1-1-1993

Measurement of teacher attitude to direct instruction

Beth P. Hands
Edith Cowan University

Edith Cowan University

Copyright Warning

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

MEASUREMENT OF TEACHER ATTITUDE TO DIRECT INSTRUCTION

by

Beth Phyllis Hands B.Ed., B.Soc.Wk., Post Grad. Dip. Ed. St. (Dist).

A thesis submitted in Partial Fulfilment of the
Requirements for the Award of

Master of Education

at the Faculty of Education, Edith Cowan University.

Date of Submission: 27th May 1993

ABSTRACT

This study investigates teacher attitude towards a teaching innovation, Direct Instruction. In particular, it is concerned with what aspects of this teaching strategy affect a teacher's decision to implement the method and the relationship between experience with Direct Instruction and teacher attitude.

A repeated measures, pretest– posttest design was employed to assess participants' attitude before and after viewing a video demonstrating the teaching method, Direct Instruction. The 40 item Attitude towards Direct Instruction (ADI) scale comprised 30 items derived from a questionnaire designed by Proctor (1989) in addition to ten new items created especially for the study. The scoring function comprised the familiar Likert rating format using five response categories; Strongly Agree, Agree, Not Sure (NS), Disagree, and Strongly Disagree.

Data was collected from two sample groups. The first group, or 'change' sample, comprised 144 primary trained teachers and was used to assess the degree of attitude change between the pretest and posttest occasion. The second group, referred to as the 'control sample', comprised 275 third year education students at Edith Cowan University and was included in the study to assess the influence of the NS response category on the precision of the measuring instrument as well as forming part of the calibration sample.

Item analyses were conducted on the ADI scale using the extended model of Rasch, an important measurement model for assessing the psychometric properties of items with ordered categories. Due to the nature of the change sample, which included teachers with no experience with Direct Instruction, it

was necessary to include the NS category despite the knowledge that it can cause problems of a measurement nature. By assessing threshold order across all items it was shown that the NS category could be employed during the initial data collection but that any responses obtained had to be suppressed and changed to missing data in the data file if the subsequent analyses were to be effective. A final instrument comprising 19 items demonstrated sound psychometric properties with high reliability and person separation capabilities.

Graphical procedures, which took account of the special features of the measurement model, were used to investigate item bias (as part of the calibration of the instrument) as well as attitude change. In both cases, the plots provided an effective and simple interpretation of information as inconsistencies amongst both items and people were readily identified. The relationship between attitude change, as a result of the demonstration of Direct Instruction, and prior teacher experience with the method was also obtained from an examination of plot shifts across the calibration range.

Generally, teachers were more likely to have a negative attitude towards Direct Instruction if they had never taught the method. They perceived Direct Instruction to be too teacher oriented and highly structured. On the other hand, teachers were more likely to have a positive attitude towards Direct Instruction if they had had the opportunity to use it. They saw Direct Instruction as effective, easy to use, and versatile.

DECLARATION

I certify that this thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any institution of higher education; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signature:

Date: 27/5/93

ACKNOWLEDGEMENTS

The completion of this thesis has been an immense personal achievement, however I would never have achieved this task without the support and encouragement of many people.

Firstly, I would like to thank Barry Sheridan, my supervisor for the last year, for his unfailing support and encouragement. Without his belief in the value of this study, his wisdom and guidance, I may never have climbed the many mountains in my path.

Secondly, I would like to acknowledge Trish Formentin, Laurie Summers, and Sybe Jongeling for their guidance in the initial stages of this study.

Thanks must also go to the many teachers who freely gave their time and energy to complete the questionnaires.

Finally, to my husband, John and children; Kim, Jarrah, and Holly, thankyou for your patience and untiring support.

To all these people, thankyou.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

The general concern of this study is with teacher attitude towards a teaching strategy. More specifically, the problem is to study what aspects of a teaching method, particularly an innovation, affect a teacher's decision to introduce the method into their classroom.

Teacher attitudes are of great significance in all educational deliberations, particularly with respect to educational change. In the study of innovations in education, the importance of the teacher has often been overlooked. In an effort to understand change, the focus of research has usually been the innovation itself or the implementation process. When innovations are unsuccessful, the teachers are diagnosed as resistant to change. It seems that a major problem contributing to the failure of many innovations is the erroneous assumption that knowledge about the planned change is sufficient for it to occur. Many curriculum developers consider teachers to be passive consumers of their products. Teachers, on the other hand, believe themselves to have authority over what happens in their classroom. It is the teacher in the classroom who is the key to the success or failure of a teaching strategy. If the teacher does not consider the innovation to be worthwhile, then it is unlikely to be successfully implemented. How a teacher decides on which teaching strategies to use in the classroom should be of prime concern to promoters of new methods.

Recently, the focus of research has shifted to examining change from the perspective of the individual teacher. Several issues commented upon by researchers regarding teacher attitude towards an innovation are of interest to this study. First, it appears certain characteristics of an innovation affect a teacher's attitude. If an innovation is perceived by the teacher to have these

criteria, implementation of the change is more likely to be successful. Second, it is suggested that an innovation cannot be fully appreciated until after it has been implemented. Teachers who have had the opportunity to put an innovation into practice are more likely to express a more positive attitude towards it. In this study, a teaching strategy that has not achieved widespread acceptance amongst teachers, Direct Instruction, is used to investigate these aspects of innovation theory. In order to do this an instrument capable of measuring teacher attitude towards Direct Instruction is required.

The measurement of social phenomena is a fundamental prerequisite to most social research. It has been argued that social and psychological variables, such as attitude towards a teaching method, cannot be observed directly and therefore cannot be empirically tested (De Fleur & Westie, 1958; Tarter, 1969). However, it is argued in this study that the development of an instrument to measure attitude of teachers, for example, is a dynamic process involving the creation of a variable and the building of explanatory theory to explain the relationships observed between the phenomena to be measured. The process of measurement is therefore involved with the construction of variables through the employment of a measurement model. It then becomes possible to differentiate between objects or people according to the degree of certain properties they possess and in so doing test hypotheses with the aim of building explanatory theory. This is important for the development of sociological theory in general, and innovation theory, the substance of this study, in particular.

The literature regarding educational change is both voluminous and diverse. Gaynor and Du Vall (1977) have described it as extensive, fragmented, largely normative and predominantly unsubstantiated empirically. While several attempts have been made to develop a comprehensive educational change theory (Thompson, 1979; Waugh, 1983), a degree of consensus has not been

reached. Of the many examples of theoretical models of change that appear in the literature, the problem solving model (Lippitt, Watson, & Westley, 1958), the research–development–diffusion model (Clark & Guba, 1965), and the social interaction model (Mort, 1964) could be cited to illustrate this point. Such models have been criticised for failing to "encapture the complexity involved in implementing changes ... (for lacking) ... sufficient generality to encompass most planned educational changes" (Waugh, 1983, p.25), and for not having an overriding educational theory (Common & Egan, 1988). In order to develop a broad theory capable of explaining the implementation process, methodical, empirical testing of theoretical concepts is required.

The testing and development of theories that bridge the gap between general sociological theory and minor day to day hypotheses that cannot be generalised are important. These theories are referred to by Merton (1968) as "theories of the middle range." Such theory, suggests Merton (1968, p.51) guides inquiry

> by developing special theories from which to derive hypotheses that can be empirically investigated and by evolving, not suddenly revealing, a progresssively more general conceptual scheme that is adequate to consolidate groups of special theories.

Essentially, theories of the middle range are the means by which a young and developing discipline is able to systematise the knowledge gained from large quantities of empirical evidence and observation. If the social sciences are to advance as a science, Merton concludes, then theories of the middle range will play an important part in the building of strong theoretical structures upon sound theoretical bases.

Common and Egan (1988) suggest that one reason for the lack of theoretical consensus amongst researchers regarding innovations, centres on their failure to acknowledge teachers to be the key agent in the change process. This study attempts to test, in an empirical manner, aspects of the innovation process

specific to the teacher, and in so doing contribute to building the bridge between the anecdotal comments and observations of researchers regarding teacher attitudes to innovations and broader sociological theory.

Statement of Purpose

As stated, the purpose of this study is to investigate teacher attitude towards Direct Instruction. Of particular interest, is the degree to which certain characteristics of an innovation, and also prior experience with the method, contribute towards teacher attitude. This will be achieved in two stages. The first stage involves the calibration of an attitude measurement scale, guided by a conceptual framework based on innovation theory. The application of Rasch measurement will enable a greater understanding of what characteristics of innovations contribute towards a teacher's attitude. The second stage uses graphical procedures to assess the degree of attitude change as a result of a brief demonstration of Direct Instruction. The degree of attitude change will be compared to the amount of prior knowledge of the method. This stage investigates both the relationship between prior knowledge of Direct Instruction and teacher attitude, and the effect of a brief encounter with the method and teacher attitude.

Significance of the Study

This study contributes to the knowledge and understanding of teacher attitude towards innovations by formalising the link between theory and measurement. This is addressed by using the latest developments in psychometric theory to construct a sound attitude measure. Such measures are not widely used nor available to researchers or educators.

The concept of change is not always investigated in an accurate or appropriate manner. This study uses graphical or pictorial presentations which are easy to interpret, yet powerful in their ability to demonstrate change.

Finally, the effect of including the 'Not Sure' or 'Uncertain' category in association with the Likert type response categories was investigated for situations where participants may be unfamiliar with the technique or strategy being investigated.

Limitations

The time limitations of the study prevented both follow up evaluation of teacher attitude, for example, several weeks after the treatment, and more extensive exposure of teachers to the teaching strategy. It is reasonable to predict that a more comprehensive demonstration of the teaching method may have resulted in different teacher attitudes. Using a video clip to define the teaching method, Direct Instruction, had the advantages of consistency and convenience, however it also presented several problems. The video could not capture the 'reality' or dynamism of a classroom. Also, by attempting to present a technically correct and unbiased view of Direct Instruction, the result may have had a negative effect on teacher attitude.

Outline of the Study

Chapter Two contains a review of the literature leading to the development of the conceptual framework of the study in which the characteristics of an innovation that affect teacher attitude are identified. This guides the operationalisation of the attitude scale designed to measure a teaching strategy, Direct Instruction, the innovation selected as the focus of this study. The chapter concludes with a restatement of the problem which resolves into a number of research questions. Chapter Three is involved with a general discussion of measurement followed by

an assessment of the measurement model used to calibrate the attitude scale. As a result, knowledge of the variable teacher attitude towards Direct Instruction can be obtained.

The design of the study provides the substance of Chapter Four. A repeated measures, pretest–posttest design is employed to assess the degree of change in attitude resulting from exposure to a demonstration of a teaching innovation. The technique involves graphical procedures which take account of the special features of the measurement model.

Chapter Five and Chapter Six report the analysis of the data. In Chapter Five the calibration of the measuring instrument is described while an assessment of teacher attitude change as a result of a brief demonstration of Direct Instruction is undertaken in Chapter Six. Of particular interest, is the relationship between prior knowledge of the teaching strategy, Direct Instruction, and teacher attitude. The study concludes with Chapter Seven which presents a summary of the findings and a discussion of their implications for practice, theory, and future research.

CHAPTER TWO
LITERATURE REVIEW AND THEORETICAL FRAMEWORK

The purpose of this study is to investigate teacher attitude towards a teaching strategy. In particular, it is concerned with what aspects of a teaching method, particularly an innovation, affect a teacher's decision to introduce the method into their classroom. By exploring innovation theory, first in general, and then in relation to a specific innovation in teaching, Direct Instruction, the conceptual framework to guide the study will evolve. In the light of the theoretical implications applied to Direct Instruction, the chapter concludes with a statement of research questions to guide the study.

Innovation in Education

Change has always been a feature of our society, and many aspects of our lives are subject to constant change. Some innovations we see as improvements and embrace them gratefully, others are greeted with consternation or indignation. In the field of education, new methods of teaching, different subject matter, and alternative aims and objectives are constantly being developed. These changes are received by teachers in a variety of ways. One innovation may be seen as worthwhile, another, a waste of time and effort. Waldrop and Adams (1988) have noted that many promising projects have not shown the benefits intended by the developers because they have not been supported by teachers in schools.

The Nature of Innovations

Innovation refers to an object, idea or practice which is perceived to be new, and the process by which the object, idea or practice is adopted. This process is

ongoing and does not end with the development of a new teaching strategy. As Hall, Loucks, Rutherford and Newlove (1975, p.52) observed "change or innovation adoption is not accomplished in fact just because a decision maker has announced it." While each situation varies, many researchers (Berman & McLaughlin, 1978; Fullan, 1991; Huberman & Miles, 1985) see three broad phases to the change process. The first phase has been variously labelled as initiation, mobilisation or adoption. The steps leading to, and the decision to adopt or implement, an innovation are included in this phase. In the second phase, labelled implementation or initial use, the innovation is first put into place. This phase may last two or three years. The third and final phase called continuation, incorporation, routinisation, or institutionalisation, involves the decision to maintain or discard the innovation. If the process has been successful, the innovation would be included into the teacher's classroom routine.

Change at the classroom level is a highly personal experience. The concerns and frustrations of individual teachers are often more critical than the technological dimension of the proposed change. As Brown and McIntyre (1982, p.35) suggest, "the nature of the relationship between attitudes and the implementation will depend on the nature of the innovation itself." One reason for the ultimate success or failure of an innovation is related to a teacher's reaction to certain characteristics of the innovation.

Characteristics of Innovations

In the first phase of the innovation process, as described earlier, the decision is made by a teacher to adopt or reject a proposed change. The support and commitment made by the teacher in this initial phase affects what happens when the actual implementation of the change begins (Berman & McLaughlin, 1976).

Brown and McIntyre (1982) found that teachers exert almost total control over classroom innovations. Professional colleagues and administrative authorities were unable to exert any effective pressure on teachers to change their attitude towards new teaching strategies. Certain aspects of the new idea, object or process appear to be significant in determining its success or failure (Fullan, 1991; Hoyle, 1976). What these are and how they affect the teacher need to be explored further.

Considered in the light of teachers' responses to innovations, several characteristics appear to be significant. In 1977, Fullan and Pomfret felt that these could be adequately described as "the explicitness or plans for explicitness associated with the innovation and the complexity or degree and difficulty of change required by the innovation." (p.368) At the same time, Doyle and Ponder (1977) suggested that a 'practicality ethic' guided teachers' reactions to change proposals. The term 'practical' is an expression of a teacher's perception of the potential consequences of introducing an innovation to their classroom. Attributes of a proposal for change which would elicit the perception of practicality from teachers were classified by Doyle and Ponder (1977) as instrumentality, congruence, and cost. Innovations with these attributes would be perceived as practical and most likely to be introduced.

Similar attributes of innovation have been given alternative labels. For example, Rogers and Shoemaker (1971) identified six characteristics: complexity, compatibility, relative advantage, trialability, communicability, and status. Recently, Fullan (1991) focussed on three principal criteria which he considered teachers use to assess an educational innovation; need/evidence, procedural clarity, and personal costs and benefit. The importance of these three characteristics was also noticed by Mohlman, Coladarci, and Gage (1982) in a comparison of results of five teacher effectiveness experiments, and by Henry

(1989) who investigated three education innovations. The three characteristics of need, clarity and cost, emerge as critical aspects of an innovation which affect teacher attitude and will now be explored in detail.

Need

Need for an innovation refers to the extent to which the innovation addresses a perceived need. Emrik and Paterson (1978), Fullan (1991), and Rosenblum and Louis (1981) all found that the relevance and importance of the innovation to the teacher affected the implementation process. The teacher also considered the response of the students, their interest level, and evidence that the proposed change worked. Where the innovation focussed on a specific, identified need the chances of an enthusiastic and active implementation process was enhanced.

Teachers will not consider a proposed change of value if it conflicts with their personal philosophies. A number of studies have found that teacher attitudes had a strong effect on teacher receptivity to change. Mohlman, Coladarci, and Gage (1982) found evidence of teachers failing to appreciate innovations due to expressed philosophical disagreements. Waugh (1983, p.15) observed that "planned educational changes which involve teachers in a conflict with their basic educational attitudes and beliefs would not be well received by those same teachers." Changes which are compatible with a teacher's existing values, past experiences or present needs have a higher chance of successful implementation. Essentially, this is because the change does not require substantial alterations to current behaviour (Print, 1987). The perceived need for change emerges as an important aspect of the successful implementation of an innovation.

Clarity

The clarity of the innovation refers to the extent its essential features and procedures are made clear to teachers. As Crandall, Eiseman, and Louis (1986, p.41) see it, "educators want to know what the innovation is all about, how it works and what it means for them in terms of time, effort and energy." However, clarity of both goals and means of implementation is a constant problem in the innovation process.

The more complex the change, the greater the importance of clarity. Fullan (1991) found that problems relating to clarity occurred in many studies of significant change. Lack of clarity, especially, unclear goals and means of implementation, presents a major concern in the early stages of the innovation process. Doyle and Ponder (1977) described several instances of innovations failing due to inadequate explicitness and ambiguity resulting in frustration or confusion for the teacher. Fullan (1991, p.71) noted that "unclear and unspecified changes can cause great anxiety and frustration to those sincerely trying to implement them."

Innovations that are presented clearly, specifically, and easily translated into classroom practices are more likely to be implemented. Henry (1989) found many teachers preferred innovations to be highly prescriptive, clearly explaining what is required. Where innovations have been perceived by teachers as highly complex, the level of implementation is low (Crowther, 1972; Henry, 1989; Paul, 1977). The clarity of goals and procedures required of an innovation emerges as a second aspect relevant to the innovative process.

Cost

Cost refers to the effort required by teachers in time, energy, acquisition of new skill, and change of classroom routine in order to introduce the innovation. If the perceived advantages of the innovation are considered to outweigh the cost, the chances of successful implementation are high (Cavanagh & Styles,1983; Nicholls, 1983). While developing a Levels of Use indicator which allowed a determination of the extent to which an innovation was being implemented, Hall, Loucks, Rutherford and Newlove (1975) observed that teachers would not progress in the implementation process if a perceived advantage was not present. These advantages may relate to greater efficiency, cost–saving, saving of time, or, in particular, educational achievement. In an extensive review of teacher attitudes and attitude change, Stern and Keisler (1977) found much evidence to support this. They stated,

> where new teaching approaches are perceived as rewarding, the teacher's attitude is apt to be modified in a positive direction. These rewards need not be in the form of salary increases, fringe benefits or professional status. Far more significant for most teachers is assurance of improved student progress and more stimulating learning conditions. (p.74)

An innovation which offers a sense of mastery and accomplishment provides a powerful incentive for the teacher to put it into practice.

In the early stages of implementation, how the change affects a teacher personally is of greater influence than descriptions of goals and benefits of the programme. The introduction of a change into the classroom often involves high personal costs and unpredictable benefits to the teacher (Doyle & Ponder, 1977; Loucks & Hall, 1979). Teachers are ultimately responsible for what happens in their classroom and a natural reaction is to resist change. As Lortie (1975, p.212) states, "the teacher ethos is conservative, individualistic and focused on the present."

Another aspect of cost relates to the extent to which the innovation can be partially implemented prior to a total commitment by the teacher. If the innovation is able to be trialled, it is seen as a factor that reduces the cost. Teachers may feel more comfortable in testing the method, without feeling committed. This is particularly important with innovations that involve a large financial outlay or major rescheduling of classroom routines. Stephens (1974) found that teachers would adopt innovations if the reward structure of the school was made contingent on innovativeness. In this case, the notions of cost and reward included recognition and student enthusiasm as well as monetary remuneration. Personal costs and benefits of an innovation to the teacher emerges as a third aspect relevant to the innovative process.

Summary of Innovation Characteristics

Three characteristics of an innovation: need, clarity and cost, appear important in the implementation of an innovation. Fullan (1991) observed that these three aspects must be favourable at an early stage of the innovative process in order for implementation to proceed.

The three characteristics of an innovation as stated provide a theoretical basis for the conceptual framework of this study. As such, innovation theory of the middle ground, as described in Chapter One, will be investigated. The conceptual framework specifies aspects of innovations as implemented through a teaching strategy, Direct Instruction. In this way, a focus for the study is provided as the framework specifies a means of guiding the empirical inquiry. Attention is now directed to Direct Instruction as a specific innovation to be implemented.

Direct Instruction

As an innovation, Direct Instruction involves the use of new materials, teaching strategies and possible alteration of the teacher's beliefs. In terms of implementation, however, Direct Instruction has achieved limited success in regular classrooms. When observing Direct Instruction for the first time, a strong negative reaction by teachers has been noted and reported by several investigators. Gersten & Guskey (1985) commented that many teachers initially felt stifled by the approach. The perceived excessive structure and regimentation of Direct Instruction was claimed by Mathes and Proctor (1988) as the reason for the negative reaction by teachers. Fields (1986) has observed similar responses from teachers in Australia. Consequently, teachers have been reluctant to introduce the method into their classrooms. Why this should be so provides a basis for the importance of this study.

To understand this behaviour, the development of Direct instruction is explored followed by an examination of the characteristics of the strategy and why it can be claimed as a worthwhile innovation. In addition, Direct Instruction is investigated as an innovation in the context of the innovation theory discussed earlier.

Development of Direct Instruction

Direct Instruction evolved from the efforts of Siegfried Engelmann and Carl Bereiter to build effective teaching principles into programmes in an experimental preschool at the University of Illinois (Corr, 1984). After initial successful results, Engelmann and Becker (who had now joined him), were asked to develop a teaching strategy that could be trialled from pre-school through to Grade Three within Project Follow Through. In the U.S.A., during the period 1968 to 1976, Project Follow Through, an extension of the Headstart pre-school programme,

investigated promising but different approaches to the education of economically handicapped children aged six to nine years.

The programmes were extensively evaluated (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977) and the outcomes regarding Direct Instruction were impressive given the consistency of results over time, the school areas included (rural and inner city), and the communities involved (Becker, 1978). Direct Instruction demonstrated positive outcomes more than any other approach in basic academic skills, comprehension, and affective measures (Becker, 1977, 1978, 1984; Carnine, 1977; Engelmann, Becker, Carnine, & Gersten, 1988). The benefits continued to be realised up to nine years after the project ended (Gersten, Keating, & Becker, 1988; Meyer, 1984).

A variety of programmes have been developed using the basic instructional model. As a teaching innovation, Direct Instruction has several distinctive features and these will be noted briefly to emphasise the innovative nature of this strategy.

