the teachers or by visiting different types of schools recording as comprehensively as possible all the teachers' and pupils' activities. The investigator is then faced with the problem of deciding which features of teacher and pupil behaviours are relevant to the objectives of the study (Hilsum & Cane, 1971). Once these are established, two important sampling decisions have to be taken: what samples of behaviour will be recorded and what samples of time. The time unit adopted, for instance, is likely to affect the description gained (Bennett & McNamara, 1979). Too many categories of behaviour as well as too brief a sample of time, moreover, are bound to lower the reliability level of the schedule. One way of allievating these problems is by adopting a continuous recording of behaviours as in the Hilsum & Cane schedule. Moreover, as Hurwitz (1973) points out, the reliability and validity of the schedule should be determined from pilot test bases before it is put into actual use. The researcher then has to train/make provisions for training observers in the use of the schedule, especially so in the classroom situation. This is indeed a crucial part of the study (Hilsum & Cane, 1971). Before and during the actual observation sessions, high inter-observer reliability ratigs must be obtained so as to ensure that the observer/s is/are recording classroom events on the same criteria. Inevitably, this exercise could add to the "inconvenience" to teachers and pupils of having a person or persons recording classroom behaviours, no matter how unobtrusively so. Although some writers (e.g. Heyns & Lippitt, 1954; Kerlinger, 1964) state that the observer has little or no effect on the situation observed, the researcher must take in consideration the possible disruptive effects that the observer's presence could have on the pupils, and how this would affect the collected data. Indeed, as Medley & Mitzel (1963) and Samph (1976) point out, some observed behaviours do change even if others do not. The ideal situation would be to collect the observation data in a non-reactive fashion (Cheyne, 1979), but as McIntyre (1980) asserts, some reactions are to be expected and that "participants reactions to being observed are likely to depend on what they see as the function of the observation" (page 16). This, of course, gives rise to the problem of how far should participants (teachers and pupils) know about the nature of the observations. Persons et al. (1976), for instance, argue that teacher anxiety arising from the collection of data from his/her class can be minmized if the teacher knows what the observer is going to look for, and has access to the recorded data. On the other hand, if the teacher is so well informed one would be increasing the incidence of the Hawthorne effect and the resultant bias of the data. Besides all this, one must also bear in mind that what the observer has been "judging may not be an individual's true ability to perform but more the ability to respond to the challenge of an observation" (Samph, 1976; page 741).

The effort that the teacher and the pupils put into a lesson depends on several factors; not least the time of day and the day of the week. Moreover, the "performance" of the teacher is often influenced by his knowledge of, and attitudes towards, the subject being taught. This, of course, is especially the case in the primary level. Systematic sampling of classroom activities is therefore essential such that the collection of such biased data is avoided or at least averaged out.

(b) Structured interviews and/or self-administered questionnaires

Continued on Page 26

Continued from Page 15

Whichever type of questionnaire is used this should enable the researcher to collect data not only on the teaching methods but also on classroom and curricular organisation, discipline, testing, marking and so on (e.g. Bennett, 1976). In developing the questionnaire the researcher should ensure that the questions are not ambiguous (that is, that the instrument is reliable), through successive piloting in the field (Kerlinger, 1964), and even with knowledgeable individuals. Piloting could also reveal questions which should have been asked but were initially overlooked. Pre-coded answers would make the instrument more "efficient" in the sense that one knows beforehand what type of answers are to be expected, and pre-

coding would be amenable for transfer on computer cards for ease of storage and retrieval (Cohen & Manion, 1980). Moreover, questions which require the teacher to answer from only two distinct categories should be avoided as these tend to force the individual to answer one way or the other even though she does not agree.

The self-administered questionnaire has two major shortcomings. First, one can never make certain that all the questions are completed (e.g. Davie et al. 1966). Second, not all the questionnaires are completed and returned. Even when a relatively high percentage are returned (like the 80% of returns in the Bennett survey, 1976; page 42), one would never know whether or not the teachers who declined to or did return the questionnaire share some common charcteristic which renders the sub-sample of returns unrepresentative in terms of one or more of the parameters under study.

One way of overcoming these problems is to opt for a structured interview (Kerlinger, 1964). This, however, apart from pre-testing in the field, would also involve the training of interviewers and the difficulty of finding the most convenient time for teachers when they can be interviewed. This would most probably involve more commitment on the teachers' part in terms of giving up part of their free time. For this reason, and in order to avoid having teachers not completing all the self-administered questionnaire because they get bored half-way through, the time taken to complete it should be reasonable.

