

Rash or Rasch?

A paper presented at the Annual Conference of the Australian Association for Research in Education. University of Newcastle, November 1994.

Ted Nettle
School of Teacher Education
The University of New South Wales

INTRODUCTION

A recent study by Dunkin, Precians and Nettle (1994) examined the effects of formal teacher education upon student teachers' cognitions regarding teaching. The study involved the use of individual interviews with a sample of 39 first year and 19 third year primary teacher education students. The interviews consisted of a series of open ended questions which were taped for subsequent analysis. The study found that there were changes in the beliefs which student teachers expressed about teaching practices. Changes were noted after a first session of practice teaching. Differences were also seen in the beliefs expressed by student teachers who had experienced different curriculum courses in their first semester and between first year and third year student teachers who were undertaking the same program.

Caution was exercised in the interpretation of the results obtained from the study because of the small sample sizes. The opportunity to use a larger sample, possibly a sample stratified across a number of courses and institutions would have required either the expenditure of financial and human resources which were beyond the scope of research budget, or the use of a different method of data gathering. In order to continue the research program it was decided that the development of a pencil and paper research tool, which could be used with larger samples, should be a priority.

Thus the aim of the present study was to develop a scale to measure beliefs about teaching practices which would enable the replication of the findings obtained from previous student interviews and the subsequent extension of the research to larger samples of student teachers and also to teachers.

A RASH MOVE?

The dominant research paradigm within the field of research on teacher thinking has been a qualitative one, ideographic in its emphasis, using small samples and a case study approach. It may appear, therefore, that a move to develop a questionnaire and conduct survey research in this field is a rather rash move. There have, however, been calls for this type of approach (Floden & Klinzing, 1990; Marland

& Osborne, 1990; Verloop, 1988). Indeed, after conducting an intensive case study of one teacher, Marland and Osborne recommended that, given the immaturity of this field of inquiry, "research into teachers' theories of action should continue to be nomothetic" (1990, p. 108).

A questionnaire developed to measure an aspect of teacher thinking would be one research tool among many which, while they may have individual limitations, will complement each other, to improve our understanding of how students learn to become teachers and how teachers think about teaching.

The issue here then is not one of the merits or demerits of the qualitative or quantitative research paradigms. Rather its one of

finding the most appropriate way to develop a questionnaire containing a scale for measuring the beliefs of teachers and student teachers.

Within the quantitative research tradition, questionnaires have been an important way of gathering data on the attitudes and beliefs of teachers and student teachers. Many different instruments have been used, but one that has had a strong influence on the research in this area is the Minnesota Teacher Attitude Inventory (MTAI) which was developed in 1951. Some consideration of its origins and use will be important here. In its original form it consisted of 150 items designed to provide a single score indicating a teacher's preference for "democratic" or for "autocratic" values in teaching. The unidimensionality of the scale was an issue for its authors (Horn & Morrison, 1965) and a common feature of its subsequent use has been the search for an underlying factor structure.

Horn and Morrison (1967) conducted a factor analytic study with 306 university education students and found five factors. Wehling & Charters, (1969) produced a questionnaire very similar in structure to the MTAI, although one with a more cognitive emphasis. They found eight dimensions of teacher beliefs after factor analysing responses from 291 classroom teachers. Wolfe and Engel (1978) combined 18 statements from the MTAI and 52 statements from the instrument developed by Wehling & Charters. They presented these to 364 education majors and the resultant factor analysis yielded three factors. More recently Bunting (1984) produced an inventory containing items similar those in the MTAI and the Wehling and Charters instrument and found four factors which in a later modified instrument (Bunting, 1985) reduced to two factors.

Each of these studies began with a large and varied collection of items and sought to determine an underlying factor structure using an exploratory factor analysis. The assumptions upon which they were based were derived from classical test theory. Within this theory, it is assumed that those items which are most highly correlated with one

another can be considered as measuring the same latent trait or factor.

Once the factors are discovered, the items which load most highly on them are considered to form a homogeneous pool which all measure the same trait in the same way.

The conceptual variety of some of the items is well illustrated by three items which form part of a single factor labelled as "directive" by Bunting:

#Students find it difficult to concentrate when there is activity in the classroom.

#If parents were firmer and more regular in the discipline of their children, the work of the teacher would be easier.

#The curriculum consists of subject matter to be learned and skills to be acquired.