Characteristics of Direct Instruction

Direct Instruction is based on four assumptions. First, all children will learn if they are taught correctly, regardless of their previous learning history. Second, learning the basic academic skills should be the prime aim of any compensatory education programme. Third, disadvantaged children tend to be behind other children in essential academic skills, and, fourth, in order to catch up, disadvantaged children must be taught more in less time.

With these assumptions in mind, Direct Instruction programmes are based on the systematic and explicit teaching of academic strategies. In developing the instructional design, cognitive learning is seen as comprising the analysis of

behaviour, knowledge systems and communication (Engelmann & Carnine, 1982). The result is a comprehensive theory of instruction.

Distinctive features of Direct Instruction, listed by Moore (1981, p.34) are:

1. Carefully structured lesson programmes designed to maximise student growth.
2. An increased amount of academic engaged time.
3. Direct Instruction teaching competencies (signalling, pacing, correction procedures, feedback, and reinforcement.).
4. Quality control procedures (scripted lesson presentation, monitoring of student progress,criterion-referenced testing, teacher training).
5. Active participation of students (moving from teacher dependent activity to independent problem solving).
6. Making instructional goals clear to both instructor and students.
7. Mastery learning.
8. Appropriate student placement in programmes, with placement tests of required skills an integral part of each programme.

A wide variety of Direct Instruction programmes are available in commercially prepared packages. Subjects include spelling, reading, expressive writing, mathematics, social studies, cursive handwriting, microcomputer programming, study skills, and social skills. Examples of programmes in specific subject areas include Spelling Mastery (Dixon & Engelmann,1981, 1990), Corrective Reading (Engelmann, Hanner, & Johnson, 1978, 1989), Corrective Mathematics (Engelmann & Carnine, 1981), Expressive Writing (Engelmann & Silbert, 1985) and more recently, Skills for School Success (Archer, 1989).

The teaching method continues to be applied to a broadening range of increasingly complex cognitive skills. Many studies have investigated the results of the implementation of these programmes. The next section looks briefly at some of the research findings.

Direct Instruction Research

An extensive body of empirical evidence attesting to the effectiveness of Direct Instruction has been accumulated. In the U.S.A., the studies are too numerous to document, but documentation is available in reviews of research by Becker (1984), Cotton and Savard (1982), Gregory (1983), and White (1988). For the purposes of this study, it is important to note that significant academic gains have been demonstrated by a variety of participants after the implementation of Direct Instruction. Examples include studies with bilingual students (Gersten, Brockway & Henares, 1983), preschool students (Weisberg, 1983, 1984; Weisberg, Packer, & Weisberg, 1981), educable mentally retarded children (Crozier, 1974), and remedial students (Branwhite, 1982; White, 1988).

In Australia, the implementation of Direct Instruction has achieved impressive results in a variety of settings. Alex Maggs and others have trialled the method with children in regular classrooms (Hawke, Maggs, & Moore, 1979; Hawke, Maggs, & Murdoch, 1978; Maggs & Moore, 1978), with retarded children (Clunies-Ross, 1979), and with immigrant, cerebral palsied, disadvantaged white, aboriginal, and Down's syndrome children (reviewed by Lockery & Maggs, 1982). In all cases, language competencies increased and, where tested, intelligence.

Despite this impressive accumulation of research, it is evident that teachers, generally, do not perceive Direct Instruction to be a worthwhile teaching innovation. The next section provides an overview of what researchers have noted about teacher attitudes towards Direct Instruction as an innovation. In particular, these observations are linked to the three major characteristics identified by Fullan (1991) as affecting teacher attitude; that is, need, clarity and cost.

Teacher attitude towards Direct Instruction

Very few studies have investigated teacher attitude towards Direct Instruction. Generally, researchers have made observations based on anecdotal evidence. Care has been taken to distinguish between comments made by and about teachers who have taught or studied Direct Instruction and those comments regarding teachers' first impressions of the approach, as an innovation.

Need to implement Direct Instruction

Teachers hold varying personal philosophies as to how children are best taught. Some believe children will learn "academic skills, creativity and self esteem if they are allowed to learn inductively – to discover rules, facts and underlying principles from guided exposure to and with language, numbers, games, and so on." (Cotton & Savard, 1982, p.2) Other teachers place more emphasis on the learning and mastery of basic skills, and consider themselves responsible for the academic achievement of their students (Fields, 1986). Tobin (1987) concluded that teachers' beliefs about how students learn and what they ought to learn had the greatest impact on what teachers did in the classroom and whether they saw a need to change. Those teachers holding more 'child–directed' philosophies will not perceive a need to implement teaching methods based on the mastery of basic skills such as Direct Instruction. Many teachers involved in the Follow Through Project reported a philosophical clash between their 'discovery learning' methods and the structured, teacher–initiated method of Direct Instruction (Gersten, Carnine, Zoref, & Cronin, 1986; Gersten & Guskey, 1985). In those series of studies the teachers had no choice in the implementation of the Direct Instruction.

No studies commenting on the implementation of Direct Instruction by teachers in regular classrooms as a result of a perceived need were found in the literature.

It is probable that only teachers whose teaching philosophy is based on mastery of basic skills would see the need for a teaching strategy such as Direct Instruction.

Clarity of the Direct Instruction strategy

A feature of the Direct Instruction method is the high degree of specificity in terms of teaching behaviours. While Doyle and Ponder (1977) describe specificity as desirable, Fullan and Pomfret (1977) refer to the 'dilemma of explicitness'. In other words, innovations that are highly specific may not be seen as appropriate for a variety of settings, and potentially overwhelm people with information (Fullan, 1982). Waugh (1983) suggests that innovations should not be so specific that they do not allow teachers to adapt them to suit their own classrooms. Direct Instruction is considered by many to be too specific. The lessons are entirely scripted and the behaviour of the teacher is carefully defined. This creates the potential for conflict, as Direct Instruction could be seen by some teachers as threatening their prerogative in decision making, creativity, and intuition (Fields, 1986). Teachers resent having little room to improvise in the lessons. Gersten and Guskey (1985) reported that teachers initially felt stifled by the approach, in which all decisions were seen to be made for them. Other researchers (Becker, 1977; Cole & Chan, 1990; Mathes & Proctor, 1988) have noted that some teachers described Direct Instruction programmes as too prescriptive, regimented or authoritarian. In this case, it appears Direct Instruction is *too* clear. A fine line exists between clarity and over prescriptiveness.

Cost of implementing Direct Instruction

The implementation of Direct Instruction programmes requires time and effort on behalf of the teacher as well as an initial expense in purchasing the materials. The programmes need planning, practice and, ideally, training prior to their implementation (Fields, 1986). This necessitates a high level of commitment and motivation on behalf of the teacher. To quote Gersten, Carnine and Woodward (1987, p.49), " becoming competent and comfortable with Direct Instruction takes considerably more time and effort than many teachers and administrators are willing to take."

Many features of Direct Instruction necessitate careful training and preparation. These include following the scripts, co-ordinating unison responses, the use of correction procedures, the maintenance of student accuracy at a fast pace, and procedures for signalling. It is these features that are the most frequently criticised by teachers (Becker, 1977; Cole & Chan, 1990; Gersten, Carnine, & Woodward, 1987). The two teaching skills of Direct Instruction found to be most strongly related to student achievement, immediate correction of student errors and maintaining student accuracy at a fast pace, are also the most difficult to master (Gersten, Carnine, & Williams, 1982). Teachers may not perceive the gains to be achieved by implementing Direct Instruction as outweighing the time and effort required.

The high cost of purchasing the commercial packages has also been criticised, and may contribute to a teacher's reluctance to initiate the programmes (Nicholls, 1980). The price of equipping a classroom with teacher handbooks, student books and consumable workbooks is considerable. The successful implementation of Direct Instruction does require a lot of time and initial effort by the teacher. Changes include new teaching materials, teaching methods and,

often, a change of personal teaching philosophy. These are steps many teachers apparently are not prepared to take.

In the light of the above discussion, Direct Instruction does not appear to have the required characteristics for an innovation to be received favourably by a teacher. However, after a teacher has taught Direct Instruction for a period of time, positive comments have been recorded (Gersten et al., 1986).

Successful Implementation of Direct Instruction

Most innovations cannot be fully assessed until after they have been implemented for a period of time. Fullan (1991, p.xi) observed that "successful innovations and reforms are usually clear after they work, not in advance." Crandall (1983) found that commitment to an innovation only came after teachers were actively involved in its use. Generally, this was after the teachers had seen successful results with their students (Berlin & Jensen, 1989; Hoyle, 1976, Print, 1987). Direct instruction is no exception.

Teachers have indicated a positive attitude towards Direct Instruction after successfully implementing the approach in the classroom, (Becker, 1984; Gersten et al., 1986; Gersten & Guskey, 1985) or after thorough training and supervised experience in teaching the method (Proctor, 1989). Proctor (1989) found that 89 per cent of undergraduates and teacher graduates in his study agreed that regular use of Direct Instruction had increased their appreciation of it. Mathes and Proctor (1988) felt that Direct Instruction could not be "fully evaluated by reading descriptions of it, looking at the materials or, even, observing it being used." (p.97) Teaching Direct Instruction for a period of time allows an accurate appraisal of the method and its underlying philosophy. Many teachers overcome their initial negative reaction to the apparent excessive structure and regimentation (Gersten et al., 1986; Mathes & Proctor, 1988). A

more positive attitude towards Direct Instruction has been expressed by teachers after seeing students achieve better academic results (Engelmann et al., 1988; Gersten et al., 1986; Gersten & Guskey, 1985).

Conclusion

This review of the literature associated with innovations and their implementation has highlighted three points. First, Direct Instruction appears a suitable example with which to examine the nature of innovations. An investigation of the aspects of need, clarity and cost in relation to teacher attitude towards Direct Instruction, would contribute to theory building by adding to knowledge of innovations as a concept through the development of explanatory theory. Second, attempts to measure attitudes to innovation are very limited and a need exists to establish sound measurement procedures in this area. To achieve this a measurement model is required which is capable of linking the theoretical deliberations espoused in this chapter to the information available within the frame of reference implicit in the above. This important aspect of the study provides the substance for the next chapter. Third, differences in attitude towards Direct Instruction due to differing degrees of exposure to the strategy can be investigated through the employment of the measuring instrument developed in accordance with the specifications of innovation theory. This investigation will be considered in Chapter Six following the development of the measuring instrument, and described in Chapter Five.

The two central aims of this study, therefore, will be to develop a sound measure of attitude towards Direct Instruction and to examine the extent of the change in attitude to Direct Instruction after exposure to the strategy, for teachers possessing differing histories relating to knowledge of Direct Instruction.

Research Questions

In line with these two central aims, the study is organised around four research questions:

1. What are the psychometric properties of the Attitude towards Direct Instruction Instrument as adopted and developed for this study?
2. To what extent does a brief demonstration of Direct Instruction affect teacher attitude towards this strategy?
3. To what extent does previous knowledge of Direct Instruction affect teacher attitude towards this strategy?
4. To what extent does previous knowledge of Direct Instruction affect teacher attitude towards this strategy following a brief demonstration of the method?

CHAPTER THREE
MEASUREMENT AND THE MEASUREMENT MODEL

In Chapter Two, the conceptual framework for the study based on innovation theory was developed. In order to test the theoretical implications of interest, an appropriate measure is required. Such an instrument, based on a measurement model, will be required to investigate attitude towards the teaching strategy Direct Instruction, and in so doing link this strategy to the more general context of innovation theory. In this chapter, the nature of measurement is explored and the steps involved in developing a sound measuring instrument through the employment of a measurement model are elucidated.

What is Measurement?

Several definitions of measurement have been provided in the literature. Stevens (1951) defines measurement as the assignment of numerals to objects, events, or persons, according to rules. Cohen and Nagel (1934) consider measurement to be the correlation with numbers of entities that are not numbers. More recently, Black and Champion (1976, p.169) defined measurement as "the assignment of numbers to, and the symbolic correlation of the numbers, with nominal social and/or psychological properties of individuals and/or groups." This latter definition places measurement squarely within a social context. These definitions, however, are very limited as they fail to embrace the basic principles of measurement when placed in a scientific, objective context.

Measurement involves comparisons. As Wright and Masters (1982, p.2) noted, "only those characteristics of an object can be measured which can be described in terms of 'more' or 'less'." These need to be described in such a way that normal arithmetic operations can be obtained from the measure, which can then

be easily placed on the linear continuum. It follows that only characteristics that can be thought of as linear magnitudes may be described by measurement, for example, height, weight, temperature, strength of feeling favourable to an object. The continuum is purely an abstract conception, with an origin, a point at which counting begins, and a unit of measurement. In summary, Wright and Masters (1982, p.3) consider the essence of measurement to involve:

1. the reduction of experience to a one dimensional abstraction,
2. more or less comparisons among persons and items,
3. the idea of linear magnitude inherent in positioning objects along a line, and
4. a unit determined by a process which can be repeated without modification over a range of the variable.

The process "which can be repeated without modification" ... (involves) ... "a theory or model for how persons and items must interact to produce useful observations." (Wright & Masters, 1982, p.3) The model is based on the fundamental requirements of measurement, which will now be discussed.

The Requirements of Measurement

In order for measurement to occur, two requirements must be met, order and objectivity.

Order

This principle requires an order of both persons and items. People with more ability must always gain a higher score than people with less ability. A person of a given ability should have a greater chance of success on an easier item than a more difficult one. For any two items, the easier item should be answered correctly more often than the more difficult item. When this requirement has been met, the set of statements constitute a unidimensional scale. This requirement was enuciated explicitly by Guttman (1950, p.62).

> If a person endorses a more extreme statement, he should endorse all less extreme statements if the statements are to be considered a scale. We shall call a set of items of common content a scale if a person with a higher rank than another person is just as high or higher on every item than the other person.

Thus for true measurement to be obtained, the response matrix or pattern must conform to a Guttman pattern. This pattern is represented in its ideal form, for seven items listed in order of increasing difficulty, in Table 1.

TABLE 1
Response Matrix conforming to the Guttman Scale

Item							Total
1	2	3	4	5	6	7	Score
1	0	0	0	0	0	0	1
1	1	0	0	0	0	0	2
1	1	1	0	0	0	0	3
1	1	1	1	0	0	0	4
1	1	1	1	1	0	0	5
1	1	1	1	1	1	0	6
1	1	1	1	1	1	1	7

Objectivity

The second requirement of measurement is objectivity. A disadvantage common to many so called forms of measurement in the social sciences is the inability to compare the score a person achieves on one scale to the score a person achieves on another. The score has meaning only in relation to scores earned by others in the sample. If a person has a score of 84 on an attitude scale it is necessary to know what scale and what group of people were involved in order for the score to make sense. However, if told a child is one hundred and twelve centimetres tall, it is possible to make sense of this information without reference to the height of other children. The measure therefore has objective meaning. Wright (1967, p.3) observed that there were two conditions necessary to achieve objectivity. He noted,

> first, the calibration of measuring instruments must be independent of those objects that happen to be used for calibration. Second, the measurement of objects must be independent of the instrument that happens to be used for measuring.

This can be achieved by calibrating items along a latent continuum on which both persons and items could be positioned. Such a technique results in 'scale-free' person measurement and 'sample-free' item calibration. The differences between pairs of person measures or item calibrations are then sample independent. This means that two items must differ by the same number of measurement units no matter what sample of persons is used. Similarly, two persons must differ by the same number of measurement units no matter what items are used in the measurement procedure. This requirement makes it possible to generalise measurement beyond the particular instrument used. Objects measured on similar but not identical instruments can then be compared.

Measurement of Latent Variables

In the social sciences, the traits we are trying to measure (for example, attitude to work) are not overt as in the physical science sense (for example, mass or length). Only the effect of these underlying traits or constructs can be observed. They are therefore more appropriately called latent traits or variables. Items used to calibrate the variable are developed from observed patterns of experiences or behaviour or from certain conceptualisations based on theory developed in that particular area. By analysing the response patterns to the items it is possible to see if the items fit together in such a way that the variable can be defined. Should an item not fit the model or response pattern it is probably not related to the variable in question within the frame of reference as defined.

Scores are achieved from the response of people to a particular provocation, for example, a pen and paper test in mathematics. The response should be such that the more ability the person has the more this provocation will reveal. The score is purely a matter of a category or number. Counting is the first step, while measurement is deduced from these raw scores using a measurement model which has the properties described above. Measurement models possessing these properties were proposed and developed by Georg Rasch in the 1950's and first introduced to the Social Sciences as the Simple Logistic Model of Rasch (1960/1980).

The Rasch Model of Measurement

The scaling tradition of measurement had its roots in the work of Thorndike (1904, 1916), Thurstone (1931), and Guttman (1950, 1954). Rasch, a Danish mathematician, reached similar conclusions as to the requirements of measurement by independent means. When confronted with the problem of equating different reading tests taken at different times by different people, Rasch realised the necessity of finding a way to make the the scores sample free and item free (Wright, 1988).

In order to do this, Rasch explored the concept of sufficient statistics, in which the total score is all that is needed to capture all information about an item or person. He discovered that, using the concept of conditional probability, if he conditioned on a sufficient statistic, the other parameter was partialled or conditioned out of the model. In other words, if he had a sufficient statistic for the person measure, he could condition out the person measure and learn more about the item parameter. Conversely, conditioning on the item sufficient statistic leaves only the person parameter in the model. In this way, Rasch developed the concept of 'specific objectivity' by following on from the pioneering work of Ronald Fisher in the 1930's, and especially the latter's notion of

maximum likelihood. With this technique, the estimates obtained for the parameters being measured are so structured as to achieve the maximum likelihood of them conforming to the model.

Rasch realised that a probability model, in which error was built into its conceptualisation, was necessary in order to account for error in fluctuations of responses. The concept of 'odds' was applied to the relationship between the ability of a person and the difficulty of an item. The odds of a person of particular ability achieving success on an item of a particular difficulty was described as the mathematical relationship: b (ability)/ d (difficulty). The odds were then translated into probability, using the relationship: odds/(1 + odds). This means that the response of a person to a particular item is governed only by the difference between the ability of the person and the difficulty of the item. This difference specifies the probability of what is supposed to happen when a person of a given ability attempts an item of a given level of difficulty. It follows that the chances of a person of greater ability getting an item correct ought to be greater than for a person of lesser ability. Also, a person of a given ability ought to have a greater chance of getting an easy item correct than a more difficult item. At the same time it ought to be that the easier the item, the more likely any person will get it correct. The result was the development of the Simple Logistic Model of Rasch (SLM).

The Rasch model is not a data model, but may be used with data. For true measurement to evolve, the data must conform to the strict requirements described above. As stated by Wright (1991, p.158), "the scientific questions are not 'Does the model fit the data?' and 'Is the model violated', but 'Can the data fit the model?' and 'Is (sic) the data useful?' "

Advantages of the Rasch Model

Rasch models develop from the fundamentals of measurement, and have been described by Wright (1988, p.28) as "a definition of measurement" or even, "*the* law of measurement." As a means of data analysis, the Rasch model offers many desirable features.

A Trait Line

The calibration of items and the measurement of persons results in a response continuum or trait line on which both persons and items can be located. Items which do not conform to the model are not measuring the same underlying variable and are removed from the measure. The continuum is ordered around a relative zero point, and items are placed onto the continuum in an increasing order of provocativity or difficulty. This means that items requiring a minimum level of ability in order to answer correctly are seen as least provocative. Items requiring a participant to have a high level of ability in order to answer correctly are seen as the most provocative.

Unit of Measurement

Estimates are in the logarithmic metric. The parametric relationships then become additive and hence the measurements, which are the parameter estimates, constitute an interval scale (Andrich, 1978b).

Error of Measurement

Measurement errors are estimated independently for each item and person. These measures are sample free and therefore test specific (Wright, 1991). They specify exactly how well the items will perform when applied to any sample.

Validity

Construct validity demonstrates whether or not a measuring instrument is in fact measuring a particular phenomenon. Gay (1987, p.131) defines it as the "degree to which a test measures an intended hypothetical construct." A construct refers to a trait that is not directly observable, such as attitude or intelligence. Therefore only its effect can be observed. It follows that research studies involving a construct are only valid to the extent that the measure of the construct is valid.

Construct validity is inherent when the data is found to fit the Rasch model, in that unidimensionality is a requirement within the context of the frame of reference as specified by the conceptual framework. The items are then all measuring the same underlying trait or construct. The extent to which the items are arranged in a logical manner along the continuum provides additional evidence for construct validity.

Reliability or Person Separation

Reliability is usually defined as the degree to which a questionnaire consistently measures whatever it measures (Gay, 1987). The Index of Person Separation is the Rasch model equivalent of Cronbach's Alpha and is a measure of internal consistency. It indicates how reliably the persons measured are separated along the continuum by the test and determines what power can be put into the tests-of-fit to the model. The greater the variability across the measure, that is, the greater the spread of persons, the more chance the test has of assessing the tests-of-fit in a meaningful manner. As a measure of internal consistency, this index provides an estimation of error variance relative to true variance. Thus, index values close to 1.00 indicate very small error variance is present and hence the instrument is displaying high reliability. An index greater than 0.90 would be considered very satisfactory in this regard.

Application of the Rasch Model

Rasch measurement models have been used successfully to develop such well known academic tests as the TORCH tests of Reading Comprehension (Mossenson, Hill, & Masters, 1987), Key Maths (Connolly, Nachtman, & Pritchett, 1976) and the Woodcock Reading Mastery Tests (Woodcock, 1987). While the SLM has dominated testing applications associated with multiple choice and dichotomous test formats, extensions to this model have been developed over the past decade for use with multiple category items, such as present with most forms of attitude questionnaires, and usually scored according to the familiar Likert format. Before examining this extended format to the Rasch model, a brief overview of attitude measurement is required.

Attitude Measurement

Over the years, many differing explanations of attitude have been offered (Ajzen & Fishbein, 1980; McGuire, 1985; Shaw & Wright, 1967; Wicker, 1969). Researchers have disagreed about the psychological locus of attitude, whether attitude should be defined as a response or as a readiness to respond, the degree to which attitudes are organised, and the extent to which attitudes are learned through previous experience. Various broad definitions have resulted in confusion as to the very essence of attitude (Greenwald, 1989), and consequently how it should be measured. In 1972, Fishbein and Ajzen found 500 different operational definitions of attitude, and in 70 per cent of the 200 studies in which attitude was defined in several ways, different results were obtained depending on which definition was used.