A major problem arising from the use of questionnaires is that what one claims to do in the classroom is not always what one actually does - how one conceives oneself to be is not always compatible with what one really is. There is, of course, little that one can do in terms of improving the instrument or its administration procedure. The concurrent use of the observation schedule and the questionnaire/s, however, is one way of overcoming this problem and establishing some validity of the teaching methods and class behaviours observed and recorded.

Achievement measures

The choice of tests used to give a measure of the influence of teaching on pupil achievement is of crucial importance. As Galton & Simon (1980) point out, there are two issues in particular which require consideration: "the relevance of the measures to what teachers and schools are trying to accomplish and the extent to which they are equally appropriate to a variety of teachers and schools, when comparisons between them are being made" (page 45). One way of overcoming the problem of choosing appropriate instruments would be, according to Cheyne (1979), to use different types of assessment procedures like teacher-based assessment, standardised and non-standardised tests on the basic skills.

The pupils would have to be tested/assessed at least twice: at the beginning of the study (to give a measure of their initial abilities), and at the end. There will be, of course, a number of pupils who would miss some of the tests. Preferably, provisions should be made to avoid this sample loss by having these pupils tested on another occasion (the earliest the better). As in the Bennett survey (1976), it would be advisable that the teachers themselves should administer the tests under normal classroom conditions "to obviate a test atmosphere and, hopefully, to reduce anxiety" (page 80). Detailed instructions should be provided so as to limit differences in the administration of the tests (including time and day when these should be held).

Each method of assessment has its particular advantages and disadvantages. Teacher-based assessment, for instance, involves a high level of subjectivity. On the other hand, this measure is a true reflection of the objectives of the teacher or the school. Besides, teacher-based assessment could give an indication of pupil achievement not only on the basic skills but also on a wider choice of subjects (e.g. Davie et al., 1966).

Non-standardised tests may also be said to reflect the teacher's/school's objectives when these are involved in their construction. However, these criterion referenced attainment tests give rise to a number of disadvantages like variations in scoring procedures, and their reliability and validity. In fact, they should be tested for their reliability and validity on appropriate groups of children before their full administration on the sample (e.g. Galton & Simon, 1980).

Standardised achievement tests would overcome all these disadvantages. What is more, they enable the researcher to compare the achievement of pupils from different schools as well as with results from other investigations. On the other hand, as Berliner (1976) and Dunkin (1976) point out, standardised tests are "inadequate" since they only measure a limited range of content and because they are not likely to be based on, or match, the content which has been taught: they lack content validity at the classroom level. Moreover, since these tests are usually group administered multiple choice type, young or lower socio-economic status children can be at a disadvantage. Indeed, as Dunkin (1976) points out, in studies of teacher effectiveness, the influence of intelligence and ethnicity on test performance should be removed or at least diminished. One way of overcoming these problems is to refine the test items: choosing items which are reactive to teaching (by consulting teachers), and choosing items that correlate weekly with a measure of general intelligence (Dunkin, 1976).

Statistical Analysis and Interpretation

It is not the purpose of this paper to delve into the choice of appropriate statistical techniques like cluster analysis, analysis of variance and co-variance, regression and other multivariage analysis of data. What is pertinent, however, is to consider some major statistical issues and their implications for the final interpretation of results.

Most of the above-mentioned techniques are based on the assumption of normality, the homogeneity of variance, the continuity and equal interval of measures. The researcher has to decide whether to use parametric statistics (cf. Kirk, 1968) or non-parametric statistics (cf. Siegel, 1956). Some writers (e.g. Warburton & Southgate, 1969) argue that the use of "crude" non-parametric techniques is not justified, even when the data is "fallible" (e.g. small amounts of data, biased data). Lewis (1967), moreover, asserts that the standard techniques are powerful and robust enough to hold in most cases, even when the above assumptions are violated.

Statistical inference would essentially be based on the statistical significance of results. The researcher can opt for a conservative interpretation of results by rejecting null hypotheses at the 0.01 level (e.g. Pringle et al., 1966) or 0.001 level instead of at the 0.05 level or lower. One must bear in mind, however, that in educational research a result may be statistically significant but not important (Warburton & Southgate, 1969). On the other hand, mean gain in attainment, for instance, under teaching style I may be only marginally greater than that under teaching styles II & III. Even though this result is not statistically significant it could have educational significance. Concomittant with this there is also the issue of generalizability of findings. An interpretation may be legitimate in the present sample but not so in the population.