(Bunting, 1984, p197)

Herein lies one of the problems with this approach to scale construction. The factors determined by correlating the responses to the items may not necessarily be measuring the same trait. They may be measuring different yet correlated dimensions of the beliefs of the teachers being surveyed.

Since the statements which defined the four pedagogical dimensions were generated by asking teachers and student teachers what they believed were the most important things they could do to enhance student learning, it is very likely that another group of teachers or student

teachers, responding to the same statements, would consider the teaching practices described in those statements as important for enhancing student learning. An exploratory factor analysis may therefore be expected to find most of the items contributing to one major factor.

In the present study the approach taken has been quite different from that taken in the studies referred to above. The aim was to develop scales for four pedagogical dimensions whose construct validity had been established from previous research. Thus, rather than discover the underlying factors, the aim was to design items and a response format which would allow the measurement of known factors.

A RASCH MOVE.

An alternative approach which would explore the relations among the items within each of the dimensions was needed. This was found in item response theory and in particular in the use of the Rasch model (Andrich, 1978; Tognolini & Andrich, in press; Wright, 1967; Wright, 1984). Whereas classical test theory focuses on the responses of a group of individuals to a whole test or subtest, item response theory

focuses on the particular interactions between individuals and items. While classical test theory attempts to summarise the overall relationship by fitting any of a range of mathematical model to the data, Rasch takes a particular mathematical model and tests whether the data fit the model. There are obvious limitations in that such an approach is only possible where the data meet the assumptions of the model. However, when they do, the model provides the means for understanding the responses of the individuals and the functioning of the items.

The Rasch model estimates the probability that a particular outcome will follow from an interaction between a person and an item, for example, the probability that a particular person will choose response 4 rather than response 3 or 5 on item 10 of an attitude scale. One of the features of the Rasch model which is important in this study is that the person's attitude can be estimated independently of the difficulty of the items and the items difficulty can be determined independently of the particular persons attempting the items.

Another feature of the Rasch model which is of importance in the present study is the requirement of unidimensionality. The Rasch model assumes that there is a unidimensional underlying trait which each of the items is measuring. Thus, if each of the pedagogical dimensions is considered to represent a single latent trait, it would be possible to develop from the items representing that dimension a scale for measuring the latent trait. Because the model requires unidimensionality and asks the question "does the item fit the model?" those items which do not fit the assumption of unidimensionality are shown up as misfitting items which are highlighted for the researcher. The researcher must then ask the question "why does this item not fit the model?"

It is at this point that the use of the Rasch model can be seen as involving aspect of both quantitative and qualitative approaches to research. The data collected in this study have been analysed by comparing their distribution with that predicted by a particular mathematical model. However, rather than rejecting some items on the basis of a number, such as an eigenvalue or a factor loading, an item does not fit the Rasch model will be subjected to further inquiry which will result in a much greater degree of understanding of the original data and will produce a more meaningful scale for measurement.

The final feature that is of importance for the present study is that the Rasch model is able to represent the ability, or in this case the affectivity, of the person and the difficulty of the item on the same interval scale. This feature of the model is of great benefit in the development of measurement scales.

The remainder of this paper will outline the way in which the Rasch model was used to assist in the development of scales for measuring teacher and student beliefs about teaching.

SCALE DEVELOPMENT

PROCEDURE

The statements made by student teachers in response to the question "What are the most important things you can do to enhance student learning?" (Dunkin, Precians, & Nettle, 1994) were collected and since they had all been classified as reflecting one or other of the four pedagogical dimensions: activity, motivation, interpersonal relationships or structuring, they formed the basic list of teaching practices for the present study. However, as the present study was also to include practising teachers, a sample of eight teachers was asked the same question. Where the responses of teachers and student teachers differed the teachers' statements were added to the list of teaching practices. Additional statements were obtained from recent literature on teaching effectiveness (Brophy & Good, 1986; Wang & Walberg, 1991). Examples of the statements used in the final questionnaire are provided in Table 1. The numbers are the numbers of the items in which the statements were placed.