Attitude, as with most social and psychological variables, is a latent variable that cannot be directly observed. Explicit attempts to link attitude theory and measurement are absent in a great deal of attitude research (Breckler &

Wiggins, 1989; Fishbein & Ajzen, 1975; Ostrom,1989). This is due mainly to the lack of consensus as to the nature of attitude and a misunderstanding of the nature of measurement. A measurement model, according to Wright and Masters (1982, p 90), extracts "from suitable data a useful definition of an intended variable and then ... measures persons on this variable." One advantage of the Rasch model when applied to attitude data is that it attempts to verify that the items on an instrument work together and define the variable.

In order, therefore, to measure attitude in a sound psychometric manner, adherence to the characteristics that allow the measures to be placed on a linear continuum must prevail. A common fault is to put items measuring several dimensions into one scale, and then analyse the data as if the scale was unidimensional (Rennie & Parker, 1984). Thurstone, a pioneer in attitude measurement, envisaged attitude as an affect which was either for or against an object. It was seen as bipolar, running on a continuum from positive to negative, or favourable to unfavourable. The continuum had a zero or neutral point and attitudes were considered to range from one extreme to the other.

Two of the more popular methods of assessing attitude are based on the understanding or assumption that attitude is a unidimensional concept. One method, Thurstone's Equal–Appearing Interval Scale (Thurstone & Chave, 1929) is based on the argument that in "all measurement we must restrict ourselves to some specified continuum along which measurement takes place." (Ajzen & Fishbein, 1980) Briefly, Thurstone's method involves assigning an affectivity value to a series of statements thus specifying their location on an evaluative continuum. This is established by the time consuming task of asking a number of judges to rate each item according to its affect it had on them. In the final questionnaire, participants are asked to agree or disagree with each statement in the scale and the attitude score calculated as the mean or median of the scale

values of statements with which the person agreed. This approach has the advantages of the provision of statement scale values and being statistically rigorous in establishing a continuum. However, it is time consuming to construct and no formal person parameter is provided in the model.

The Likert method of Summated Ratings (Likert, 1932), on the other hand, asks participants to indicate the intensity with which they agree or disagree with a number of statements. The statements are selected to represent the range of attitude from the favourable to the unfavourable end of the continuum. A person's score is calculated by assigning a numerical value to each of the responses and then summing the values for all questions. As a consequence, this procedure does not provide statement scale values, but is relatively simple to use in terms of construction and interpretation on a superficial level. While statistical rigour is not involved in developing a continuum, direct estimates of person attitude values are obtained without resorting to a measurement model.

The Extended Model of Rasch

The generalisation of the SLM of Rasch to accommodate item responses to multiple categories provides a measurement model that captures the essence of both the Thurstone format of comparative judgements involving sequences of agree or disagree and the Likert format of degrees of agreement or disagreement (Andrich, 1975, 1982). When applied to attitude measurement, the extended model of Rasch overcomes many of the problems associated with the Thurstone and Likert methods, yet retains the best theoretical and practical features of both. It provides a method to unify the Thurstone goal of item scaling and the Likert procedure for attitude measurement (Andrich, 1978d) enabling both items and persons to be placed on the same scale (Rost, 1988). With this extended model, the affective value of an item and the attitude of a person are each

parameterised. As a result, objective measures for both sets of parameters can be estimated from the data collected using this measurement model while taking account of the extended information available in the different categories associated with each item. When working with attitude measures, item difficulty is usually referred to as item affectivity and person ability as person attitude.

Andrich (1978c) first formalised an extension to the SLM of Rasch for ordered response categories of attitude, or Likert–type, questionnaires by specifying a set of thresholds associated with the item categories. Subsequent reparameterisation of the item category coefficients (Andrich, 1982, 1985, 1988) has shown that this model offers several desirable features which support its use in attitude measurement. The extended model of Rasch is therefore considered a suitable measurement model to be employed in this study for the development of the Attitude towards Direct Instruction instrument.

An important feature of this model is that as the number of categories per item increases beyond two (for the dichotomous case), it is possible to parameterise a threshold structure, associated with the categories, to obtain a clear perspective on the item structure based on a hierarchical progression across the categories. A set of thresholds are conceptualised as a set of boundaries located between the response categories of an item and specify the change in probability of a response occurring in one or the other of the two categories separated by the threshold. An important property of these thresholds is that they will be ordered in a hierarchical fashion, from lowest to highest, only when the data fit the construction of the model (Andrich, 1985).

The extended model takes the general form when person n of attitude β_n responds to item i of affectivity δ_i for each item i where there are m ordered thresholds τ_{ki}, for k = 1, m, on the measurement continuum:

$$\Pr\{X = x;\beta_n, \delta_i, \tau_{ki}\} = \exp\left\{x(\beta_n - \delta_i) - \sum_{k=1}^{x} \tau_{ki}\right\} \Big/ \gamma_{ni} \qquad (1)$$

where the score $x \in \{0,1,...,m\}$ and the normalising factor is

$$\gamma_{ni} = 1 + \sum_{k=1}^{m} \left\{\exp k(\beta_n - \delta_i) - \sum_{j}^{k} \tau_{ji}\right\}$$

The constraints $\sum_i \hat{\delta}_i = 0.0$ and $\sum_k \hat{\tau}_{ki} = 0.0$ are imposed, without loss of generality, for each item *i* in estimating these parameters. The category coefficient for score *x* is defined as

$$\kappa_{xi} = -\sum_{k=1}^{x} \tau_{ki} \quad , \qquad \text{where } \kappa_{0i} = 0$$

The thresholds can also be reparameterised to form a hierarchy of parameters which are directly related to the Guttman (1950, 1954) principal components, where the number of parameters is governed by the number of categories in the scoring function.

$$\begin{aligned} \Pr\{x;\beta,\delta,\theta,\eta,\psi\} = \frac{1}{\gamma}\exp\{ & -x\delta \\ & + x(m-x)\theta \\ & + x(m-x)(2x-m)\eta \\ & + x(m-x)(5x^2 - 5xm + m + 1)\psi \\ & + x\beta \ \} \end{aligned} \qquad (2)$$

For example, with four categories, three item parameters can be estimated. To date, four parameters have been identified and clarified, although it is possible to have more, provided the number of categories per item is greater than five. These parameters are labelled, in hierarchical order, as *location* (δ) *scale* (θ), *skewness*(η), and *kurtosis* (ψ). In a real sense, the higher order parameters (from *scale* onwards) qualify the location of an item on the latent trait continuum, with the second order parameter, *scale*, defining the unit of measurement for that item.

While the extended model possesses all the desirable measurement features discussed earlier for the SLM, additional features need to be addressed. The model is available in a computer program, ASCORE (Andrich, Lyne, & Sheridan, 1991) which will be employed for all item analyses reported in this study.

Threshold Estimates

A feature of the extended model is that it can account for correct scoring of the response categories, but only if the threshold estimates are correctly ordered. As the affectivity of an item increases, the probability of a response in a higher category for a given person decreases. For a person possessing a low attitude towards the trait under consideration, the most probable responses to an item of high affectivity value would be in the first category. A person possessing a higher attitude value would most probably endorse a higher order category for the same item, so that one or more thresholds will be exceeded. The logic of this situation means that a person of higher attitude would endorse successively higher placed categories for items of increasingly lower affectivity values, than would be the case for a person of lesser attitude toward the trait. If this consistency is observed across the person response patterns for all items, then the Guttman requirement for objective measurement will prevail and the item thresholds will be in an ordered sequence.

If the threshold estimates appear in ordered sequence from low to high then the item can be considered as behaving in the fashion intended, according to the Likert scale format. However, if thresholds are disordered then the probability of a response in one category may never be greater than the probability of a response in at least one other category and the scoring for the categories is not acting as expected. The nature of this order then needs to be investigated as a priority. The usual diagnosis is that dependencies are present in the data. This

means that the respondents are favouring specific categories to a much higher degree than would be expected in relation to other items reportedly measuring the variable as defined. Andrich (1982) noted that only if the thresholds were ordered was the model distribution strictly unimodal. If threshold estimates are disordered then the data do not fit the construction of the model.

Scale Estimates

Returning to the reparameterised alternative to the threshold estimates and as specified earlier in (2), while the first order item parameter specifies the average location of the item on the continuum, the second order item parameter is associated with the dispersion of the item categories and has been labelled as the *scale* parameter. It provides information about the average spread of the thresholds for each item and can be estimated if three or more categories per item are present.

The third order item parameter estimates the degree of skewness associated with the item categories. Andrich and Schoubroeck (1989) have demonstrated how this parameter can detect and also provide information pertaining to the presence of a response set across items. Because of the nature of the problem addressed in this study , attention will be focussed on the *location* and *scale* parameters only. This is especially so as the emphasis is on comparisons across scales, for which attention to the *scale* parameter is of particular importance.

Tests-of-fit

When the item and person estimates have been established, and the thresholds are ordered, it is important to examine the degree to which these estimates are in accord with the conceptual framework. This is accomplished by substituting the parameter estimates back into the model and examining the degree to which

the expected values predicted from the model differ from the observed data values. Different tests-of-fit are then employed to assess the degree to which the original data satisfied the requirements for objective measurement as specified by the measurement model. In this context, no significant difference between the original data and the expected scores across both items and persons should exist. If the tests-of-fit reveal there are significant differences then the items and persons involved need to be reassessed in conjunction with the original conceptualisation as specified by the study. Item contents are examined and reasons explored concerning why these items are causing problems. Similarly, the response patterns for aberrant persons are examined to determine the extent of their variation from the expected Guttman pattern. In this way a better understanding is obtained regarding the variable being constructed.

Two main tests-of-fit are available with the ASCORE program, the item-trait interaction and the item-person interaction tests.

Item-Trait Interaction

The item-trait test-of-fit statistics are used to examine the consistency of the item parameters over the range of person estimates. For this analysis, the persons are divided into several segments, or class intervals, of approximately equal size to create contingency tables of expected and observed values. A chi-square statistic is then derived to assess the probability of the degree of divergence between observed and expected values occurring by chance alone. The number of segments is determined by the size of the calibration sample and an approximate rule-of-thumb involves 60 persons per segment. In addition to individual statistics available for each item, these data can be pooled across all items to obtain an overall test-of-fit statistic. This index indicates the degree of consensus displayed collectively by all items of the instrument across persons located at differing attitude levels.

Item-person Interaction

This test-of-fit examines the degree to which persons are responding to items of differing affectivity value in a logical and consistent manner and relates directly to the consistency of individual person and item response patterns. There are two aspects to this test-of-fit, one involving the response pattern of individual persons across all items, and the other, the pattern for each item across all persons. The fit statistic assesses the degree of divergence, or the residual, between the expected estimate and actual data values for each person-item combination when summed over all items (for a given person) or summed over all persons (for a given item). Further details regarding the derivation of these statistics are provided in Andrich and Sheridan (1980).

These fit statistics approximate a t-distribution when the data fit the model. Thus the overall distribution for the item statistics and for the person statistics should each have a mean of zero and a standard deviation of one. As a consequence, any individual item or person statistic which is greater than plus or minus a particular desired limit (usually set at two) is considered to be diverging from the model in a systematic way that is not accountable by chance alone (at the 5 per cent level of significance). Decreasing or negative values indicate a person or item pattern response that is overdiscriminating or fitting the model very closely, whilst increasing or positive values indicate underdiscrimination or poor fit to the model. The former case indicates that dependencies are present in the data whilst the latter informs that the data are largely 'noise' and contributing nothing of significance to knowledge or an understanding of the variable being constructed.

Attaining fit to the model

The aim of an item analysis is to provide knowledge about the variable created for the study and to develop an instrument possessing sound psychometric properties. It may be necessary to remove items and specific data responses of persons that preliminary analysis indicates are not conforming to the model. This decision is based on accumulated evidence from the sources discussed above. An examination of those items and person data responses removed provide valuable insight as to the nature and meaning of the variable being measured. The items are not necessarily removed from any future version of the scale, but may be retained, modified and resubmitted to another sample at some future point of time. In addition, persons (whose response data do not conform to the model) could be interviewed. Further analysis of the results of these procedures would provide a better understanding of the theoretical construct being measured and so contribute to the aim of this research in the building of explanatory theory.

The process of data analysis and the controlled modification of the datafile continues until the set of items demonstrates a satisfactory level of fit to the model. Simultaneously, increasing understanding of the meaning of the variable becomes possible. At this point, a measuring instrument has been constructed.

Conclusion

This chapter has been concerned with measurement, and in particular, measurement of attitude. In order for a construct to be measured the means chosen to do so must meet the fundamentals of measurement. The Rasch model meets these requirements. In this study, the extended model of Rasch will be used to analyse data collected from teachers responding to a questionnaire entitled Attitude towards Direct Instruction. Item statements for

the questionnaire were developed from several sources in accordance with the conceptual framework based on the innovation theory discussed in Chapter Two. Before considering the item analysis, the design of the study will be described and this provides the substance for the next chapter.

CHAPTER FOUR
RESEARCH DESIGN

This is an empirical study designed to assess the influence of the level of prior knowledge of a teaching strategy on attitude towards the strategy. A considerable portion of the study is devoted to the operationalisation and calibration of the instrument designed to measure the variable of interest, Attitude towards Direct Instruction. The conceptual framework guiding the operationalisation of the instrument draws on innovation theory, as described in Chapter Two, and the calibration is achieved by employing a measurement model which takes account of recent developments in psychometric theory, as described in Chapter Three. The other major portion of the study examines the relationship between prior knowledge of Direct Instruction and the influence this knowledge has on attitude towards this strategy as an innovation in teaching. To assess this, the study employs a repeated measures, pretest–posttest design based on graphical procedures which take account of the special features associated with the measurement model.

This chapter is concerned with the study's design. A description of the plan for assessing change in attitude is followed by the strategy employed for operationalising the variable, attitude towards Direct Instruction. The chapter concludes with details of the sample and data collection procedures.

The Plan

The overall purpose of the study is to investigate teacher attitude towards a teaching strategy, in particular the effect on attitude of prior knowledge of the method. A repeated measures, pretest– posttest design is employed to assess participants' attitude before and after viewing a video demonstrating the teaching

method, Direct Instruction. This design is chosen to provide evidence of any change of attitude from one occasion to the next, and if teachers classified according to experience with, and the degree of exposure to, the teaching strategy differ in the intensity of the attitude expressed.

The Video as the Treatment

In order to evaluate teacher attitude to an innovation, it is important that all participants respond to a consistent demonstration of the strategy in question. In this study, a video of a West Australian teacher using Direct Instruction principles to teach five students how to locate a birth notice in the local daily paper is used. A potential difficulty with presenting information to participants is the risk of a 'halo effect'. Therefore, care was taken to ensure the information was presented in a factual and unbiased manner. Two experts in the educational field; Mr D. Leach, Murdoch University and Dr. D. Evans, Edith Cowan University examined the video for content validity. They were asked to address a checklist of all observable features of Direct Instruction, a copy of which appears under the heading "Attitude to Direct Instruction Survey" in Appendix A. Both judges agreed the video showed Direct Instruction in a fair and reasonable manner and included the principal features of the method.

Item Bias

An integral part of the item analysis for this study is to ensure that commensurate measures for both the pretest and posttest occasions are employed. Consequently, a systematic analysis is undertaken to check for bias in the data by assessing the presence or otherwise of interaction effects among the item estimates for the pretest and posttest data.

Assessing the stability of item estimates between the two occasions involves both statistical and graphical techniques, with the emphasis placed in this study on the latter. These techniques take advantage of the special feature of Rasch models whereby the degree of concurrence between any two sets of item estimates can be assessed using standard errors on an individual basis. With the graphical approach, the degree of concurrence between the set of item estimate coordinates for the pretest and posttest occasions relative to an hypothesised identity line can be assessed within a confidence band of a known significance level (say 68 per cent, 95 per cent). Any item whose coordinates are located outside the confidence bands provides evidence of interaction effects associated with this item.

The presence of bias so detected would then need to be investigated in relation to the measure of the variable. The inclusion of these items with the other items would detract from the precision of the measure being developed and thus obscure the meaning of the variable involved. In this way, a claim of true commensurate measures can be established. While a demonstration of this technique is provided in Chapter Five, a detailed discussion is available in Wright and Stone (1979, 1991).

Attitude Change

When the requirements for commensurate measures have been established, the data for both pretest and posttest occasions can then be pooled and precise person estimates obtained for the next stage of the analysis, assessment of attitude change. The responses for each person, therefore, appear as two separate records within the same data file: one for the pretest and one for the posttest occasion. Thus, if the instrument is a true measure of attitude towards Direct Instruction then the response-pairs for each person would represent, in

effect, two levels of attitude (if a change in attitude towards Direct Instruction has in fact occurred). In effect, the two attitude measures refer to *different* persons *with respect to the variable being measured*, as the person before the event (in the pretest occasion) can be considered a different person from that in the posttest occasion (they have different attitudes!). This contrasts with the technique of modelling measurement of change using log linear models which has been a feature of the Rasch research in Europe (see Fischer, 1976, for example). With the latter approach, the change process is usually modelled formally by building change parameters into the design. This has the effect of creating a second set of items which are employed to account for any change that may be present.

The same graphical technique described earlier for item bias also provides a sound and valuable method of assessment of change between two occasions, this time using the set of two attitude measures obtained for each person. If a plot shift occurs with respect to the location of the hypothesised identity line, then a change of attitude is revealed. This method provides both an overall, or mean, assessment of attitude on a group basis (according to the degree of shift of the hypothesised line), and on an individual basis. The advantage of the latter will be discussed in association with the data analysis reported in Chapter Six.

Teacher Experience with Direct Instruction

Finally, in order to investigate the relationship between previous knowledge of Direct Instruction with attitude, and attitude change, three groups of participants are required. The first, Group One, have had no exposure to the method, Group Two have heard about the method but never taught it, whilst Group Three have taught or are still teaching it. Differences between each group can be assessed

by examining separate plots for each group using the same graphical technique described earlier. Details on this method also appear in Chapter Six.

Operationalisation of the Variable

In order to examine the presence of change, a measuring instrument must be provided that faithfully reflects the concepts specified by the theoretical model. This is achieved through the establishment of a conceptual framework from which items can be derived which will measure what is intended to be measured, and as discussed in Chapter Two. The items are used to provoke from the respondents a reaction that will produce a consistent pattern of outcomes that can be translated using a measurement model into a measure of the variable of interest to the study.

The Measuring Instrument

The items selected for the Attitude towards Direct Instruction questionnaire derive from two sources. A questionnaire designed by Proctor (1989) to investigate teacher attitude toward Direct Instruction drew on aspects of innovation theory similar to those aspects of innovation theory discussed in Chapter Two. The thirty items selected from this source were supplemented by ten new items created especially for the study and based on the same conceptual framework, thus creating an initial questionnaire of 40 items. The complete set of statements is reproduced in Appendix A.

An advantage of using different sources for the selection of items is in the establishment of construct validity. If the items generated from Proctor's (1989) survey and those developed especially for this study are found to be located on the same trait line, evidence is provided for the construct validity of the measure. This is because the same conceptual framework is employed as a theoretical

base to generate the forty items in accordance with the three aspects of the variable as identified in Chapter Two: cost, clarity, and need. An attempt was also made to ensure the items covered the range of affectivity values for the variable. While most items were easily categorised, several did not fall easily into any one group. These items were placed into a fourth group and labelled 'other'.

- Items were stated in both the negative and positive form to minimise response sets. This notion is based on concern over individual differences in response style (Cronbach, 1946). Overall, the questionnaire included 21 items stated in the positive form, and 19 items stated in the negative. It is now common practice to have approximately equal numbers of positively and negatively worded statements so that tendencies to favour either the 'Agree' or 'Disagree' response, to the exclusion of the other, would balance out. Unusual response tendencies can be exposed using the differently worded formats through the employment of Rasch models (Wright & Masters, 1982). This is especially evident with the extended model where attention is directed to the third order item, or *skewness*, parameter (see Andrich & Schoubroeck, 1989, for details). Table 2 shows a breakdown of the items into each category, by positively and negatively worded format.

Each item employs the familiar Likert rating format using five response categories; Strongly Agree, Agree, Not Sure, Disagree, and Strongly Disagree. In this format, the 'Not Sure' (NS) could also have the effect of a neutral response because of its location between the agree and disagree categories. Because the study is investigating the nature of attitude change, the use of the NS category in the original instrument is critical, especially for respondents at the pretest occasion who have had no prior knowledge of Direct Instruction. Some

TABLE 2
Distribution of items across innovation characteristic
by negative and positive wording

Wording Mode	Cost 'C'	Need 'N'	Clarity 'C'	Other 'U'	Total
Positive	4	10	4	3	21
Negative	11	5	0	3	19
Total	15	15	4	6	40

items in the questionnaire have a high degree of specificity to Direct Instruction, and these people would not be in a position to rate these items. It is important from the validity point of view, therefore, that the NS category be provided as an outlet, so that these participants could respond to such items in a non threatening manner. In the early stages of the development of an instrument, the NS category can also provide an avenue for highlighting items that may be phrased in an awkward or clumsy manner, or even ambiguously worded.

However, the role of the NS or neutral category in the Likert format has been questioned. In 1975, DuBois and Burns suggested that respondents may answer in this category for a number of different reasons, not necessarily because of a neutral attitude towards the statement. For example, the respondent may not understand the statement, may be indifferent to it, may know nothing about it, or possibly react to an awkward phrase or word in the item content. Andrich (1978a) has shown that this category will provide problems of a measurement nature if it is included as part of the measure. Further discussion of the role of the NS category and its implications for this study appears later in this Chapter, and also in Chapter Five.

The Sample

Data were collected from two sample groups. The first group, to be referred to as the 'change sample', is used to assess the degree of attitude change between the pretest and posttest occasion resulting from the viewing of a video tape demonstrating Direct Instruction. The second group, to be referred to as the 'control sample', was included in the study to assess the influence of the NS response category on the precision of the measuring instrument.

Change Sample

This sample consisted of 144 primary school trained teachers located in Perth, Western Australia. The entire teaching staff from three schools participated, as well as teachers attending postgraduate courses at Edith Cowan University. The remainder were teachers approached by the researcher on a personal basis.

In an attempt to assess the influence of prior knowledge on attitude, as predicted by innovation theory, teachers with varying degrees of experience with, and background knowledge of, Direct Instruction were selected. When grouped according to level of previous exposure to Direct Instruction, Group One (no prior exposure) contained 79, Group Two (some exposure) had 45, while Group Three (considerable exposure) had 20. All teachers in the sample were post service and, as such, have been faced with the necessity of deciding on teaching strategies for use in their classrooms. They would also be aware of the constant barrage of alternative teaching methods and subject matter, and the consequent need to make choices.