References

1. Bennett, N. (1976) *Teaching Styles and Pupil Progress.* London: Open Books.
2. Bennett, N. and McNamara, D. (Eds.) (1979) *Focus on Teaching.* London: Longman.
3. Berliner, D.C. (1976) "Impediments to the study of teacher effectiveness" *Journal of Teacher education,* 27, (1), 5-13.
4. Carroll, J.B. (1963) "A model of school learning" *Teachers College Records,* 64, 723-733.
5. Cheyne, W.M. (1979) "Problems of design and analysis in the investigation of pre-school education" in Clark, M.M. and Cheyne, W.M. (Eds.) Studies in Pre-School Education, Britain: Hodder & Stoughton/S.C.E.R.
6. Cohen, L. and Manion, L. (1980) *Research Methods in Education* London: Croom Helm.
7. Davie, R., Butler, N. and Goldstein, H. (1972) *From Birth to Seven* London: Longman/N.C.B.
8. Dunkin, M.J. (1976) "Problems in the accumulation of process-product evidence in classroom research" *British Journal of Teacher Education* 2, (2), 175-187.
9. Dunkin, M.J. and Biddle, B.J. (1974) *The Study of Teaching* New York: Holt, Rinehart & Winston.
10. Eggleston, J.F., Galton, M. and Jones, M.E. (1976) *Processes and Products of Science Teaching* Britain: McMillan/Schools Council.
11. Evans, K.M. (1968) *Planning Small Scale Research* Revised ed. 1978. Britain: N.F.E.R.
12. Galton, M. and Simon, B. (Eds.) (1980) *Progress and Performance in the Primary Classroom* London: Routledge & Kegan Paul.
13. Galton, M., Simon, B. and Croll, P. (1980) *Inside the Primary Classroom* London: Routledge & Kegan Paul.
14. Hamilton, D. and Delamont, S. (1974) "Classroom research: a cautionary tale" *Research in Education,* 11, 1-15.
15. Heyns, R.W. and Lippitt, R. (1954) "Systematic observational techniques" in Lindzey, G. (Ed.) *Handbook of Social Psychology* Vol. I. Cambridge Massachusetts: Addison-Wesley.
16. Hilsum, S. and Cane, B.S. (1971) *The Teacher's Day* England & Wales: N.F.E.R.
17. Hurwitz, R.F. (1973) "The reliability and validity of descriptive-analytic systems for studying in classroom behaviour in Bennet, N. and McNamara, D. (Eds.) (1979) *Focus on Teaching* London: Longman.
18. Kerlinger, F.N. (1964) *Foundations of Behavioural Research* Second ed. 1973. Tokyo: Holt, Rinehart & Winston.
19. Kirk, R.E. (1968) *Experimental Design Procedures for the Behavioural Sciences* Belmont, California: Brooks/Cole.
20. Lewis, D.G. (1967) *Statistical Methods in Education* London: University of London Press.
21. Lewis, D.G. (1968) *Experimental Design in Education* London: University of London Press.
22. McIntyre, D.I. (1980) "Systematic observation of classroom activities" *Educational Analysis,* 2,(2), 3-30.
23. Medley, D.M. and Mitzel, H.E. (1963) "Measuring classroom behaviour by systematic observation" in Cage, N.L. (Ed.) *Handbook of Research on Teaching.* Chicago: Rand McNally.
24. Persons, W.S., Brassel, W.R. and Rollins, H.A. (1976) "A practical observation procedure for monitoring four behaviours relevant to classroom management" *Psychology in the Schools,* 13,(1), 64-71.
25. Pringle, M.L. Kellmer, Butler, N.R. and Davie, R. (1966) *11,000 Seven Year Olds.* London: Longman/N.C.B.
26. Rosehine, B. (1970) "Evaluation of Classroom Instruction" *Review of Educational Research,* 40, 279-300.
27. Rosenshine, B. and Furst, N.F. (1973) "The use of direct observation to study teaching" in Travers, R.M.W. (Ed.) *Second Handbook of Research on Teaching.* Chicago: Rand MacNally.
28. Samph, T. (1976) "Observer effects on teacher verbal classroom behaviour" *Journal of Educational Psychology* 68, (6), 736-741.
29. Siegel, S. (1956) *Non-Parametric Statistics for the Behavioural Sciences* New York: McGraw Hill.
30. Stallings, J.A. (1976) "How instructional processes relate to child outcomes in a national study of follow through" *Journal of Teacher Education,* 27,(1), 43-47.
31. Vasta, R. (1979) *Studying Children* San Francisco: W.H. Freeman.
32. Warburton, F.W. and Southgate, V. (1969) *I.T.A.; An Independent Evaluation* London: Murry & Chambers.