Table 1 Statements used in the development of a Scale for Measuring Beliefs about Teaching Practices.

| ACTIVITY | MOTIVATION | RELATIONS | STRUCTURING |
|---|------------|--|-------------|
| 2b | | | |
| providing opportunities for themselves. | | students to think through problems for themselves. | |
| 2a | | | |
| saying something positive about each student's work. | | | |
| 2d / 16c | | | |
| communicating with students on the students' own level. | | | 2c / 8c |
| ensuring that material is presented at an appropriate level for students. | | | 8a / 10a |
| allowing students to question what they are learning. | | | 10d |
| encouraging students when they participate well in lessons. | | | 10b |
| 10b | | | |
| accepting and responding to students' feelings. | | | 10c |
| monitoring the progress of individual students. | | | 7c |
| asking students to find answers to questions from books in the library. | | | 7b |
| 7b | | | |
| awarding house points when students work well in class. | | | 7d |
| discussing students' personal problems with them. | | | 7a |
| demonstrating tasks before students are asked to do them. | | | |

As the aim of the study was to develop particular scales measuring the four dimensions of activity, motivation, interpersonal relationships and structuring it was important at the outset to establish the construct validity or unidimensionality of the items in each of the four scales. This issue was addressed by submitting all of the items to a panel of six teacher education experts who were given definitions of the four pedagogical dimensions and asked to classify the statements into these four categories.

Items on which at least four of the six judges categorised the statements the same way, were used in the development of the next instrument. Some items were modified as result of the feedback obtained and one item was classified as belonging to a different dimension.

A questionnaire was developed with 56 items, 14 in each of four scales. This was called the Teaching Practices Questionnaire (TPQ) and asked students to rate each of the items in terms of its importance for enhancing student learning. The response format therefore reflected that of the original study.

Earlier trials had indicated that responses to this form of questionnaire were strongly skewed towards the "important" end of the response continuum. Therefore, respondents were told that each of the teaching practices listed was believed to be important for enhancing student learning. However they were asked to indicate their perception of the degree of importance of the teaching practices on a six point scale where 5 represented "one of the most important" and 1 represented "one of the least important". The 0 (zero) was to be used if the respondent did not believe that a particular teaching practice was important for enhancing student learning. In this way it was hoped that respondents would be encouraged to use the whole of the scale from 5 to 1. Respondents were to write their answers in a box at the right of each statement.

The Teaching Practices Questionnaire was administered to 128 first year and 111 third year student teachers from a major metropolitan

university and to 147 primary school teachers teaching in metropolitan schools (n=386). The student questionnaires were administered to students at the beginning or end of a lecture session. The teacher questionnaires were posted to teachers along with a reply paid envelope. The response rate for the students was 99% (only two questionnaires were returned uncompleted). The response rate for the teachers was only 59%.

Each of the sub scales representing the pedagogical dimensions of activity, motivation, interpersonal relationships, and structuring was

analysed separately using the ASCORE program (Andrich, Sheldon, & Lyne, 1991) at the Educational Testing Centre of the University of New South Wales. This program used the Rasch model to estimate the importance of each item and the belief strength of each person. Because the Rasch model assumes unidimensionality of the items within the scale it was possible to examine those items which did not fit the model and to determine their appropriateness for that scale. All 56 items were also analysed using the ASCORE program to determine the extent to which all of the items could be considered to be unidimensional. Such a finding would be quite consistent with the manner in which the items for the questionnaire were initially gathered.

Finally a factor analysis (SPSS-X FACTOR program) was conducted to check the similarity between the factor structure obtained and the four pedagogical dimensions upon which the scales in the questionnaire were based. Again the finding of a single major factor would be consistent with the manner in which the items for the questionnaire were initially gathered.

RESULTS

The first set of analyses of the responses to the TPQ were conducted within each of the four pedagogical dimensions. The dimensions varied in the degree to which items fitted the model. Table 2 shows the number of items which fitted the assumptions of the Rasch model when each of the four dimensions was analysed in three different ways. The first analysis used the raw data from the questionnaires. The second analysis suppressed, for the purpose of scaling the items, those individual responses which were more than 2.5 standard deviations away from the mean. Results are shown for this analysis where there was a significant improvement in the fit of the items. The third analysis which was conducted for the motivation dimension only involved the deletion of two items which were consistently fitting very poorly.

Table 2 Number of Items Fitting the Rasch Model in each Pedagogical Dimension for Three Different Analyses

Table 2 indicates that it was possible to develop unidimensional scales, with a minimum of 9 items for each of the four pedagogical dimensions. However inspection of the location of the scale as defined by the items and the location of the people on the same scale indicated a major problem with the possible use of the questionnaire. The extent of this problem is shown in Table 3.