In accordance with the pretest–posttest design employed, data were collected from the participants using the Attitude towards Direct Instruction and this constituted the pretest occasion. After viewing the video, the attitude towards

Direct Instruction was again collected from those participants, this time as the posttest occasion. The same instrument was employed for both occasions.

The nature of the investigation proposed for this study provided an interesting design problem. On the one hand, it was necessary to include in the change sample teachers with no prior knowledge of Direct Instruction, but provide, on the other hand, every opportunity for them to answer the questionnaire in a non-threatening manner. This meant allowing for the option of NS responses, especially to items of a highly specific nature. In this way, these teachers were provided with a sensible presentation which allowed them to respond to the remaining less specialised items, involving more general references to innovations, and in accordance with the proposed conceptual framework. By this means, it became possible to obtain measures for this group of teachers that would maximise the validity of their special situation. The mechanism for handling the NS category responses, involving the creation of missing data is discussed in more detail later in this Chapter and in Chapter Five.

The confounding influence of the NS category was apparent when, due to an unforeseen delay in the data collection, the opportunity arose to commence a preliminary analysis on 119 sets of data in the change sample. It revealed a large number of respondents using the NS category for many items. Indeed, problems of a measurement nature were now obvious if responses to this category were left in the data sets, especially as this category was not used necessarily for a neutral response rating.

One advantage of commencing the data analysis before all data were collected was the ability to gain an insight into the behaviour of certain items and the overall fit to the model. One result, the problem of including the NS category response as part of the measure of attitude towards Direct Instruction, was apparent. Consequently, when data were collected from the remaining 25

participants in this sample group, feedback as to the role of the neutral category was sought. Three main reasons for a NS category response were given. First, uncertainty as to the meaning of Direct Instruction; second, irrelevance of the question to their situation; and third, ambiguity of the question. Use of the NS response category to indicate a neutral attitude did not appear as an alternative.

One way of examining the influence of this category on the quality of measurement is through an examination of threshold order across the items. The stategy involves suppressing all responses to the NS category by creating missing data in the data records and then comparing the effect on the threshold order both before and after the replacement. This procedure is possible because the estimation algorithm used for the extended model is capable of handling missing data routinely.

It is important that the nature of the missing data, deliberately created as part of the design, be assessed thoroughly to ensure that the validity of the remaining scores is sound, even though sufficient responses are available for the remaining items to provide reasonably precise estimates of the item and person parameters. Because the NS responses were eliminated *after the data had been collected*, it would be provident to ensure that this technique is justified from a measurement perpective by examining responses to these same items with the NS not provided as part of the response sequence; that is, eliminating (or suppressing) the NS category *prior to respondents rating these items.* A control sample was therefore employed to investigate if the strategy of eliminating NS responses after the event (as for the change sample) is a legitimate means of calibrating the measuring instrument.

Control Sample

The control sample consisted of 275 third year education students at Edith Cowan University and a small number (14) of post service teachers. The pre-service teachers had all taught for at least ten weeks during their final practicuum and were two weeks away from graduation. The participants could be considered similar to many of the teachers included in the change sample, who were in the first year of service. An advantage of their inclusion in the sample was the easy accessibility for the collection of data. Further, the participants had had varying degrees of exposure to Direct Instruction.

For this sample, only the posttest occasion was invoked, with the questionnaire being completed after viewing the video tape. However, two different forms of the questionnaire were used. Approximately half of the sample were asked to attempt all items without access to the NS category, using the four normal response categories of Strongly Agree, Agree, Disagree, and Strongly Disagree. This provided data in which the NS category *is removed before the event.* The other half were provided with a NS response category placed to the right of the normal four response categories and asked to use the former response category only if absolutely necessary. Comparisons between the pretest and posttest occasions for the change sample, and across the change sample and the control sample would then provide evidence of the effect on the measurement process of the NS category.

Data Collection

Data were collected in two stages. The change sample data were collected over a period of two months. Staff from three primary schools participated in the study on Staff Development days. Other participants completed questionnaires in their homes or while attending classes at Edith Cowan University. In each

case, the researcher introduced the study and presented instructions regarding completion of the questionnaire. In order to assist in the item analysis, and in an effort to understand the variable of interest, participants were encouraged to make comments against any item they felt to be unclear or ambiguous.

Participants in the control sample watched the video and completed the questionnaire during lectures or tutorials at Edith Cowan University. In most cases, the researcher introduced the video and gave instructions regarding the completion of the questionnaire. For the remainder, a tutor supervised the completion of the questionnaires after being briefed by the researcher, and being provided with an instruction sheet detailing the completion of the questionnaire.

CHAPTER FIVE
INSTRUMENTATION

The analysis of the data collected for this study is in two main stages. The first, presented in this chapter, involves the calibration of the measuring instrument. This stage will employ the measurement model described in Chapter Three and provide knowledge of the variable, Attitude towards Direct Instruction based on the conceptual framework presented in Chapter Two. The second stage, covered in Chapter Six, deals with the assessment of attitude change related to the level of knowledge of the teaching strategy, Direct Instruction.

The first task addressed in this chapter is an examination of the role of the 'Not Sure' (NS) category in the construction of the measuring instrument. Then follows a detailed description of the calibration of the instrument, including an assessment of its use as a commensurate measure for the repeated measures analysis to be described in Chapter Six.

The 'Not Sure' Category

A preliminary examination of responses for the change sample revealed that the NS category attracted considerably more responses in the pretest than in the posttest, which indicated a greater commitment to rating the items was present for the posttest than for the pretest occasion. This confirmed that the first step in the item analysis must involve an assessment of the NS category. It is known that this type of response category can cause problems with measurement because a response in this category is often due to reasons quite different from the trait under consideration. A background discussion relating to why this category was employed, and the logic relating to the use of the 'change' sample and 'control' sample in this strategy was provided in Chapter Four. Attention is

now focussed on an assessment of this category in relation to measurement as a preliminary to the main calibration process.

'Not Sure' as a Rating Category

The first step in the item analysis involved the two data sets for the pretest and posttest occasions of the change sample where the NS category was employed as the middle category. An examination of the threshold estimates showed a large number of items exhibiting reversed thresholds. A summary listing of items with threshold reversal for each data set is displayed in Table 3, while the details for each item are listed in Table B–1 in Appendix B. This meant that the distribution of responses was not in a logical order as conceptualised for this

TABLE 3
Items with Threshold Disorder by Sample Data Set

Sample	Items with Threshold Order reversed	
Date Set	NS category included	NS omitted
Change (pre)	1, 4, 7, 9, 13, 14, 30, 35	1, 30
Change (post)	1 4, 7, 9, 10, 12, 13, 14, 16, 18, 20, 22, 23, 26, 30, 33, 38,	4, 14, 30
Control (with NS)	** no relevant data **	14, 26, 30, 39
Control (no NS)	** no relevant data **	9,

measure. As discussed in Chapter Three, disordered thresholds need to be investigated as a priority. It was suspected that the NS category played a major part in such reversals, over and above any contributing factors such as ambiguities or difficult words or phrases associated with any particular item.

If the NS category operated in a manner consistent with the other categories, then the probability of a response in the NS category for any item would form part of a logical relationship across all categories. In particular, the NS category

should be neither over-represented nor under-represented as any disproportion amongst categories can contribute to the reversing of thresholds (Andrich,1985, 1988; Andrich & Schoubroeck,1989).

'Not Sure' Category as Missing Data

The second step in this preliminary item analysis was to evaluate the strategy of converting NS scores to missing data. This involved suppressing the scoring value for that category and replacing it with a blank space. All responses scored as three (the NS category score) were made blank and scores of 4 (Agree) were recoded to 3, and scores of 5 (Strongly Agree) recoded to 4. The end result was a data file containing response codes from 1 to 4 to represent the sequence Strongly Disagree, Disagree, Agree, and Strongly Agree. The missing data then represented items where participants were unable to make a decision within the sequence Strongly Disagree to Strongly Agree. As explained in Chapter Four, the extended model of Rasch is well suited to handling missing data in a routine manner.

The item analysis was then re-run on the two revised datafiles and the threshold order re-examined. A dramatic change in the behaviour of the threshold order was observed, as the summary listing in the first two rows of Table 3 reveals and as detailed in Table B-2 of Appendix B. The number of items exhibiting reversed thresholds had now reduced to between two and three only in a set of 40 items.

An item analysis was next undertaken on the control sample, where the NS category was either *not* employed (about half of the sample) or when responses to the NS category (the other half of this sample) were suppressed as missing-data. A similar low number of items was found to exhibit threshold reversal for the control sample as for the change sample. Rows three and four of Table 3,

and Table B-3 of Appendix B, provide details in this regard. Because comparable outcomes have resulted from a manipulation of the NS category both before and after the event, the technique of replacing scores in the NS category with missing data, and using the responses from the remaining four categories to calibrate the instrument, is justified. The small number of items exhibiting threshold reversal can now be examined as part of the more familiar item analysis strategy associated with the item and person tests-of-fit described in Chapter Three. A more detailed description of the measurement implications of this NS category using the strategy employed is available in Sheridan (1993).

When the NS category was converted to missing data, it was noted that some respondents had very few scores remaining in the Strongly Disagree to Strongly Agree categories across the 40 items. A decision was made to eliminate from future analyses, the data of those respondents who had less than 20 responses in their data set. This procedure resulted in the removal of 25 records from the pretest change sample, but only 3 from the posttest change sample. Upon closer examination, all of the teachers eliminated were from Group One, those who had never seen Direct Instruction. This outcome is consistent with the logic of the situation in that many participants were unable or unprepared to comment on Direct Instruction prior to the viewing of the video. This procedure ensured that the precision of the estimates obtained in subsequent analyses was not compromised by using records containing too many missing data points while at the same time it has been possible to accommodate genuine 'not sure' responses to the remaining more general items. In this way efforts to capture sufficient and reliable information from Group One participants to examine for change between pretest and posttest occasions would appear to have succeeded.

Calibration of the Instrument

To undertake a calibration of an instrument using the analytical procedures that follow in this chapter, it is critical that the calibrating sample reflect the frame of reference as specified by the conceptual framework. The calibration sample consisted of 409 teachers, comprising the change sample (posttest) and the control sample. The participants sampled for this study were selected from teachers possessing differing levels of experience with the teaching strategy of interest, Direct Instruction. As a significant number of respondents in the pretest of the change sample had not seen the method, this meant that a number of items could not be rated at this stage. In order to maximise the precision of item estimates, it was decided that the pretest data of the change sample would not be used in the calibration of the items. However, the change sample (posttest) was suitable for this purpose and would also provide an anchor between analysis used solely for calibration of the items and analysis assessing change.

All 40 items of the instrument were submitted to an item analysis and the fit of the data to the measurement model examined for items and persons not fitting the model. These tests-of-fit, discussed in Chapter Three, are the item-trait and the item-person interaction tests. The item-trait interaction test-of-fit uses a chi-square statistic to assess the probability of a chance occurrence of the degree of divergence between observed and expected values for persons located along the latent trait continuum. Individual statistics for each item are provided, as well as an overall test-of-fit. The item-person interaction fit statistics, on the other hand, provide information relevant to each person response pattern and how this accords with the Guttman pattern in association with item order along the latent trait continuum. The statistics approximate a t-distribution when the data fit the model. Therefore the overall distribution for the item or person statistics should have a mean of zero and a standard deviation of

one. Any individual item or person statistic whose value lies outside a specified limit indicates a source of misfit to the model.

When examining item fit, often the behaviour of one particularly bad item can contaminate the behaviour of other items. The removal of such items often allows the remaining items to exhibit confirmation to the model consistent with the conceptual framework. The test-of-fit statistics and threshold estimates for all forty items are listed in Table B-4 of Appendix B. By considering the accumulated evidence from the various fit statistics, the suitability of each item to remain in the analysis is assessed.

In a similar manner, some persons in the sample may contaminate the results through the presence of extemely unlikely responses to some items or worse, by responding in a mischievous, haphazard way overall. These people may respond in such a manner due to a poor response set to the situation, feeling tired or unconcerned about the questionnaire. The complete response set for a person who responds in an extreme manner across a number of items is usually removed from the data file during the calibration stage. Other persons may simply respond in an unlikely or unusual manner to isolated items, which may be problematical anyway. In this case, the response to the particular item is suppressed by creating a missing data entry as described earlier for the NS category and as elaborated further in Sheridan (1993). It should be stressed at this point, that the removal of extremely unlikely responses to specific items, or removing some persons from the calibration sample completely, is to improve the precision of the calibration of the instrument. Table 4 shows the mean and standard deviation of the person fit statistics for the set of 40 original items where item misfit was evident, compared to the 21 items which did fit the model.

TABLE 4

Mean and Standard Deviation of each distribution of the Person Fit statistics for the original sample and when the extremely unlikely responses are removed for both the original (40 items) and final version (21 items) of the ADI Survey

Statistic	All 40 Items		Final 21 Items	
	Original Form	Edited Form	Original Form	Edited Form
Mean	-0.69	-0.44	-0.69	-0.24
Std Devn	2.31	1.48	1.73	0.98

The effect of extremely unlikely responses by participants can be observed from these two sets of statistics, where the situation for the 21 items is now close to the prescribed t-distributiion. Attention can be focussed at a later stage on those persons exhibiting unlikely responses, to determine the nature of their misfit, including resorting to interviewing each person if possible. By this means, additional insights can be obtained into the meaning and nature of the variable under construction.

A form of step-wise analytic procedure is thus undertaken between items and persons exhibiting misfit to the model until a set of items and persons remain who provide evidence of the data conforming to the strict requirements of the model. At the same time, the primary aim of constructing knowledge of the variable in a meaningful way must be satisfied by examining the contents of items misfitting the model and also the items that remain and comprise the measurement continuum. A number of items and persons were subsequently eliminated culminating in a set of 21 items that together demonstrated an acceptable level of fit to the model. The fit details and threshold estimates are displayed in Table 5. The overall test-of-fit indicates the degree of consensus displayed collectively by all items of the instrument across persons located at differing attitude levels. For this set of items, the overall test of fit for item-trait interaction had a probability of no misfit of 0.19 and, apart from item 29, the

TABLE 5

Threshold Estimates and Item Fit for Attitude to Direct Instruction under the hypothesis of no misfit for 21 items (N = 382)

Item	Item–trait Interaction**		Standardised Residual	Threshold Estimates		
Code*	χ^2 (5 df)	Pr(χ^2)	(361 df)	1	2	3
CP01	6.43	0.24	0.15	–3.65	–0.28	3.92
CP03	3.19	0.66	–0.27	–3.70	–0.10	3.80
UR05	1.95	0.85	–0.73	–4.59	–0.35	4.94
NP06	1.86	0.86	–0.26	–3.22	–0.13	3.36
CR12	4.35	0.49	0.13	–3.59	–0.59	4.18
LP15	2.40	0.79	–1.53	–3.77	–0.01	3.78
NP16	1.50	0.91	–0.03	–2.18	–1.05	3.23
UR17	8.42	0.11	–0.41	–2.52	–0.91	3.43
UP19	5.02	0.40	1.03	–3.51	–0.67	4.18
NP21	4.11	0.52	–1.80	–4.10	–0.84	4.94
CR22	3.53	0.61	0.50	–3.83	–0.95	4.78
NR23	5.19	0.38	–0.35	–4.16	0.20	3.96
NR26	1.89	0.86	–0.10	–3.74	0.19	3.55
NP27	8.06	0.13	0.83	–3.85	–0.02	3.87
NP28	4.41	0.48	–1.96	–4.20	–0.31	4.51
NR29	19.68	0.00	1.63	–3.18	–0.99	4.16
NP31	4.61	0.45	–2.73	–2.96	–1.83	4.79
CR32	7.68	0.15	0.18	–3.67	–1.64	5.31
LP33	7.58	0.16	0.23	–3.93	–0.20	4.13
NP36	5.52	0.34	–1.87	–3.62	–0.44	4.06
CR40	5.14	0.38	–0.35	–3.38	–0.81	4.18

*Key for Labels: 1st character – N = Need; L = Clarity; C = Cost; U = Other.
2nd character – P = Positive wording; R = Negative wording.
Numbers refer to items in the original 40 item ADI Scale.

**The overall χ^2 is 112.5, which has a probability of 0.19 on 100 degrees of freedom.

individual items fit the model well. For the item–person interaction, or standardised residual, test-of-fit, only item 31 had a statistic outside the nominated range of –2 to +2. Also, the threshold estimates are correctly ordered for all items, demonstrating that the model distribution for the threshold estimates is unimodal. Finally, the Person Separation Index of 0.95 indicates that this

instrument has a high resolution for separating the teacher attitude estimates along the continuum and that the power of the tests-of-fit is also high.

Person range

Generally, the sample has a broad range of attitude and is well targetted as the distribution of person and item location estimates in Figure 1 reveals. The

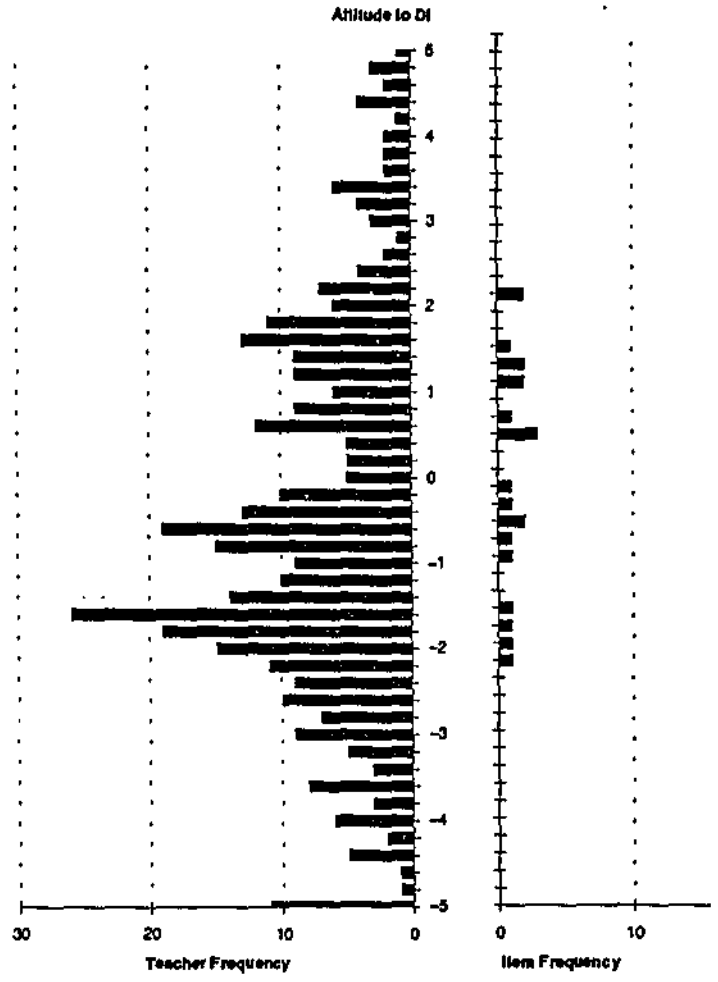

FIGURE 1: Distribution of Teacher Attitude and Item Location Estimates for the ADI Scale (N = 382)

sample is reasonably centred around zero, with a slight bias towards the negative end, as an overall mean of –0.24 and a standard deviation of 0.98

confirms. This means that the test, on average, is a little difficult for the participants to agree with. In addition, a considerable tail is noted with participants from both extremes of agreement and disagreement with the strategy well represented. In selecting the sample, efforts were made to include teachers with a range of exposure to Direct Instruction. According to innovation theory, teachers unfamiliar with an innovation are likely to be fairly antagonistic towards it. At the same time, teachers familiar with a teaching strategy are more likely to be supportive of it. As Figure 1 indicates, the sample has a sufficient range of attitudes to provide the level of variance required for the experience groups when assessing attitude change to Direct Instruction, and presented in the next chapter.

Item range

As Figure 1 demonstrates, the items are spread over a range exceeding four logits which is a good result for attitude type measures.

Construct validity

The 21 items identified derive from two sources, 15 from Proctor's (1989) questionnaire, and six developed especially for the study. As these items are located on the same measurement continuum and created in different ways from the same conceptual framework, evidence of construct validity has therefore been established.

Teacher attitude towards Direct Instruction

It is now possible to gain a greater understanding of what characteristics of Direct Instruction contribute to a measure of teacher attitude towards this teaching strategy. The set of 21 items identified in Table 5 can now be

considered to define the variable. An examination of item content and item arrangement on the continuum is now considered.

Location of items on the continuum

An examination of the location of the 21 items on the latent trait continuum provides information on the nature of the variable created. The display in Figure 2 lists each item by its specification code, number location in the original 40 item scale, and an abbreviated statement of the content. This same information can be rearranged and displayed according to the specification code only, where the first letter indicates which of the innovation characteristics of need ('N'), clarity ('L'), or cost ('C') the item is describing; a 'U' specifies those items for which one of these classifications was not possible. Figure 3 provides a segregation of the same 21 items by innovation characteristic, this time displaying the code only. A similar distribution is possible for the positively and negatively (or reverse scored) worded items where the second letter in the code, 'P' and 'R' respectively, would be used and is reproduced in Figure C--1 in Appendix C.

The items that participants found easiest with which to agree relate to the effectiveness of Direct Instruction (Items 1and 31) and the ease with which it can be implemented (Item 32). Item 19, which relates knowledge about Direct Instruction to support for the method is also one of the least provocative items. On the other hand, the most provocative items relate to the teacher directed (Item 12) and structured nature of the method (Item 26). Many participants found it extremely difficult to agree that these aspects of the method are advantageous. It appears that while many teachers consider Direct Instruction to be an effective teaching strategy, its highly structured nature would act as a deterrent to its implementation. The overall location of the items on the continuum seems logical given the observations and comments of researchers

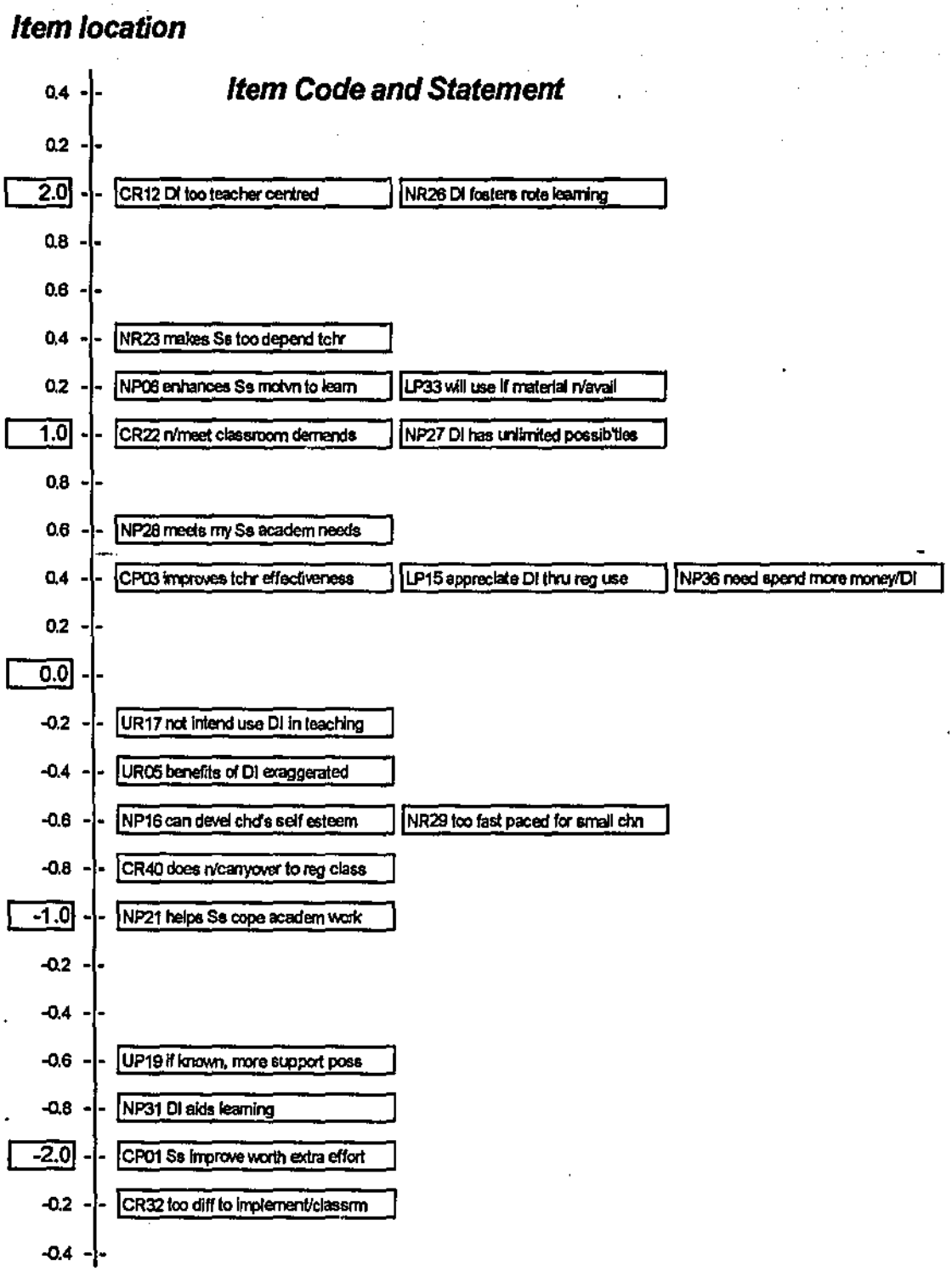

FIGURE 2: Distribution of Items along the Attitude to Direct Instruction Continuum displaying innovation code and statement content

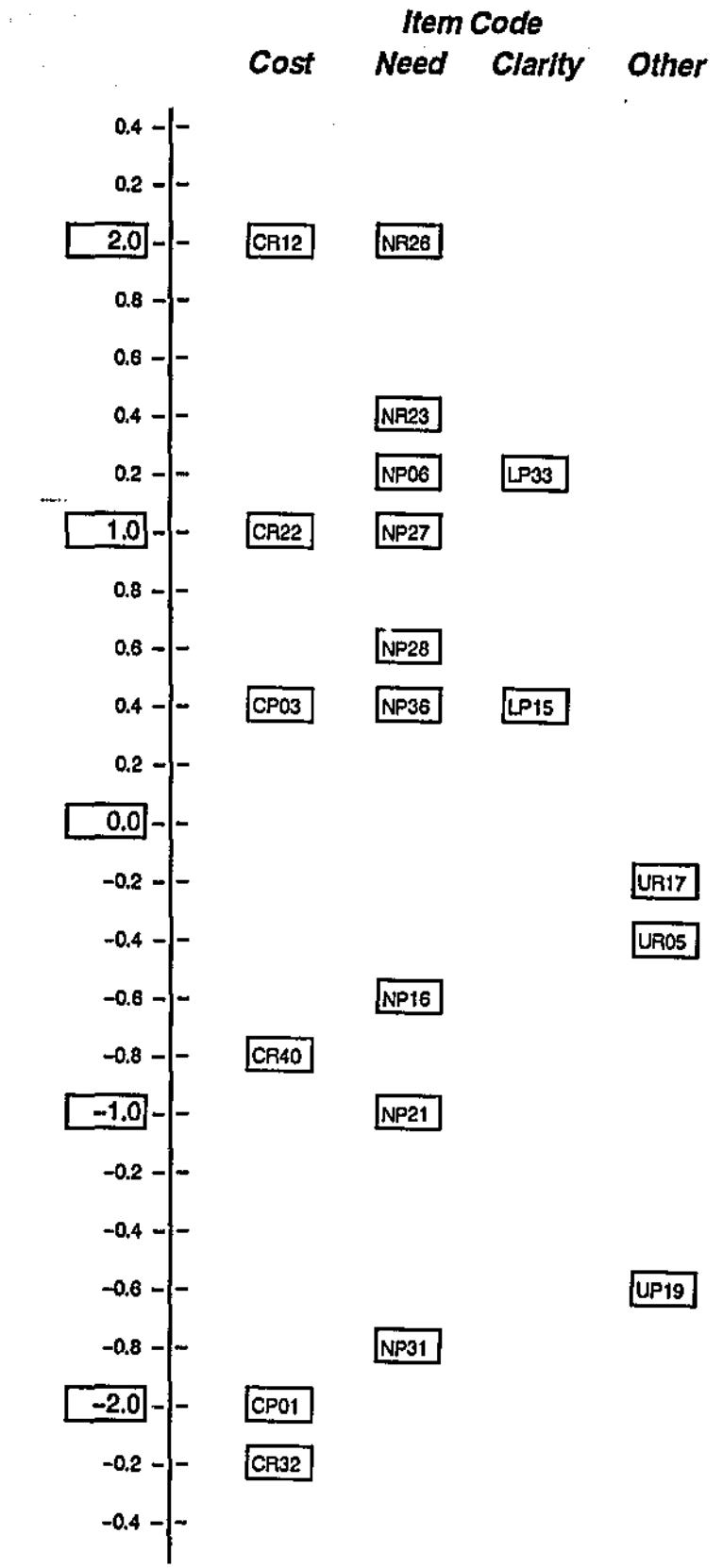

FIGURE 3: Distribution of items along the Attitude to Direct Instruction Continuum arranged by innovation characteristic code

as reported in the literature and noted in Chapter Two, in addition to comments made by teachers participating in this study.

With respect to the distribution along the measurement continuum, it would appear that items stated in the positive form are generally easier to agree with than those possessing a negative orientation. Figure 2, and Figure C–1 in Appendix C, display details of this trend. The three most difficult or prococative items on the continuum are worded negatively, whilst the majority of items clustered towards the less provocative end of the continuum are worded positively.

There does not appear to be any differential trend or clustering of items regarding the location along the continuum of one innovation characteristic in preference to another. As Figure 3 reveals, both less provocative and more provocative items appear representative of all three characteristics. However, whilst it is difficult to reach any firm conclusions based on two items only, items relating to clarity appear to be more difficult with which to agree.

Items removed from the Scale

An examination of the statements for items misfitting the model provides additional insights into the nature and meaning of the variable, Attitude to Direct Instruction, where several trends are apparent. Items 9 and 35 refer to the emphasis of Direct Instruction in teacher training, and as Direct Instruction is not taught during most teacher training courses, many participants found these two items difficult to answer. Consequently, answers were inconsistent and attracted many qualifying comments. Where the NS response category was available, a large proportion of participants answered in this category for these problem items. A further observation is that while the two items are similar in content, where an element of feeling, such as "I am glad ..." is involved (as with Item 35),

then the item became more difficult with which to agree. This problem would appear to occur through a mismatch in the target between items and persons and that the items did not necessarily reflect the frame of reference for the setting of this variable.

In several cases, the wording of an item appears to have contributed to the misfit. All items which include words referring to emotion were misfitting, and in these cases, were found to be overdiscriminating, which lead to responses that were *too* predictable or too easy to agree to -- a sort of self fulfilling prophecy! The three items involved were; Item 35, ("I am glad ..."), Item 24 ("I am looking forward ..."), and Item 39 ("I hope ..."). Two of these items, 24 and 39, were also expressed in the future tense, a possible further contributory factor to the misfit.

One example of a poorly worded item demonstrating misfit is

Item 30: "All teachers should be prohibited from using Direct Instruction."

The word 'prohibited' is highly emotive and attracted many qualifying comments from participants. Other misfitting items had wording that may be interpreted in a variety of ways. Examples are:

- Item 34: "Direct instruction creates a positive attitude in the classroom."

  Does it refer to the teacher's attitude or the child's? A positive attitude towards what? and

- Item 13: "Direct Instruction is boring."

  Boring for whom? The teacher or the child?

Another item:

- Item 4: "Direct Instruction is primarily beneficial in 1:1 and small groups."

was written for Proctor's (1989) survey as a negatively worded item. In his study, this item did not perform as he expected. A possible reason for inconsistent responses to this item could be related to teacher preference of

teaching style. Not all teachers may see the use of Direct Instruction in small groups as an advantage.

Defining Attitude towards Direct Instruction

By examining item content and item location on the latent trait continuum, it is possible to describe the characteristics of the variable and provide meaning to just what it is that Attitude towards Direct Instruction is measuring.

The characteristics of Direct Instruction that are less provocative, and that most teachers would generally agree with, relate to the method's effectiveness as a teaching strategy and ease of implementation. The need to know more about Direct Instruction was acknowledged as important, as would be the case with any new teaching strategy. The more provocative aspects of Direct Instruction, with which only the most ardent supporters would agree, relate to the teacher-oriented, repetitive nature of the approach. These aspects of Direct Instruction are also the more noticable when observing the method for the first time.

Before concluding this first stage of the analysis devoted to item calibration, it is necessary to check that commensurate measures are present for both the pretest and posttest occasions. This will involve an examination of the 21 items selected for the instrument, but this time focussing only on the change sample.

Commensurate Measures for the Pretest and Posttest

To establish if commensurate measures are present across two occasions, it is necessary to demonstrate that the item location estimates remain invariant for the two samples involved. In addition, if three or more categories are involved in the item scoring, then it is important to ensure that the scale parameter estimates for all items and for both samples are equal. Otherwise, direct

comparison of the item location estimates across the two occasions would be affected. This is a direct result of the interrelationship between the hierarchy of item parameters, as described in Chapter Three, and the distance between thresholds which in turn influences the value of the location estimate as an average value across all category values.

One of the features of the ASCORE program employing the extended model is the facility to restrain the scale parameter to equal value for all items. This process can result in a loss of information, especially if the contribution of the higher order (*skewness*) parameter is significant. However, a greater loss would be entailed if *unequal* scale parameters are employed. When the equal scale option was employed for the pretest and posttest occasions, average scale estimates of 1.40 and 1.36 respectively were obtained. These values were sufficiently close, within the range specified by the standard errors (0.02) to satisfy the requirements that the scale values present for both occasions were equal. The location values could now be compared directly using the graphical procedure described in Chapter Four and elaborated upon in the next chapter.

The plot of item location estimates for both pretest and posttest occasions is presented in Figure 4. In an ideal situation, all item location pairs would lie along the hypothesised straight line which is inclined at 45 degrees to the axes and passes through the origin of both axes. Because a probability model is involved, limits must be established on either side of the hypothesised line to account for chance fluctuations that are acceptable within the limits as specified. For this study, a limit was set at the 5 per cent level of confidence.

As the plot in Figure 4 reveals, most items are located around the hypothesised central line, and within the confidence limits as set, thus providing evidence of a general lack of item bias across the pretest– posttest occasions. However, three

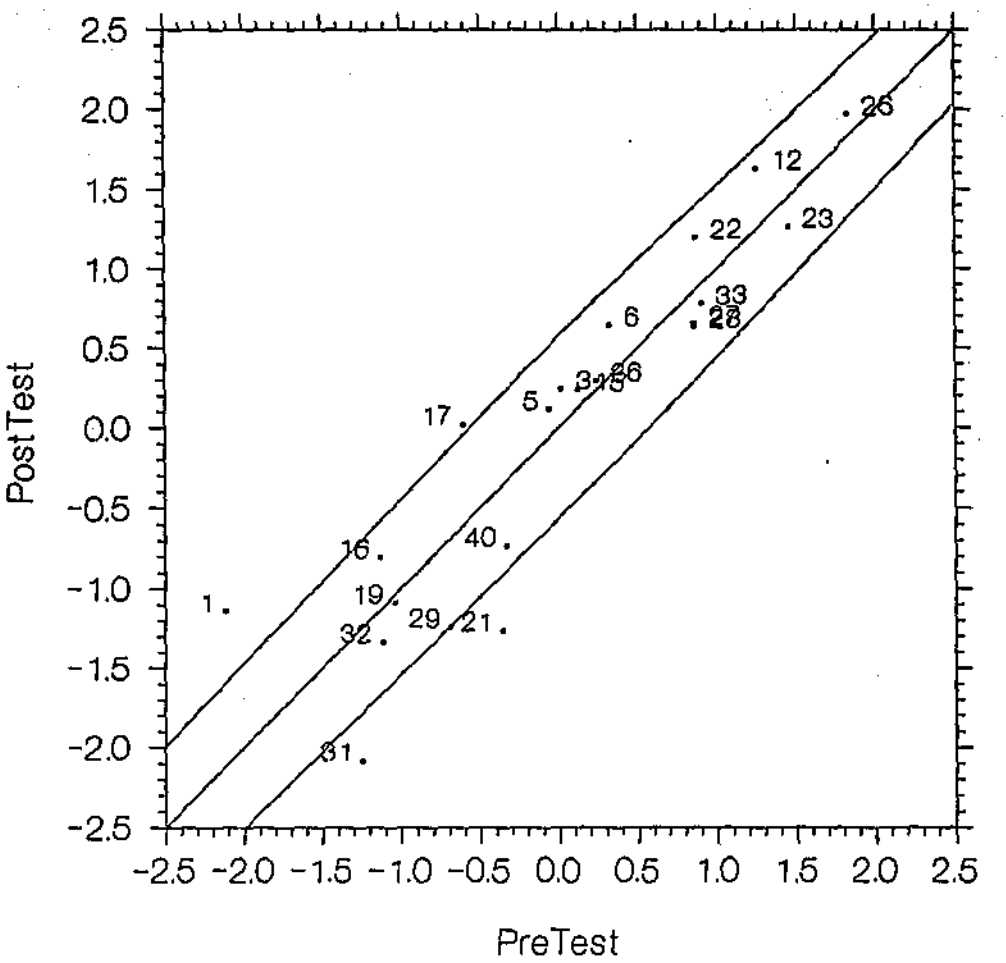

FIGURE 4: Plot of Item Location Estimates (in logits) for the Pretest and Posttest for the ADI Scale

items lie outside this band, and need to be reassessed before proceeding with stage two of the analysis. Items 21 and 31 both revealed a decrease in difficulty, *relative to the other items collectively*, in moving from the pretest to the posttest. Conversely, Item 1 became more difficult to agree with, *relative to the other items collectively*, in the posttest. Attention is now directed to an examination of the content of these items in an attempt to better understand the nature of the bias and to gain further insights into the meaning of the variable, Attitude towards Direct Instruction.

Consider the wording of the items whose affectivity decreased for the posttest occasion:

Item 21: "Direct Instruction helps students cope with academics."

Item 31: "Direct Instruction aids learning."

The large number of participants who had not previously seen Direct Instruction would probably be unsure as to what the method involved. It appears that after viewing the video many participants were able to agree that Direct Instruction did assist children in learning. Also, for the pretest occasion the wording of the items may have been a problem. These items are, upon reflection, loosely worded and could be interpreted in a number of ways: What does 'academics' cover? How do you define 'cope'? 'Learning' what?

Both items are similar in content and represent the 'need' characteristic of an innovation. This suggests that many participants were able to agree that Direct Instruction was an effective teaching strategy as both items were located towards the less provocative end of the continuum. After viewing the video, the perceived personal need for using Direct Instruction was increased in a differential manner. This trend is supported by the observation that Item 29, also representing the 'need' characteristic of an innovation, was only marginally within the confidence band, having also become an easier item with which to agree in the posttest occasion.

The remaining item exhibited an increase in affectivity for the posttest occasion. This item represented the 'cost' characteristic of an innovation.

Item 1: "Student improvement is worth the extra effort of using Direct Instruction."

Participants in the pretest occasion, particularly those who were not familiar with Direct Instruction, appear unsure as to the effort required by a teacher in implementing the approach. After viewing the approach in action, many participants found it much harder to agree with the view that the extra effort of using Direct Instruction is worth potential student improvement. Alternatively,

participants may have felt that the implementation of Direct Instruction involved a disproportionally greater cost than was initially conceived.

Following an appraisal of these three items, it was decided to remove items 1 and 21 from the final instrument, but retain Item 31. Item 21 appeared more awkwardly worded than Item 31, and as the location of Item 31 was at the less affective end of the continuum for the final set of items, the latter was retained to maintain as large a range in affectivity values as possible. In addition, Items 21 and 31 were both related to learning aspects of the strategy, so retention of one of these was considered desirable to maintain the depth of range and content of items along the continuum. The final form of the scale used to examine change in attitude for the different experience groups therefore comprised 19 items.

It is now possible to address the first research question.

Research Question 1: What are the psychometric properties of the Attitude towards Direct Instruction Instrument as developed for this study?

The final scale of 19 items has a high overall test–of–fit ($p = 0.74$), demonstrating a satisfactory degree of consensus between all items across persons located at different attitude levels. The Index of Person Separation statistic of 0.95 indicates the instrument has high reliability and that the power in the tests–of–fit is high. Threshold estimates for all items are ordered, hence the model distribution is unimodal. Thus, it can be confidently stated that the instrument constructed to measure teacher attitude towards Direct Instruction has sound psychometric properties.

In the next chapter, this final instrument comprising 19 items will be used in stage two of the analysis to assess the attitude change of participants as a result of exposure to the teaching innovation of interest, Direct Instruction. In addition,

the attitude of participants grouped according to degree of exposure to, or experience with, Direct Instruction will be examined.

CHAPTER SIX
ATTITUDE CHANGE

In the previous chapter, the calibration of the Attitude towards Direct Instruction instrument was reported and meaning ascribed to the variable to be measured. This instrument was also shown to provide commensurate measures for the pretest and posttest occasions. In this chapter, the attitude measures available are used to investigate the change of teacher attitude as a result of knowledge of the teaching strategy, Direct Instruction. Of particular interest is the relationship between experience with Direct Instruction and teacher attitude. In so doing, the research questions stated in Chapter Two will be addressed and hypotheses generated from these questions tested.

Attitude Measures

Before commencing the analysis for this second stage, the pretest and posttest data were pooled, the NS responses suppressed and replaced by missing data. As described in Chapter Four, this action creates two records for each person within the one data file, one record for the pretest occasion and one for the posttest. This technique takes advantage of the special properties of the measurement model to account for commensurate measures and allow for an efficient method of obtaining person attitude estimates for both occasions. Each pair of attitude estimates for the two occasions are now directly comparable as they are derived from one set of items whose contribution to a unidimensional measure of Attitude towards Direct Instruction has been established. The attitude measures can then be selected in accordance with the individual groups specifying different levels of knowledge and experience towards Direct Instruction.

Assessing Change in Attitude

The research questions posed in Chapter Two provide the focus for stage two of the analysis devoted to attitude towards Direct Instruction. Each research question is capable of generating hypotheses that can be tested.

The same graphical procedure described in Chapter Four, and again in Chapter Five for testing item bias, will be employed for testing the hypotheses associated with change in attitude. In the present case, the attitude measures for pretest and posttest occasions replace the two sets of item location estimates. If no change in attitude is present between the two occasions, the hypothesised line will pass through the origin, as was the case for the item bias situation. Otherwise, if a change in attitude has occurred between the pretest and posttest occasions, the hypothesised line will shift away from the origin. The direction of the attitude change is indicated by the shift along one of the axes from the origin, and the size of this change is determined by the distance moved away from the origin. It should be noted that this plot shift relates to the mean change in attitude for the group concerned.

Research Question 2: To what extent does a brief demonstration of Direct Instruction affect teacher attitude?

In Chapter Two, it was observed that researchers investigating Direct Instruction found that teachers only observing and not teaching the method usually expressed a negative response. They reacted to the "excessive structure and regimentation" (Mathes & Proctor, 1988. p. 97) of the strategy. In consideration of this observation, the question generates the following hypothesis:

Hypothesis 1: Teachers will respond in a negative manner to a brief demonstration of Direct Instruction.