Table 3 shows the means and standard deviations for the locations of both items and people within each of the four pedagogical dimensions. While the standard deviations are similar in all cases, the means for the people are approximately two standard deviations above the means for the items. In other terms, this table indicates the skewed nature

of the distribution resulting from the items being considered by the students and teachers as "important" for enhancing student learning.

The result of this difference in the means for items and persons is that the scale would not provide a very accurate measure of the beliefs of the people. To use this scale would be like trying to measure the finals of the high jump at the Olympic Games with a high jump stand that was only graduated up to 1.5 meters. The measuring scale may be very accurate up to 1.5 meters but that scale would be of limited value for measuring the jumps of people who can jump beyond that height.

Table 3 Mean Location Values and Standard Deviations for Items and People within each Pedagogical Dimension

The Rasch analysis of the TPQ data had produced four measurement scales using most of the items within each of the four pedagogical dimensions.

However, in that form they were not likely to discriminate accurately between the beliefs of the majority of the students and teachers. The aim was ultimately to measure the beliefs of the students and teachers, therefore, it was necessary to find a way to discriminate between the individuals in terms of their preference for one or other of the pedagogical dimensions.

Since students and teachers saw most of the statements of teaching practices as being important, it was decided that the best way to ensure that the students and teachers distinguished between the dimensions was to produce a ranking instrument which forced the respondents to choose among the four dimensions. The resultant ipsative data were considered acceptable since the questionnaire was only seeking to determine the relative values given to the four scales by each individual, not to compare one scale with another.

The Pedagogical Dimensions Questionnaire (PDQ), developed for this purpose, required respondents to rank the four alternatives within each item from 1 to 4 where 1 was most important and 4 was least important in terms of the importance of that alternative for enhancing student learning. An individual's score on any given dimension would be the sum of his or her first choices. To ensure that sufficient choices could be made, it was decided to increase the number of items from the minimum of 9 which had been found to be unidimensional. This meant either repeating a lot of items or using more of the original items. Ultimately all but one of the original items were used and two, and in one case three of the items, were repeated in each dimension giving the PDQ a total of 16 questions. Table 4 shows the item numbers from the PDQ, arranged in columns representing the four dimensions and arranged

vertically on a Logit scale determined from the administration of the TPQ. The statement at the top of the Relations column marked with the symbol ## was deleted from the final questionnaire.

It can be seen that the items were chosen in order that each question would contain an item of much the same level of affectivity from each dimension. This ensured as far as possible that the respondents' choices would be based on their beliefs about the dimensions and not on differing levels of affectivity.

Examination of Table 4 also shows that the Motivation items were distributed differently from the other items with a definite gap from 0 to -1 on the logit scale. Because of this that it was decided to retain the two items which had been deleted in the earlier analysis using ASCORE (items 7b and 15b). This avoided too much repetition of items, a feature which initial trialling had shown students found irritating while maintaining a similar level of affectivity on the logit scale for each question. The item from the TPQ that was deleted from the PDQ was the most positive item in the relationship scale. It can be seen that three items were grouped together at the positive end of this scale such that to use all three would have imbalanced the level of affectivity of some of the questions.

Table 3 Items within Dimensions, arranged according to placement on Logit Scale after administration in TPQ

The Rasch analysis of all the items in the TPQ found that 30 of the 56 items fitted the model suggesting that these items represented a single latent trait. The factor analysis of the TPQ indicated that a single factor accounted for 35% of the variance and had loadings of 0.5 or over from 23 of the 56 items. Twelve items were common to the two groups. These findings were as expected given the manner in which the original items were collected.

TRIAL OF THE PDQ

PROCEDURE

The PDQ was administered to all of the students in the first year of the primary teacher education course at a large metropolitan university (n=83) both before and after a three week period of practice teaching.

Interviews were conducted with a sample of 12 of the first year students after they had completed practice teaching and after they had completed the second administration of the PDQ. The students were asked: "What are the most important things you can do to enhance student learning?" (Dunkin, et al., 1994) thus providing an opportunity

to examine the extent to which the PDQ could replicated findings obtained from individual interviews. The students' responses to this question were categorised separately by the author and one of the researchers who classified the student responses in the original interview study.