This hypothesis is tested by examining the plot of attitude measures for both occasions for all teachers. A general plot shift relative to the origin has occurred,

as Figure 5 reveals, and the direction moved by the hypothesised identity line indicates that the set of items has become more difficult to agree with for the participants in the posttest relative to the pretest occasion. This means that the

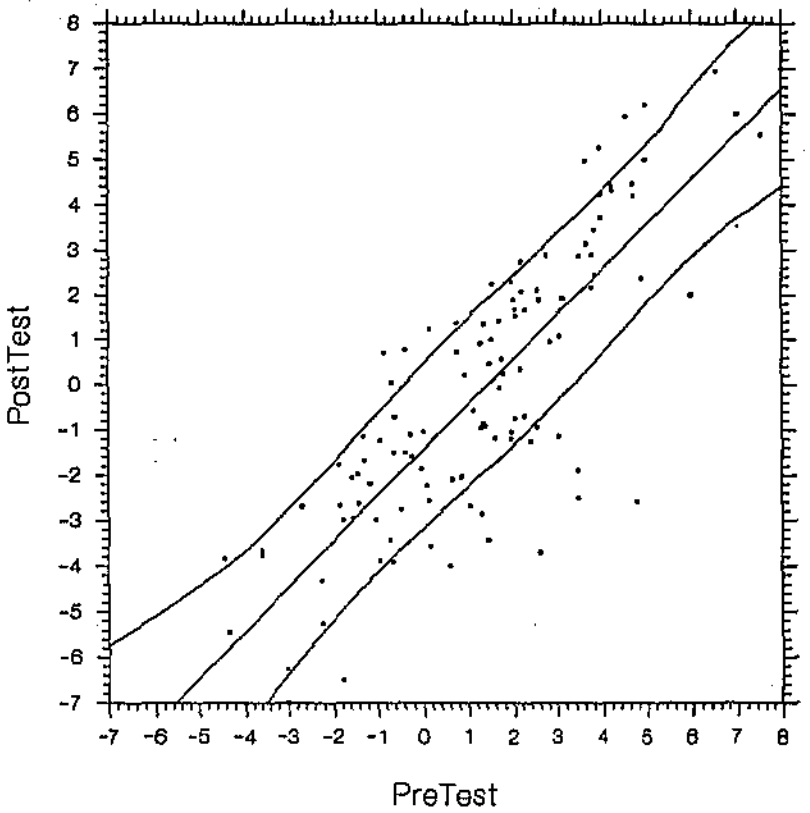

FIGURE 5: Plot of Teacher Attitude Estimates for the Pretest and Posttest for the ADI Scale for all experience groups combined (N = 109)

attitude to Direct Instruction is lower on average on the posttest occasion. Therefore , a change towards an increasingly negative attitude has occurred in teachers following a demonstration video and the hypothesis is supported.

Participants located outside the confidence bands have demonstrated a behaviour significantly different from that expected relative to the majority of teachers. This behaviour needs to be examined on an individual basis, and the outcomes can usually provide diagnostic information on the person as well as information relevant to the meaning of the variable with the attendant implications this can have for the theory involved. The approach adopted for this

study was to consult any written comments teachers provided at the time of data collection and this will be addressed later in this section.

While a change in attitude is evident overall, it would be instructive to examine the nature and extent of this change for the three different experience groups specified in Chapter Four.

Research Question 3: To what extent does previous knowledge of Direct Instruction affect teacher attitude?

The participants were divided into three groups according to their degree of exposure towards Direct Instruction: Group One had never seen the method, Group Two had seen the approach but had never taught it, and Group Three had taught it.

According to innovation theory, teachers tend to be very cautious about any teaching strategy they have not put into practice. It is therefore to be expected that people who have never seen Direct Instruction would find it more difficult to agree with many of the more provocative items in the scale. On the other hand, teachers who have successfully implemented Direct Instruction are more likely to express a positive attitude towards the approach. Teachers who have taught any teaching strategy for a length of time have had a chance to observe potential student benefits and to gauge its compatibility with their personal teaching style. This question generates three hypotheses:

Hypothesis 2: The attitude estimates of teachers who have never seen Direct Instruction will be located towards the less provocative end of the continuum.

Hypothesis 3: The attitude estimates of teachers who have taught Direct Instruction will be located towards the more provocative end of the continuum.

Hypothesis 4: The attitude estimates of the teachers who have seen but not taught Direct Instruction will not show any distinct trend.

Separate plots of the pretest–posttest attitude measures for members of each of the three groups reveal that a differential shift in attitude has taken place. In Figure 6, the attitude estimates of Group One participants who had never seen

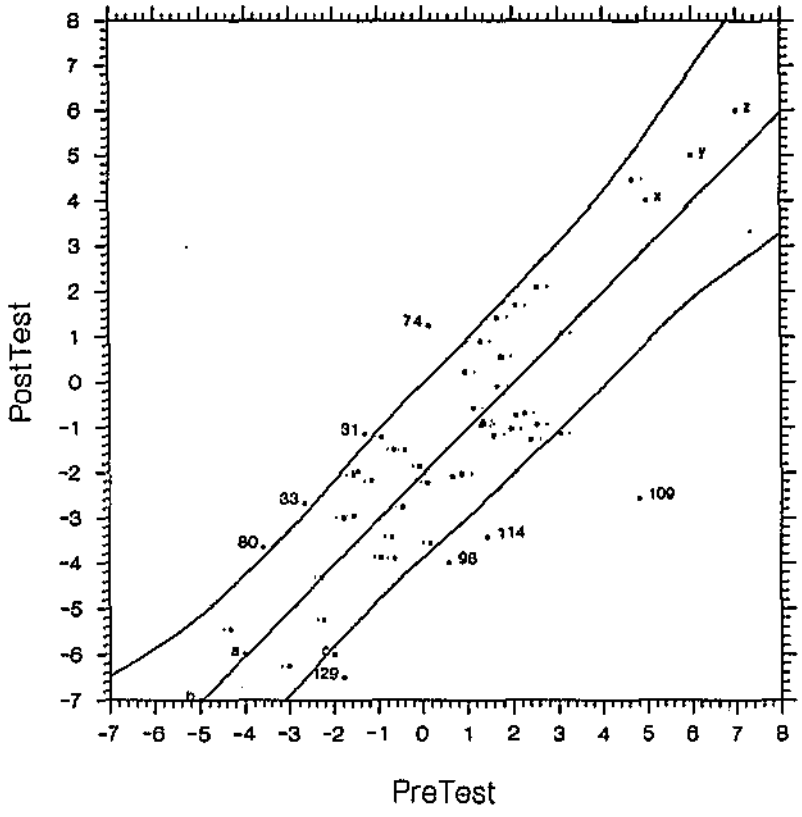

FIGURE 6: Plot of Teacher Attitude Estimates for the Pretest and Posttest for the ADI Scale for Group 1 teachers claiming no experience with the strategy (N = 50)

Direct Instruction prior to viewing the video tape are congregated toward the less provocative end of the identity line, indicating they are generally less supportive of Direct Instruction.

The Group Three attitude estimates in Figure 8, are located mostly towards the more provocative end of the continuum demonstrating that the teachers found the scale items easier to agree with, and can be considered to hold more positive attitudes toward Direct Instruction. The attitude estimates of the middle

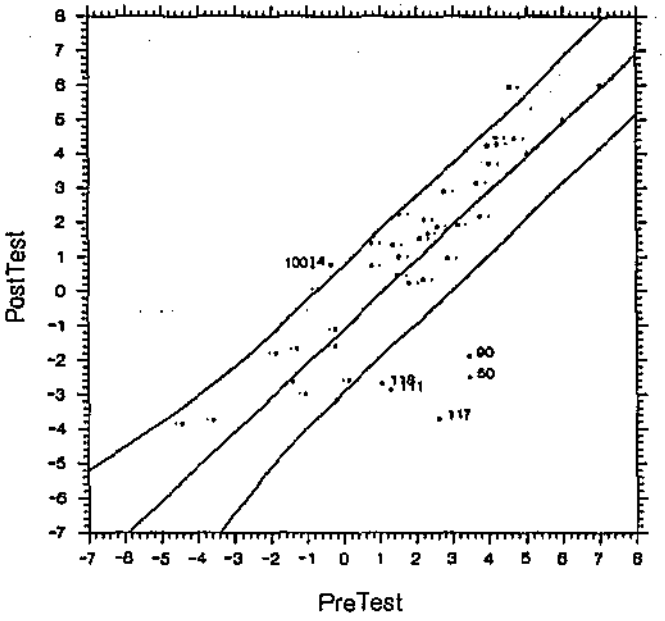

FIGURE 7: Plot of Teacher Attitude Estimates for the Pretest and Posttest for the ADI Scale for Group 2 teachers claiming limited experience with the strategy (N = 40)

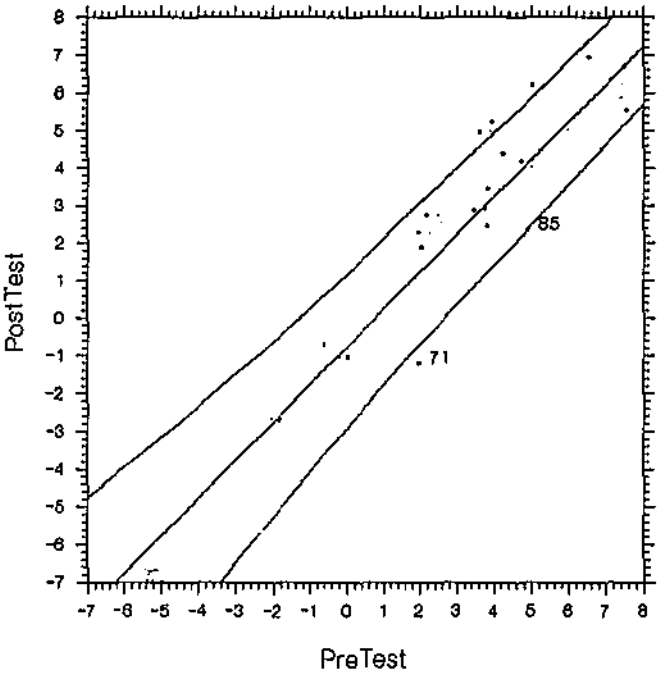

FIGURE 8: Plot of Teacher Attitude Estimates for the Pretest and Posttest for the ADI Scale for Group 3 teachers claiming considerable experience with the strategy (N = 19)

group, Group Two, shown in Figure 7, are located between the other two groups, with a larger grouping towards the more provocative end. Therefore all three hypotheses are supported by these data.

Research Question 4: To what extent does previous knowledge of Direct Instruction affect teacher attitude after a brief demonstration of the method?

This question addresses the degree of attitude change as a result of the brief demonstration of Direct Instruction. It is expected that Group one participants who are seeing Direct Instruction for the first time are likely to react to the apparent structure and teacher directedness of the method. Whereas Group Three participants would be less likely to be affected by a brief exposure, of any kind, to Direct Instruction. Their attitude has already been formed from their time spent teaching the method. Three hypotheses are tested:

Hypothesis 5: Group One participants will demonstrate a significant attitude change, the posttest becoming harder to agree with.

Hypothesis 6: Group Two participants will not demonstrate a significant attitude change.

Hypothesis 7: Group Three participants will not demonstrate a significant attitude change.

Group One participants show a very significant, negative shift in attitude, as demonstrated by the shift of the hypothesised identity line in Figure 6. It shows that the set of items became much harder for the participants to agree with on the posttest occasion, relative to the pretest.

In Figure 7, Group Two participants demonstrated a small negative shift in attitude with a greater number of participants acting in an unpredictable manner. Some degree of uncertainty is to be expected with this group, because many have had only a small amount of exposure to the method.

Group Three participants showed a very slight shift in attitude between the pretest and posttest, as shown in Figure 8. Their attitude towards Direct Instruction did not appear to be affected by the demonstration video. As a result of these examinations of the plots in Figures 6, 7 and 8, hypotheses 5, 6, and 7 are all supported by these data.

Comments for Extreme Individuals

Group One participants showed a significant shift in attitude in that the set of items became harder for the participants to agree with on the posttest occasion. Their written comments confirmed a strong reaction to the highly specific, teacher oriented aspect of Direct Instruction. For example:

- "... there is no room for discussion, exchange of ideas, questioning or creativity."
- "... rote, mechanical- ideal for learning block amounts of rote facts-robotic in nature."

The location of participants labelled 98, 109, 114, 129 in the plot of Figure 6 revealed that the posttest was significantly harder with which to agree than the pretest, over and above the average shift in attitude present. Their written comments suggest this extreme shift was because the four teachers had a different idea of Direct Instruction from that depicted in the video. For example:

- "Video was too long. My idea of 'DI' was quite different from the video. I would not personally use the method for a complete lesson."

Four Group One participants (31, 33, 74, 80) demonstrated a differential positive increase in attitude between pretest and posttest. Their comments indicated that they would like to know more about the method, and felt unable to comment more fully on a method with which they were unfamiliar.

The Group Two participants demonstrated a slight change in attitude with a great number acting in an unpredictable manner. Several participants found the

posttest-significantly harder with which to agree relative to the rest of the group. Their comments suggest their idea of Direct Instruction differed from that demonstrated in the video. For example, participant 111 commented

- "After watching the video I am left wondering just how much knowledge the children are left with ..."

On the other hand, participant 30, located at the more provocative end of the continuum, compared to the rest of the group, commented:

- "I believe that it is a very beneficial approach which should be used more often, especially in the regular classroom. More emphasis and teacher awareness training is required."

This positive comment is consistent with the participant's attitude, which was estimated as 2.1 logits.

While teachers belonging to Group Three demonstrated that their attitude generally towards Direct Instruction had already been formed and so change was minimised, two participants (71 and 85, as displayed in Figure 8) found the posttest much harder with which to agree than for the pretest, relative to the rest of the participants in this group. Their comments indicated that they felt the video did not depict Direct Instruction in a favourable manner. For example:

- "I feel that this approach works effectively with small groups of remedial children. It helps them to gain confidence in their ability. I feel that that video was not a very good example."

An immmediate advantage in seeking follow-up information in this way provides explanations of why persons exhibit extremely unlikely behaviour. It also provides additional insights into the nature of the variable under review.

These findings are consistent with studies relating prior knowledge and pre-existing attitudes to the acquisition of new knowledge (Hollingsworth & Reutzel, 1990). Lipson (1983) found that when text information was in conflict with existing values and knowledge, prior knowledge actually inhibited the acquisition of new knowledge. It follows that when the information is consistent with a

person's existing values and knowledge then they are more receptive. It stands to reason that Group One participants would probably experience more conflict between the information presented in the video and their personal experiences, than was the case for Group Three participants, and therefore, be less receptive to the teaching strategy.

This concludes the second stage of analysis which has assessed teachers' attitude change after a brief exposure to Direct Instruction. In the final chapter, the findings of this study are summarised followed by a discussion of the implications of the study for future research, theory, and practice.

CHAPTER SEVEN
SUMMARY, FINDINGS AND CONCLUSIONS

This study set out to investigate teacher attitude towards Direct Instruction, a teaching innovation that has not been widely implemented by teachers in regular classrooms. The two central aims were to investigate the psychometric characteristics of an Attitude towards Direct Instruction Scale (ADI) and to study the relationship between teachers' prior experience with the method and their attitude towards it. The theoretical basis for this study was developed from an examination of innovation theory, in which several characteristics of innovations thought to contribute to teacher attitude were identified.

The data used to test the relationships presented in this study were obtained from two samples. A 'change' sample was used to investigate change in attitude after observing a brief demonstration of Direct Instruction on a videotape. A 'control' sample was used to investigate the contribution to measurement of the Not Sure (NS) response category in the Likert type attitude scale as well as providing part of the calibration sample for assessing the psychometric characteristics of the ADI Scale.

Data analysis was conducted in two stages. The first stage involved the calibration of the measuring instrument, and as a result, the nature and meaning of the variable Attitude towards Direct Instruction was established. The second stage of the analysis assessed, through the use of data plots, the attitude change of participants after a brief exposure to Direct Instruction. This graphical procedure also enabled a study of the relationship between differing degrees of teacher experience with Direct Instruction and attitude change.

Summary of Findings

There are five main findings which come out of the analysis.

1. The first relates to the psychometric properties of the Attitude towards Direct Instruction Scale. The final scale of 19 items is constructed in the Likert format using response categories of Strongly Agree, Agree, Disagree, and Strongly Disagree and exhibits a strong 'sample-free' character. The threshold estimates for all items are correctly ordered; hence the model distribution for these parameters is strictly unimodal. The Index of Person Separation statistic of 0.95 indicates the instrument has high reliability and is capable of separating the persons along the latent continuum relative to the item location estimates. In addition, a value of this magnitude means that the power of the tests-of-fit is high. An overall test-of-fit statistic with a probability of 0.74 indicates a satisfactory degree of consensus exists between all items across persons located at differing attitude levels.

   Consequently, the variable, Attitude toward Direct Instruction has clear meaning as defined in terms of the item order along the latent trait continuum. Teachers who have a negative attitude towards Direct Instruction are more likely to be those who have never taught the method. They are likely to perceive Direct Instruction as an effective teaching strategy, but would consider it to be too teacher oriented and structured. On the other hand, teachers who have a positive attitude towards Direct Instruction are generally those who have had the opportunity to use it. They see Direct Instruction as effective, easy to use and able to be widely applied to a variety of subjects.

2. The use of the extended model of Rasch for assessing threshold order especially in relation to the 'Not Sure' or neutral type of category is an important tool for building meaningful variables whose measure is required

within the frame of reference as specified. The method of employing the NS category to improve the quality of measures in a repeated measures design where representatives at the pretest occasion have limited, or no, knowledge of the context of the variable was justified. Reverse thresholds were largely overcome by treating responses in the NS category as missing data and using responses from the other categories to calibrate the instrument.

It was apparent that the location of the NS category in the middle of the response continuum: Strongly Agree, Agree, Disagree, and Strongly Disagree was not appropriate. In future situations where the NS category is employed in an attitude scale, it should be located to one side of, and separate from, the response continuum, and the responses to that category treated as missing data.

3. The use of graphical procedures as a technique to investigate item bias and attitude change has several advantages. In general, the method is useful for detecting the presence of interaction effects across subsamples within a calibrating sample as well as providing ready access to simple yet revealing means of examining changes in learning or behaviour at both group and individual level. In this study, which involved a repeated measures design, the graphical procedure was used to assess item bias between two occasions, as well as attitude change by treating the sets of person responses on the two occasions as eminating from two different sets of persons. In both cases reported in this study, the plots provided an effective and simple interpretation of the information. Items or people that responded in an inconsistent manner between pretest and posttest occasions were easily identified thus allowing for meaningful diagnostic action to be undertaken.

4. Teacher attitude towards Direct Instruction is influenced by the amount of previous knowledge and experience with the method. This study found that teachers who had no experience with Direct Instruction were more likely to have a negative attitude towards it, and teachers who have taught Direct Instruction are more likely to hold a positive attitude towards it. This finding is consistent with the theory of implementation of innovations in that teachers are unwilling to express a strongly supportive attitude about an untried teaching method.

5. Direct Instruction, as a teaching strategy, does not create a favourable first impression. Teachers observing Direct Instruction for the first time are likely to react negatively to the apparent excessive structure and regimentation. This finding is consistent with comments made by various researchers. While conceding that Direct Instruction may be an effective teaching strategy, many teachers initially consider the effort involved in implementing the approach not to be worthwhile.

Implications

The implications of these findings, and of the study itself, are now discussed in relation to three areas: future research, theory, and practice.

For Future Research

To test the hypotheses posited by this study, an instrument was developed to measure teacher attitude towards Direct Instruction. This scale would be of value to practitioners, teacher educators, and researchers. For example, if Direct Instruction is considered for implementation by teachers, the information provided by the Scale would be of value as a starting point for training programmes. The scale, known as the Attitude to Direct Instruction Scale (ADI

Scale) has 19 items. The response categories range from Strongly Agree, Agree, Disagree, to Strongly Disagree. A Not Sure response category could be provided, but this must be located to one side of the normal Likert response sequence. Responses to the NS category would then be treated as missing data. The questionnaire takes approximately ten minutes to complete.

A number of implications for future research arise in relation to the measurement of Attitude towards Direct Instruction. First, the importance of threshold order for multiple category items was identified as critical in obtaining a set of items conforming to the Rasch measurement model. The capacity of the extended model to explore and account for the characteristics of data scored using multiple category items and its relationship to a logical threshold structure for the scoring function is an important advance for psychometric theory. This model is thus capable of providing person estimates of considerable precision and in a manner not possible with non-Rasch measurement models.

Secondly, this study has highlighted the influence of a NS category within the Likert response format and the effect that the presence of this category has on the measurement of attitude. In situations where the provision of a NS category is important to maximise validity, this study has provided a technique for handling the calibration of the measuring instrument. The strategy involves suppressing responses to the NS category and treating them as missing data. One advantage of using the estimation algorithm associated with the extended model is its ability to handle such data routinely.

Thirdly, the graphical procedures employed to investigate item bias and attitude change are simple to use and interpret, but powerful, and provide information at both group and individual level. The method provides an opportunity to investigate individual aberrant behaviour in a meaningful manner, and as a result

increase the opportunity for understanding the meaning of the variable to be measured.

Fourthly, a meaning of teacher attitude towards Direct Instruction as an innovation was defined in terms of three characteristics: cost, need, and clarity. Through use of the extended model of Rasch to calibrate the measuring instrument, it was possible to establish that items relating to all three characteristics lay on a single measurement continuum. As a consequence, this aspect of innovation theory was consolidated.

Finally, the analysis of the items in terms of the strict requirements of objective measurement has, once again, emphasised the importance of the wording of item statements. It was noted that items including emotive words or phrases did not fit the model, that is, they did not attract consistent responses from participants. The distribution of responses to items expressed in the future tense were also erratic. These trends suggest that the wording of items in attitude questionnaires should be without affective phrases and, if possible, in the present tense.

The Attitude towards Direct Instruction Scale could benefit from the inclusion of additional items drawn from two areas. First, items relating to the *general* characteristics of innovation of cost, clarity, and need may increase the precision of information regarding the relative importance of each one. By including items relating to specific aspects of a strategy, such as Direct Instruction, with a core of general items as stated, a versatile battery of measures of attitudes to innovation in its different manifestations could be constructed. This approach would also assist persons unfamiliar with a specific technique or strategy by providing them with an increased opportunity to respond in a meaningful way in the pretest occasion of an investigation.

A second area to consider is that the present scale has three significant gaps between items on the continuum. These occurred between items located at –1.0 and –1.6 logits, between –0.2 and 0.4 logits, and between 1.4 and 2.0 logits. Such gaps would increase the standard error and therefore, decrease the precision of the instrument because of some mismatch in targetting persons to items. The closing of these gaps through the creation of additional items located in these positions would enhance the usefulness of the Attitude to Direct Instruction Scale by accounting for these present gaps in our knowledge of the variable. To determine the substance of such items, the content of nearby items would need to be examined. Subsequent analyses employing these additional items would also establish further the construct validity of the instrument.

On a more general note, this study raised the issue of how to promote a new teaching strategy. The videotape used to demonstrate Direct Instruction was not intended to promote the method. In fact, it is suspected that the videotape as a presentation over and above the method being presented may have contributed disproportionally to the negative reaction by many participants, particularly those in Group One. Several Group Three participants commented that the tape was not a good demonstration , however technically accurate. By adapting and expanding the generality of the Attitude towards Direct Instruction Scale as described earlier, investigations of various types of promotional material could be undertaken and the degree of attitude change assessed along the lines presented in this paper.

For Theory

Implications arising from the study which related to theory covered several aspects. One important implication of the results of the present study is that attitude to an innovation, as measured by Attitude towards Direct Instruction

Scale, depends on degree of experience with the method. This indicates that teachers are cautious in forming an opinion about a new teaching method. They are more likely to hold a positive attitude towards a teaching strategy they have had the opportunity of teaching for a period of time.

There is no single, clearly articulated, generally agreed upon educational theory upon which innovations can be based (Common & Egan, 1988). At the moment, various schools of thought exist as to the best way to teach a child, how to teach different information, differing learning styles and so forth. Therefore most teachers use methods that are in accord with their own personal philosophies regarding children's learning styles. This study highlighted the personal nature of teachers' choice of teaching strategies.

Finally, this study shows the need to incorporate a measurement model with the development of variables. This has implications for theory in that the process of theory building is an integral part of the development of variables which are a meaningful part of explanatory theory itself. The two are complementary, not separate processes. The Rasch model is especially suitable for this process.

For Practice

Several issues relating to the successful implementation of Direct Instruction were highlighted by this study. The fact that attitude towards Direct Instruction, as an innovation, depends upon background knowledge, has important implications for practice. Generally, for a teaching innovation to be successfully implemented, the principal focus needs to be the teacher in the classroom. Many innovations have failed due to insufficient time and effort spent with the people who put them into practice, the teachers. Most research on implementation of successful educational programmes in the last few years has indicated that for programmes to succeed, an individual with instructional

expertise must serve as leader, manager, problem solver, facilitator, classroom coach (Crandall, Baucher, Loucks, & Schmidt, 1982; Hall & Hord, 1986; Showers, 1985; Stallings, 1980). In a study of high school teachers, Berlin and Jensen (1989) found few teachers capable of taking in a new idea independently, internalizing it and creating a new structure. Most teachers require feedback, encouragement and support if change is to result. With respect to Direct Instruction, successful implementation has occurred after extensive staff training sessions (Fields, 1986), or supervision by experts in the method (Gersten et al., 1986). One factor contributing to the success of Direct Instruction in Project Follow Through was thought to be the role of the facilitators (Gersten, Carnine, & Woodward, 1987). The necessity for the "development and refinement of sensitive, sensible inservice and professional development activities for teachers and instructional aides" has been acknowledged as a priority over altering curriculum materials in promoting Direct Instruction. (Gersten et al., 1987, p.49) Clearly, the teacher training component has been underestimated. In this study Direct Instruction was not widely known amongst teachers in local primary schools. Of those participants who were familiar with the method, most had seen it used only in a remedial situation. Longer term motivation sessions and inservice provisions, and the offer of support to teachers in the early stages of the implementation of Direct Instruction, should be considered.

Teachers are introduced to educational innovations through a variety of ways: professional literature, workshops, conferences, visitations, observing peers. The way in which the teacher first has contact with the innovation appears to be critical. Many innovations cannot be fully appreciated until they have been used for a while. Worthwhile teaching strategies have been ignored or shunned by teachers after a brief demonstration. This study showed that a short exposure to Direct Instruction was sufficient to cause a significant change in attitude. Doyle

and Ponder (1977) believe teachers tend to make judgements rapidly with minimal experience or evidence based principally on the characteristics of need, clarity and cost. Direct Instruction has many characteristics that actively discourage teacher enthusiasm, for example, its highly structured, repetitive nature. These aspects are very obvious in a brief observation of the method. Therefore, the failure of Direct Instruction to make a favourable first impression needs to be considered by practitioners considering its implementation.

The extensive body of research attesting to the effectiveness of Direct Instruction did not emerge as a factor contributing to teachers' attitude. Gersten et al. (1987) admit that they naively assumed the powerful body of empirical evidence would be sufficient to ensure the widespread use of Direct Instruction. However, a large number of participants in the study had never heard of Direct Instruction. Hollifield (1982) and Marchant (1989) observed that many teachers do not to keep abreast of current educational literature. Ornstein (1989, p.95) suggests that many teachers "have little reason for reading the research, lack research knowledge and are unable to understand the data, or feel that the research is not relevant to the practice of teaching." Richardson (1990) felt that while many innovations are based on sound theoretical frameworks, teachers will embed the new activity into their own set of premises which may be unrelated. It is the degree of congruence between their personal needs regarding classroom management and content coverage, and the innovation itself, that has the most affect on a teacher's attitude. Brown and McIntyre (1982) suggest more enquiry-oriented teacher education programmes may encourage teachers to increase their critical consciousness.

REFERENCES

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour.* Englewood Cliffs, N.J.; Prentice- Hall.

Andrich, D. (1975). *The Rasch Multiplicative Binomial Model: Applications to attitude data.* Research Report Number I. Measurement and Statistics Laboratory, Department of Education, U.W.A.

Andrich, D. (1978a). A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology, 31,* 84-98.

Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594.

Andrich, D. (1978c). A rating formulation for ordered response categories. *Psychometrika, 42,* 561-573.

Andrich,D. (1978d). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement, 38,* 665-680.

Andrich, D. (1982). Using latent trait measurement to analyse attitudinal data: A synthesis of viewpoints. In D. Spearitt (Ed.), *The improvement of measurement in Education and Psychology* (pp. 89-126). Melb. Vic.: ACER Ltd.

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological methodology* (pp. 33-80). San Francisco: Jossey-Bass.

Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education, 1*(4), 363-378.

Andrich, D., Lyne, A., & Sheridan, B. (1991). *ASCORE: A FORTRAN IV program for the general analysis of dichotomous and graded response data to achievement and attitude instruments.* Perth: School of Education, Murdoch University.

Andrich, D., & Schoubroeck, W. (1989). The general health questionnaire: A psychometric analysis using latent trait theory. *Psychological Medicine, 19,* 469-485.

Andrich, D., & Sheridan, B. (1980). *RATE: A Fortran IV program for analysing rated data according to a Rasch model* (Research Rep. No. 5). Perth: University of Western Australia, Department of Education, Measurement and Statistics Laboratory.

Archer, A., & Gleason, M. (1989). *Skills for school success.* Mass: Curriculum Associates

Becker, W.C. (1977). Teaching reading and language to the disadvantaged: What we have learned from Field Research. *Harvard Educational Review, 47*(4), 518–543.

Becker, W.C. (1978). The National Evaluation of Follow Through. *Education and Urban Society, 10*(4), 431–458.

Becker, W.C. (1984). *Direct Instruction – A 20 year Review.* Paper presented at the XVI Annual Banff International Conference on Behaviour Sciences, Banff, Canada.

Berlin,B., & Jensen, K. (1989). Changing teachers. *Education and Urban Society, 22*(1), 115–120.

Berman,P., & McLaughlan,M. (1976). *Federal programs supporting educational change: Volume One.* Santa Monica, California: Rand Corp.

Black, J., & Champion, D. (1976). *Methods and issues in social research.* New York: John Wiley & Sons.

Branwhite, A. (1982). The singer or the song? Does the programme make a difference with a good teacher? *Direct Instruction News, 1*(2),6–7.

Breckler, S.J., & Wiggins, E.J. (1989). On defining attitude and attitude theory: Once more with feeling. In A.R.Pratkanis, S.J. Breckler, & A.G. Greenwald (Eds.). *Attitude Structure and Function.* Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Brown, S., & McIntyre,D. (1982). Influence upon teachers' attitudes to different types of innovation. *Curriculum Inquiry, 12* (1), 35–51.

Carnine, D. (1977). Direct Instruction – DISTAR. In N. G. Haring & B. Bateman (Eds.), *Teaching the learning disabled child.* Englewood Cliffs, N.J.: Prentice Hall.

Cavanagh, G., & Styles,K. (1983). The implementation of new curriculum guidelines and policies. *Education Canada. 23*(1), 9–15.

Clark, D., & Guba, E. (1965). The re-examination of a test of the research and development model of change. *Educational Research Quarterly, 8*(3), 92–103.

Clunies–Ross, G. (1979). Accelerating the development of Downs Syndrome infants and young children. *The Journal Of Special Education, 13* (2),

Cohen, M., & Nagel, E. (1934). *An introduction to logic and scientific method.* New York: Harcourt Brace.

Cole, P.G., & Chan, L.K.S. (1990). *Methods and strategies for special education.* New York: Prentice– Hall.

Common, D., & Egan,K. (1988). The missing soul of models of curriculum implentation – educational theory. *Curriculum Perspectives, 8*(1), 1–10.

Connolly, R.G., Nachtman, W., & Pritchett, E.M. (1976). *Key Math.* American Guidance Service.

Corr, M. (1984). Direct Instruction – A viable educational option. *Australian Journal of Remedial Education, 16*(2), 1–4.

Cotton,K., & Savard, W.G. (1982). *Direct Instruction: Topic Summary Report. Research on School Effectiveness Project.* Northwest Regional Educational Laboratory, Juneau, Alaska. (ERIC document No. ED214909).

Crandall, D.P. (1983). The teacher's role in school improvement. *Educational Leadership, 41*(3), 6–9.

Crandall,D., Baucher,J., Loucks, S., & Schmidt,W. (1982). *Models of the school improvement process: Factors contributing to success.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Crandall,D.P., Eiseman, J.W., & Louis,K.S. (1986). Strategic planning issues that bear on the success of school improvement efforts. *Educational Administratative Quarterly, 22*(3), 21–53.

Cronbach, L. (1946). Response sets and test validity. *Educational Psychological Measurement, 6,* 475–494.

Crozier, M. (1974). *An attempt to accelerate low–achieving elementary students using the DISTAR Instructional System.* Report to the Edmonton Public School System.

Crowther, F. (1972). *Factors affecting the rate of adoption of the 1971 Alba Social Studies curriculum for Elementary schools.* Unpublished master's thesis, University of Alberta, Canada.

DeFleur, M., & Westie,F. (1958). Verbal attitudes and the salience of attitudes. *American Sociological Review, 23*, 66–63.

Dixon,R., & Engelmann, S. (1981, 1990). *Spelling Mastery.* Chicago: SRA.

Doyle,W., & Ponder, G.A. (1977). The practicality ethic in teacher decision–making. *Interchange, 8*(3), 1–12.

DuBois, B., & Burns, J.A. (1972). An analysis of the meaning of the question mark response category in attitude scales. *Educational and Psychological Measurement, 35*, 869–884.

Emrik, J., & Paterson, S. (1978). *A synthesis of findings across five recent studies in educational dissemination and change.* San Francisco: Far West Laboratory for Educational Research and Development.

Engelmann,S., Becker,W., Carnine,D., & Gersten, R. (1988). The Direct Instruction Follow Through Model: Design and outcomes. *Education and Treatment of Children, 11*(4), 303–317.

Engelmann,S., & Carnine,D. (1981). *Corrective mathematics.* Chicago: SRA

Engelmann,S., & Carnine, D. (1982). *Theory of Instruction; Principles and Applications.* New York: Irvington Publishers.

Engelmann,S., Hanner,S., & Johnson,G. (1978, 1989). *Corrective Reading Programme.* Chicago: SRA.

Engelmann,S., & Silbert, J. (1985). *Expressive writing.* Chicago: SRA

Fields, B.A. (1986). Direct Instruction and the practicality ethic: A case study and position in curriculum innovation. *Australian Journal of Remedial Education, 18*(1), 17–22.

Fischer,G.H. (1976). Some probabilistic models for measuring change. In Cronbag, DeGruijter, & Van der Kamp (Eds.) *Advances in psychological and educational measurement* (pp.97–110). NY: J. Wiley.

Fishbein, M., & Ajzen,I. (1972). Attitudes and opinions. *Annual Review of Psychology, 23*, 487–544.

Fishbein,M., & Ajzen, I. (1975). *Belief attitude, intention and behaviour: An introduction to theory and research.* Reading, Mass: Addison–Wesley.

Fullan, M. (1982). *The meaning of educational change.* Columbia University, N.Y.: Teachers College Press.

Fullan, M. (1991). *The new meaning of educational change.* London: Cassell Educational Ltd.

Fullan, M., & Pomfret,A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research, 47*(2), 335–397.

Gay, L.R. (1987). *Educational research: Competencies for analysis and application* (3rd ed.). Columbus, OH.: Merrill Pub. Coy.

Gaynor, A., & Du Vall, L. (1977). *The dynamics of educational innovation: Theory and practice.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Gersten, R., Brockway, M., & Henares, N. (1983). The Monterey DI Program for students with limited English (ESL). *Association for DI News, 2*(4), 8–9.

Gersten, R., Carnine, D., & Williams, P. (1982). The principal as instructional leader: A second look. *Educational Leadership, 40*(3), 47–50.

Gersten, R., Carnine, D., & Woodward, J. (1987). Direct Instruction research: The third decade. *Remedial and Special Education, 8*(6), 48–56.

Gersten, R., Carnine, D., Zoref, L., & Cronin, D. (1986). A multifaceted study of change in seven inner-city schools. *The Elementary School Journal, 86*(3), 257–276.

Gersten, R., & Guskey, T. (1985). *Transforming teacher reluctance to teacher commitment.* Paper presented at the Annual Meeting of the American Educational Research Association. (ERIC document No. ED260068).

Gersten, R., Keating,T., & Becker,W. (1988). The continued impact of the Direct Instruction Model: Longitudinal studies of Follow Through students. *Education and Treatment of Children, 11*(4), 318–327.

Greenwald, A.G. (1989). Why are attitudes important? In A.R. Pratkanis, S.J. Breckler, & A.G. Greenwald. (Eds.). *Attitude Structure and Function.* Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Gregory, R.P. (1983). Direct Instruction, disadvantaged and handicapped children: A review of the literature and some practical implications. Parts 1 and 11. *Remedial Education, 18*(3), 108–114; 130–136.

Guttman,L. (1950). The problem of attitude and opinion measurement. In S.A. Stouffer and others (Eds.), *Measurement and Prediction.* New York: Wiley.

Guttman, L. (1954). The principal components of scalable attitudes. In P.F. Lazarsfeld (Ed.), *Mathematical thinking in the Social Sciences.* New York: Free Press.

Hall, G., & Hord, S. (1986). *Configurations of school-based leadership teams.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Hall, G.E., Loucks, S.F., Rutherford, W.L., & Newlove, B.W. (1975). Levels of Use of the innovation: A framework for analysing innovation adoption. *Journal of Teacher Education, 26*(1), 52–56.

Hawke, H., Maggs, A., & Moore, J. (1979). 'Disadvantaged' turned into 'advantaged': A three year study. *Australian Educational Researcher, 6*(3), 12–21.

Hawke, H., Maggs,A., & Murdoch, R. (1978). The effects of teaching morphographics on learning to spell. *Special Education, 13* (3), 59–63.

Henry, M.I. (1989). The implementation of three innovations in Home Economics in Queensland: A case study. *Curriculum Perspectives, 9*(4), 58–63.

Hollifield, J. (1982). Teachers tell all. *Today's Education, 71*(1), 54–56.

Hollingsworth, P., & Reutzel,R. (1990). Prior knowledge, content related attitude, reading comprehension: Testing Methewson's affective model of reading. *Journal of Educational Research, 83*(4), 194–199.

Hoyle, E. (1976). The parameters of change. *Innovation: Problems and Possibilities* (pp. 27–29; 36–54). London: Open University

Huberman, A.M., & Miles, M.B. (1984). *Innovation up close.* New York: Plenum Press.

Likert, R. (1932). A technique for measuring attitudes. *Archives of Psychology, 140*, 1–55.

Lippitt, R., Watson, J., & Westley, B. (1958). *The dynamics of planned change.* New York: Harcourt, Bruce and World

Lipson, M. (1983). The influence of religious affiliation on children's memory for text information. *Reading Research Quarterly, 18*(4), 418–45.

Lockery, M., & Maggs, A. (1982). Direct Instruction research in Australia: A ten year analysis. *Educational Psychology, 2*(3–4), 263–287.

Lortie, D. (1975). *School teacher: A sociological study.* Chicago: University of Chicago Press.

Loucks, S., & Hall, G. (1979). *Implementing Innovations In Schools: A concerns based approach.* Paper presented at the Annual Meeting of the American Educational Research Association.

Maggs, A., & Moore, J. (1978). Teaching academic skills by direct instructional methods to school-aged Aboriginal and non-Aboriginal children with learning problems. *The Exceptional Child, 25* (2), 12-136.

Marchant, G.J. (1989). How many and why not more: A survey of issues in education (ERIC document No. ED 314 384).

Mathes, P.G., & Proctor, T.J. (1988). Direct Instruction for teaching hard to teach students. *Reading Improvement, 25*(2), 92-97.

McGuire, W.J. (1985). In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology : Vol. 11.* (3rd ed.). (pp. 233-346). New York: Random House.

Merton, R. (1968). *Social theory and social structure.* New York: Collier-MacMillan

Meyer, L. (1984). Longterm academic effects of the DI project Follow Through. *Elementary School Journal,* 87 (4), 380-394.

Mohlman, G.G., Coladarci, T., & Gage, N.L. (1982). Comprehension and attitude as predictors of implementation of teacher training. *Journal of Teacher Education, 33*(1), 31-36.

Moore, J. (1981). Direct Instruction: Moving towards faultless instruction in basic and advanced literacy and numeracy. *Reading Education, 6*(1), 34-41.

Mort, P. (1964). Studies in education innovation from the Institute of Administrative Research. In M. Miles (Ed.) *Innovation in Education.* New York: Bureau of Publications, Teachers College, Columbia University.

Mossenson, L., Hill, P. & Masters, G. (1987). *TORCH - Test of reading comprehension.* Melbourne: ACER.

Nicholls, A. (1983). *Managing educational innovations.* London: Allen & Unwin.

Nicholls, C. (1980). Kentucky Fried Reading Scheme. *Radical Education Document,* U.S.A., 18-22.

Ornstein, A.C. (1989). Theoretical issues related to teaching. *Education and Urban Society, 22*(1), 95-104.

Ostrom, T.M. (1989). Interdependence of attitude theory and measurement. In A.R. Pratkanis, S.J. Breckler, & A.G. Greenwald (Eds.). *Attitude Structure and Function.* Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Paul, D.A. (1977). Change processes at the Elementary, Secondary and Post Secondary levels of education. In N.Nash & J. Culbertson (Eds), *Linking processes in educational improvement: Concepts and applications.* Columbus, Ohio: The University Council for Educational Administration, 7–73.

Print, M. (1987). *Curriculum development and design.* Sydney: Allen & Unwin

Proctor, T. (1989). Attitudes toward Direct Instruction. *Direct Instruction News, 9*(1), 11–15.

Rasch, G. (1960/80). *Probabilistic models for some intelligence and attainment tests.* (Expanded ed.). Chicago: University of Chicago Press.

Rennie, L., & Parker, L. (1984). Problems in the interpretation of attitude data. In *Research and Educational Futures – Technology, Development and Educational Futures.* Collected papers from the 1984 Conference of the Australian Association for Research in Education. (pp.718–725). Perth, Western Australia.

Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational Researcher, 19*(7), 10–18.

Rogers, E., & Shoemaker, F. (1971)., *Communication of Innovations.* New York: Free Press.

Rosenblum, S. & Louis, K. (1981). *Stability and change.* New York: Plenum.

Rost, J. (1988). Measuring attitudes with a threshold model drawing on a traditional scaling concept. *Applied Psychological Measurement, 12*(4), 397–409.

Shaw, M.E., & Wright, J.M. (1967). *Scales for the measurement of attitudes.* New York: McGraw-Hill Book Coy.

Sheridan, B. (1993). *Threshold order and Likert-style questionnaires.* Paper presented at the Seventh International Objective Measurement Workshop, Annual Meeting of the American Educational Research Association, Atlanta, April 10–11.

Showers, B. (1985). Teachers coaching teachers. *Educational leadership, 42*(7), 43–48.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher, 9*, 11–16.

Stebbins, L., St.Pierre, R., Proper, E., Anderson, R., & Cerva,T. (1977). *Education as experimentation: A planned variation model (Vol.1VA-D).* Cambridge, MA: Abt Associates.

Stephens, T. (1974). Innovative teaching practices: Their relation to system norms and rewards. *Educational Administration Quarterly, Winter,* 10,17-25.

Stern, C. & Keisler, E.R. (1977). Teacher attitudes and attitude change: A research review. *Journal of Research and Development in Education, 10,* 63-76.

Stevens, S. (1951). Mathematics, measurement, and psychophysics. In S. Stevens (ed). *Handbook of Experimental Psychology.* (pp 1-49). New York: Wiley.

Tarter, D. (1969). Toward prediction of attitude-action discrepancy. *Social Forces, 4,* 398-405.

Thompson, B. (1979). A theory of teacher change developed from teachers of writing. (ERIC document No. 188197).

Thorndike, E. (1904/1916). *An introduction to the theory of mental and social measurements.* New York: Teachers College, Columbia University.

Thurstone, L.L., & Chave, E.J. (1929). *The measurement of attitude.* Chicago: University of Chicago Press.

Tobin, K. (1987). Forces which shape the implementation curriculum in high school science and mathematics. *Teaching and Teacher Education, 3,* 287-298.

Waldrop, P.B., & Adams, T.M. (1988). Overcoming resistance to the use of instructional computing in Higher Education (ERIC document No. 296656).

Waugh, R. (1983). *Teacher receptivity to system wide change,* Unpublished Ph.D. thesis, University of Western Australia.

Weisberg, P. (1983). Education and enrichment approaches. In C.E. Walker & M.C. Roberts (Eds), *Handbook of Clinical Psychology* (pp 1123-1143). New York: John Wiley.

Weisberg, P. (1984). Reading instruction for poverty-level preschoolers. *Association for Direct Instruction News, 3*(2), 16-19.

Weisberg, P., Packer, R.A., & Weisberg, R.S. (1981). Academic training. In J.L. Matson & J.R. McCartney (Eds.), *Handbook of behaviour modification with the mentally retarded.* New York: Plenum.

White, W. (1988). A meta-analysis of the effects of direct instruction in special education. *Education and Treatment of Children, 11*(4), 364–374.

Wicker, A.W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioural responses to attitude objects. *Journal of Social Issues, 25*(4), 41–78.

Woodcock, R.W. (1987). *Woodcock Reading Mastery Tests–Revised.* Circle Pines, MN: American Guidance Service

Wright, B.D. (1967). *Sample-free test calibration and person measurement.* Paper presented at the Invitational Conference on Testing Problems.

Wright, B.D. (1988). Georg Rasch and measurement. *Rasch Measurement, 2*(3), 22–28.

Wright, B.D. (1991). Scores, reliabilities and assumptions. *Rasch Measurement, 5*(3), 157–158.

Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis.* Chicago: Mesa Press.

Wright, B.D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA Press.

Wright, B.D. & Stone, M. H. (1991). Control lines for item plots. *Measurement Primer No. 13.* Chicago: MESA Press.

APPENDIX A

The Measuring Instrument

Attitude to Direct Instruction

All items for the Attitude to Direct Instruction Survey are listed on the following pages.

The two columns to the left of the item numbers contain the following information:

Column 1 Indicates the characteristic of an innovation examined by that statement in accordance with the conceptual framework established in Chapter Two and coded as:

N : Need
C : Cost
L : Clarity
U : Other

Column 2 Indicates the mode of the statement wording and coded as::

P : Positive wording
R : Negative wording (or reversed scored)

The responses for each item statement were obtained using the following set of categories according to the sample as specified:

Change Sample – for the Pretest and Posttest Occasions

Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree

Control Sample – with the NS category included at margin

Strongly Disagree	Disagree	Agree	Strongly Agree	Not Sure

– with the NS category excluded completely

Strongly Disagree	Disagree	Agree	Strongly Agree

ATTITUDE TO DIRECT INSTRUCTION SURVEY

Innovation Code	Wording Mode		Item Statement
C	P	1.	Student improvement is worth the extra effort of using Direct Instruction.
C	R	2.	Direct Instruction restricts a teacher's initiative.
C	P	3.	Direct Instruction improves my effectiveness as a teacher.
C	R	4.	Direct Instruction is primarily beneficial in 1:1 and small groups.
U	R	5.	The benefits of Direct Instruction have been exaggerated.
N	P	6.	A student's motivation to learn is enhanced with Direct Instruction.
C	P	7.	I feel well prepared to use Direct Instruction.
N	R	8.	Direct Instruction places too much emphasis on academics.
U	R	9.	There is too much emphasis on Direct Instruction in teacher training.
N	R	10.	Direct Instruction cannot be used with all children.
C	P	11.	Direct Instruction is an easy teaching approach to master.
C	R	12.	Direct Instruction is too teacher centred.
C	R	13.	Direct Instruction is boring.
L	P	14.	To fully appreciate Direct Instr. the underlying theory needs to be understood.
L	P	15.	Regular use of Direct Instr. with students has increased my appreciation of it
N	P	16.	Direct Instruction can develop a child's self esteem.
U	R	17.	I do not intend to use Direct Instruction when I teach.
N	P	18.	Student attention is maintained with the Direct Instruction method.
U	P	19.	There would be more support for Direct Instruction if people knew more about it.
C	R	20.	Direct Instruction is too mechanistic.

Innovation Code	Wording Mode		Item Statement
N	P	21.	Direct Instruction helps students cope with academics.
C	R	22.	Direct Instruction is unable to meet the complex demands of the classroom.
N	R	23.	Direct Instruction makes students too dependent on the teacher.
U	P	24.	I am looking forward to using Direct Instruction.
N	P	25.	Direct Instruction improves overall classroom conditions.
N	R	26.	Direct Instruction fosters rote learning.
N	P	27.	Direct Instruction has unlimited possibilities.
N	P	28.	Direct Instruction meets my students' academic needs.
N	R	29.	Direct Instruction is too fast paced for small children.
C	R	30.	All teachers should be prohibited from using Direct Instruction.
N	P	31.	Direct Instruction aids learning.
C	R	32.	Direct Instruction is too difficult to implement in the classroom.
L	P	33.	I will use Direct Instruction techniques even if materials are not available.
N	P	34.	Direct Instruction creates a positive attitude in the classroom.
U	P	35.	I am glad Direct Instruction is emphasized in teacher training.
N	P	36.	More money should be spent on Direct Instruction programmes.
L	P	37.	Direct Instruction can be used in all subject areas.
C	R	38.	Direct Instruction is excessively slow paced.
C	R	39.	I hope I do not have to use Direct Instruction when teaching.
C	R	40.	Direct Instruction does not carry over into the regular classroom.

Checklist of Observable Features of Direct Instruction

A Direct Instruction lesson should contain the following observable features. These have been noted by Mathes and Proctor (1988) and Gersten and Carnine (1982). Please circle the appropriate response for each feature.

1. Scripted presentation of the lesson, based on the instructional design principles developed by Engelmann and Carnine (1982). Yes/No

Comment: _____

_____

2. Teacher–directed small group instruction Yes/No

Comment: _____

_____

3 A signalling system for prompting unison responses to questioning of students. Yes/No

Comment: _____

_____

4. Positive reinforcement Yes/No

Comment: _____

_____

5. Fast pacing of lessons Yes/No

Comment: _____

_____

6 Immediate corrective feedback Yes/No

Comment: _____

_____

Any further comments please: _____

_____

_____

APPENDIX B

Tables

This appendix contains all tables referred to in the text.

TABLE B-1

Threshold estimates for existing data when the NS category is included as part of the scoring function in the sequence Strongly Disagree, Disagree, NS, Agree, Strongly Agree.

Item		Pretest (N=144)					Posttest (N=144)			
		1	2	3	4		1	2	3	4
1	*	-1.16	-1.84	0.72	2.28	*	-2.10	0.03	-0.31	2.39
2		-2.07	-0.49	-0.22	2.78	*	-1.73	1.61	-1.66	1.78
3		-2.00	-0.89	0.64	2.25	*	-2.29	0.12	-0.17	2.34
4	*	-1.55	0.04	-0.23	1.74	*	-1.91	1.34	0.10	0.47
5		-3.51	-1.27	1.46	3.32		-3.08	-1.10	0.76	3.41
6		-2.52	-0.61	0.37	2.77		-1.79	-0.31	-0.25	2.35
7	*	-1.04	-0.34	-0.66	2.05	*	-1.51	0.24	-1.18	2.45
8		-3.47	-0.47	0.65	3.28		-2.15	-0.15	0.21	2.08
9	*	-1.27	-2.13	1.12	2.28	*	-1.47	-2.15	1.38	2.24
10		-1.53	-0.32	-0.21	2.06	*	-1.36	0.14	-1.24	2.46
11		-1.47	-1.12	-0.15	2.75		-2.15	-0.35	-0.11	2.61
12		-1.87	-0.49	-0.43	2.78	*	-2.34	0.99	-1.63	2.98
13	*	-2.57	-0.29	-0.32	3.18	*	-1.96	0.51	-0.87	2.32
14	*	-1.33	0.00	-0.70	2.04	*	-0.89	-0.95	-1.05	2.89
15		-1.91	-1.81	0.67	3.05		-2.35	-1.38	1.41	2.33
16		-1.62	-1.01	0.20	2.43	*	-1.25	-1.26	-0.13	2.64
17		-2.48	-0.78	0.24	3.01		-1.78	-0.60	-0.08	2.46
18		-2.03	-0.97	0.25	2.75	*	-1.87	0.27	-0.94	2.54
19		-2.13	-0.86	0.29	2.69		-2.68	-0.68	0.71	2.66
20		-3.22	-0.84	-0.07	4.13	*	-2.23	0.68	-1.12	2.66
21		-2.05	-1.08	0.23	2.90		-2.51	-1.26	0.22	3.54
22		-2.68	-0.78	0.64	2.82	*	-2.53	-0.25	-0.45	3.22
23		-2.79	-0.90	0.11	3.59	*	-2.69	0.46	-0.64	2.87
24		-2.11	-1.67	0.20	3.58		-2.18	-0.60	0.04	2.74
25		-2.66	-1.92	0.54	4.05		-2.88	-0.85	0.40	3.33
26		-2.68	-0.23	-0.12	3.04	*	-2.43	1.00	-1.16	2.60
27		-2.56	-1.37	0.57	3.36		-2.54	-0.71	0.80	2.45
28		-2.33	-1.71	0.51	3.53		-2.46	-1.00	0.79	2.67
29		-3.03	-1.51	1.19	3.35		-2.11	-0.83	0.13	2.80
30	*	-0.72	-2.25	0.67	2.29	*	-0.73	-1.88	-0.01	2.63
31		-1.70	-1.65	-0.05	3.40		-1.99	-1.96	0.43	3.52
32		-1.90	-1.06	0.17	2.78		-1.73	-0.88	-0.43	3.04
33		-1.89	-1.26	0.31	2.84	*	-2.50	-0.06	-0.10	2.66
34		-2.54	-1.87	0.80	3.61		-2.24	-0.96	0.69	2.51
35	*	-1.67	-2.06	0.81	2.91		-3.05	-1.04	1.12	2.97
36		-2.13	-1.42	1.03	2.53		-2.37	-0.92	0.48	2.81
37		-2.51	-0.87	0.47	2.90		-2.05	-0.65	-0.58	3.29
38		-2.89	-0.95	0.82	3.01	*	-2.40	0.06	-0.45	2.80
39		-1.93	-1.17	0.06	3.05		-1.49	-0.73	0.02	2.21
40		-2.24	-1.34	0.46	3.12		-2.18	-0.88	-0.22	3.28

* indicates disordered thresholds

TABLE B-2

Threshold estimates for existing data when the NS category is converted to missing data for the pretest and posttest occasions

Item		Pretest (N=119)				Posttest (=141)		
		1	2	3		1	2	3
1	*	-1.17	-1.40	2.57		-2.17	-0.37	2.54
2		-2.24	-0.65	2.88		-1.96	0.13	1.84
3		-1.92	-0.32	2.24		-2.49	0.07	2.42
4		-1.69	0.37	1.32	*	-2.21	1.57	0.64
5		-3.65	0.06	3.60		-3.36	-0.23	3.59
6		-2.28	-0.22	2.50		-2.16	-0.36	2.52
7		-1.17	-0.80	1.97		-1.84	-0.68	2.52
8		-3.42	0.17	3.25		-2.11	-0.09	2.20
9		-2.06	-0.95	3.01		-1.31	-1.18	2.49
10		-1.50	-0.32	1.82		-1.78	-0.63	2.42
11		-1.32	-1.24	2.56		-2.28	-0.63	2.91
12		-1.84	-0.54	2.38		-2.82	-0.32	3.14
13		-2.45	-0.29	2.74		-2.40	-0.10	2.50
14		-1.86	-0.73	2.59	*	-1.08	-1.92	3.00
15		-2.33	-0.86	3.19		-2.63	0.18	2.46
16		-1.31	-1.10	2.41		-1.55	-1.16	2.70
17		-2.46	-0.50	2.96		-2.11	-0.57	2.68
18		-2.20	-0.75	2.95		-2.04	-0.89	2.92
19		-2.30	-0.49	2.79		-3.34	0.12	3.22
20		-3.22	-0.69	3.91		-2.73	-0.10	2.83
21		-1.97	-0.78	2.75		-2.93	-1.02	3.95
22		-2.61	0.02	2.60		-2.99	-0.61	3.59
23		-2.97	-0.46	3.43		-3.20	0.21	2.99
24		-2.22	-1.02	3.25		-2.51	-0.26	2.77
25		-2.99	-0.95	3.95		-3.33	-0.14	3.47
26		-2.69	-0.09	2.78		-2.87	0.16	2.71
27		-2.59	-0.40	2.99		-2.78	0.36	2.42
28		-2.68	-0.64	3.32		-2.89	0.15	2.74
29		-3.03	-0.63	3.66		-2.41	-0.65	3.07
30	*	-0.32	-2.87	3.19	*	-0.63	-2.39	3.02
31		-2.06	-1.87	3.92		-2.52	-1.52	4.04
32		-2.11	-0.75	2.86		-1.90	-1.36	3.26
33		-1.78	-0.63	2.41		-2.76	0.05	2.72
34		-2.64	-0.92	3.55		-2.69	-0.06	2.75
35		-1.61	-1.17	2.78		-3.27	0.19	3.09
36		-2.09	-0.21	2.31		-2.63	-0.15	2.79
37		-2.23	-0.29	2.52		-2.28	-0.91	3.18
38		-3.04	-0.16	3.20		-2.39	-0.52	2.91
39		-1.63	-1.13	2.76		-1.73	-0.52	2.25
40		-2.25	-0.73	2.97		-2.43	-0.95	3.38

* indicates disordered thresholds

TABLE B-3

Threshold estimates for the control sample when the NS category is omitted or the NS responses collected are converted to missing data

Item		NS data suppressed (N=167)				NS not included (N=108)		
		1	2	3		1	2	3
1		-1.51	-0.47	1.98		-4.29	0.26	4.03
2		-1.44	0.26	1.18		-2.94	0.19	2.76
3		-2.09	-0.05	2.14		-3.70	-0.18	3.88
4		-1.48	0.33	1.14		-3.50	0.39	3.12
5		-2.52	0.40	2.12		-4.15	-0.36	4.50
6		-1.43	0.39	1.04		-3.37	-0.56	3.94
7		-1.50	-0.31	1.80		-2.47	-0.19	2.67
8		-2.01	0.21	1.80		-4.53	-1.37	5.90
9		-1.15	-0.80	1.96	*	-2.22	-2.74	4.96
10		-1.58	-0.30	1.88		-1.90	-0.40	2.30
11		-0.99	-0.65	1.64		-3.34	-1.20	4.54
12		-1.55	-0.13	1.68		-3.24	-1.42	4.66
13		-1.05	0.09	0.96		-2.77	-0.98	3.75
14	*	-0.69	-1.47	2.16		-1.65	-1.60	3.25
15		-1.71	0.35	1.36		-3.81	-0.25	4.06
16		-0.73	-0.54	1.28		-2.29	-0.71	3.00
17		-1.02	-0.84	1.85		-2.18	-0.57	2.75
18		-1.13	-0.63	1.76		-2.36	-0.45	2.81
19		-1.12	-0.90	2.03		-3.45	-0.21	3.66
20		-1.49	0.14	1.35		-3.54	-0.85	4.40
21		-2.16	-0.28	2.44		-3.61	-0.72	4.33
22		-1.72	-0.18	1.90		-3.20	-1.02	4.22
23		-2.26	0.64	1.62		-3.66	-0.37	4.02
24		-1.52	0.18	1.33		-2.76	0.41	2.34
25		-1.96	-0.03	1.99		-4.84	0.29	4.56
26	*	-1.73	1.14	0.59		-3.29	-0.95	4.24
27		-1.75	-0.10	1.86		-3.97	-0.01	3.97
28		-1.99	0.24	1.75		-4.03	-0.96	4.99
29		-1.63	-0.17	1.80		-2.82	-1.15	3.97
30	*	-0.57	-1.18	1.74		-1.73	-1.57	3.30
31		-1.39	-1.26	2.65		-2.24	-2.00	4.24
32		-1.57	-0.61	2.18		-3.47	-2.59	6.06
33		-1.86	0.07	1.79		-3.92	-0.45	4.37
34		-1.84	-0.23	2.07		-4.41	-0.43	4.83
35		-2.02	-0.68	2.70		-3.84	-0.70	4.53
36		-1.30	-0.26	1.56		-4.18	-0.22	4.40
37		-1.11	-0.64	1.75		-2.42	-1.16	3.58
38		-1.30	-0.54	1.84		-1.74	-1.14	2.87
39	*	-0.60	-1.08	1.68		-2.36	-0.36	2.72
40		-1.65	-0.43	2.07		-3.31	-0.42	3.73

* indicates disordered thresholds

TABLE B–4

Threshold Estimates and Item Fit for Attitude to Direct Instruction under the hypothesis of no misfit for 40 items (N = 409)

Item	Item–trait Interaction**		Standardised Residual	Threshold Estimates		
Code*	χ^2 (5 df)	Pr(χ^2)	(394 df)	1	2	3
CP01	10.25	0.04	–1.01	–3.25	–0.18	3.43
CR02	10.91	0.02	0.74	–3.00	–0.17	3.17
CP03	4.82	0.42	–1.34	–3.17	–0.06	3.23
CR04	78.92	0.00	4.14	–3.29	0.39	2.91
UR05	8.61	0.10	–1.99	–4.05	–0.13	4.18
NP06	3.32	0.64	–1.33	–2.63	–0.28	2.91
CP07	9.44	0.07	1.92	–2.41	–0.40	2.82
NR08	8.86	0.09	0.41	–3.75	–0.65	4.40
UR09	99.11	0.00	3.44	–1.95	–1.33	3.28
NR10	7.87	0.14	0.98	–2.23	–0.50	2.72
CP11	31.68	0.00	3.32	–2.84	–0.81	3.65
CR12	1.48	0.91	–1.76	–3.09	–0.79	3.88
CR13	12.65	0.00	–3.41	–2.62	–0.50	3.12
LP14	61.47	0.00	2.54	–1.35	–2.06	3.41
LP15	11.13	0.02	–2.42	–3.11	–0.07	3.18
NP16	6.61	0.23	–0.59	–1.77	–0.88	2.65
UR17	10.83	0.03	–2.40	–2.08	–0.77	2.85
NP18	19.32	0.00	1.81	–2.27	–0.66	2.93
UP19	3.10	0.68	0.50	–2.97	–0.42	3.39
CR20	13.08	0.00	–3.21	–3.22	–0.47	3.69
NP21	8.10	0.13	–1.83	–3.77	–0.55	4.32
CR22	3.46	0.62	–0.22	–3.37	–0.86	4.23
NR23	7.75	0.15	–2.16	–3.51	0.09	3.42
UP24	22.55	0.00	–6.47	–2.65	0.04	2.61
NP25	18.47	0.00	–4.73	–3.78	–0.21	3.99
NR26	1.90	0.86	0.36	–3.26	0.17	3.09
NP27	3.96	0.54	–0.83	–3.30	–0.05	3.36
NP28	10.34	0.04	–3.05	–3.62	–0.49	4.11
NR29	2.90	0.71	0.75	–2.66	–0.72	3.38
CR30	3.58	0.60	–1.37	–1.74	–1.46	3.20
NP31	5.80	0.31	–3.22	–2.56	–1.54	4.11
CR32	12.44	0.00	0.06	–2.85	–1.45	4.30
LP33	12.91	0.00	–2.47	–3.38	–0.19	3.57
NP34	15.34	0.00	–3.84	–3.33	–0.31	3.64
UP35	15.26	0.00	–3.48	–3.49	–0.54	4.04
NP36	10.35	0.04	–3.37	–3.02	–0.41	3.43
LP37	26.30	0.00	2.65	–2.38	–0.84	3.23
CR38	25.02	0.00	2.81	–2.53	–0.65	3.18
CR39	18.33	0.00	–3.80	–1.90	–0.76	2.66
CR40	2.22	0.81	–1.94	–2.89	–0.67	3.56

*Key for Labels: 1st character – N = Need; L = Clarity; C = Cost; U = Other.
2nd character – P = Positive wording; R = Negative wording.

**The overall χ^2 is 640.4, which has a probability of 0.00 on 195 degrees of freedom.

APPENDIX C

Figures

This appendix contains all figures referred to in the text.

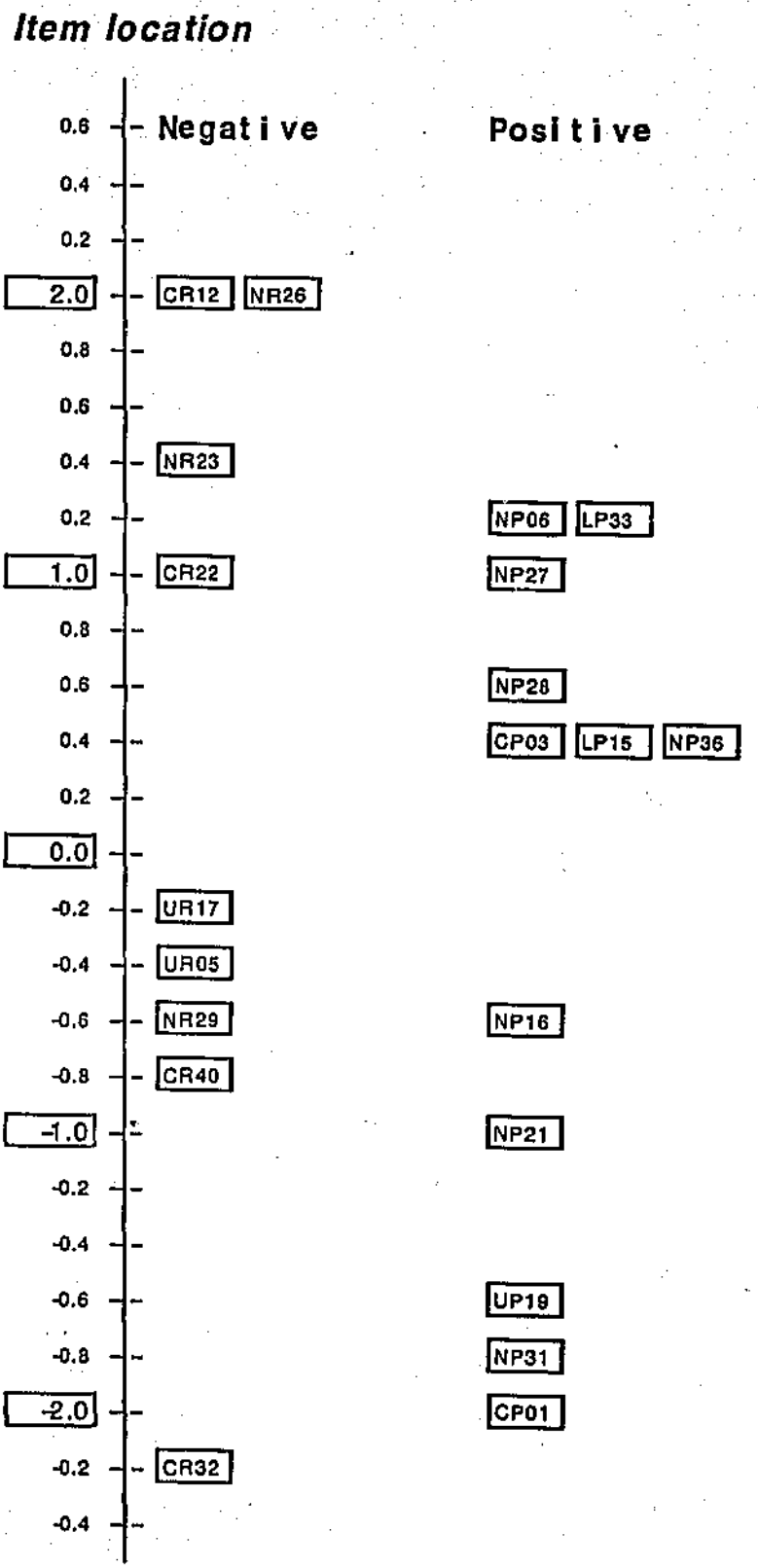

FIGURE C–1: Distribution of items along the Attitude to Direct Instruction Continuum arranged by negative or positive wording of statement