Data from the 12 interviews were analysed in the same manner as those collected by Dunkin, Precians and Nettle (1994) in order to provide a comparison with the data collected from the same students using the PDQ. This was intended as a validity check, indicating the extent to which the two approaches, interview and questionnaire, could produce similar results. A further analysis was undertaken of the results obtained by using the results obtained from using the PDQ before and after practice teaching to determine if any change in beliefs could be discerned. A separate check was made to ascertain that there was no practice effect when the PDQ was used a second time after a space of five to seven weeks.

RESULTS

Dunkin, Precians and Nettle (1994) reported the results of their interviews in terms of the mean number of comments students made in each pedagogical dimension as well as for a combination of cognitive and affective dimensions. Table 5 provides the same information for the 12 students who were interviewed after they had completed practice teaching.

Table 5 Post-Practice Teaching Responses to PDQ and Interviews

The differences in the means is an artefact of the different data gathering techniques. The PDQ figures were obtained by summing the number of times a statement in any dimension was chosen first in any of the 16 questions. The total number of responses across the four dimensions is 16. The interview figures were based on the total comments supplied by the students in response to the question asked.

The important feature of these data is that the rank order of the four dimensions is the same in the PDQ and the interview. In each case structuring was chosen or mentioned more frequently than the other

dimensions and relationships were the least frequently chosen or mentioned.

The analysis of changes in student beliefs from before practice teaching to after practice was also examined by Dunkin, Precians and Nettle (1994) and comparison with that result is also possible here. Table 6 contains these results. The means in Table 6 are the means of the number of times each student ranked the particular dimension first

in the PDQ. The maximum possible number of first rankings was 16 but the scores in the PDQ ranged from 12 to 0.

Table 6 Pre-Practice Teaching and Post-Practice choices of Students on PDQ

Dunkin Precians and Nettle (1994) found changes in student teachers beliefs after practice teaching. They found that while there were no statistically significant results, there were moderate effect sizes which indicated an inclination for the students to make more references to structuring and activity after practice teaching. Both of these trends were evident in the data obtained using the PDQ although they were neither statistically significant nor marked by anything other than a very weak effect sizes. However in the present study the drop in responses to relationships was significant with a probability of only 0.006. thus the pattern of change toward a more custodial approach after practice teaching (Dunkin, et al., 1994; Hoy & Woolfolk, 1990; Weinstein, 1990) was again confirmed in this study.

SUMMARY AND CONCLUSION

The aim of this study was to develop a scale that would measure student teachers' beliefs about teaching practices and replicate the findings of an interview study. The results above indicate that this aim has been achieved. The Pedagogical Dimensions Questionnaire provided a means of measuring student teachers beliefs about teaching and those beliefs were essentially similar to those obtained by interview questioning. In addition, the PDQ was able to detect changes in student teachers beliefs about teaching that were similar to those obtained in the earlier study by Dunkin, Precians and Nettle (1994).

The development of a measurement scale was achieved in the Teaching Practices Questionnaire although it was found that this scale was unsuitable for measuring the beliefs of student teachers and teachers because it was calibrated to measure a different range on the scale of affectivity than that range which was relevant to the students and the teachers. However, the procedure used to develop that scale, involving the use of the Rasch model, proved to be valuable for arranging the items in such a way that each question in the final questionnaire contained items at a similar level of affectivity.

In the context of research on teacher thinking, the use of the Rasch model also proved valuable as a means of undertaking what some may consider to have been a rash task: the development of a questionnaire to gain information about teacher and student teacher beliefs about teaching practices. While the Rasch model provides a very quantitative approach to measurement it also forces the researcher to adopt a

qualitative approach in trying to understand the measuring scale that is being developed and the data that have been collected. This instrument will be used in further research and it is hoped that the nett result will be a better understanding of the role of the beliefs about teaching practices held by students and teachers in both preservice and inservice teacher education.

REFERENCES

- Andrich, D. (1978). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement*, 38, 665-680.
- Andrich, D., Sheldon, B., & Lyne, A. (1991b). ASCORE: Manual of Procedures. Unpublished Manual
- Brophy, J., & Good, T. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 328-375). New York: Macmillan.
- Dunkin, M. J., Precians, R. B., & Nettle, E. B. (1994). Effects of formal teacher education upon student teachers' cognitions regarding teaching. *Teaching and Teacher Education*, 10(4), 395-408.
- Floden, R. E., & Klinzing, H. G. (1990). What can research on teacher thinking contribute to teacher preparation? A second opinion. *Educational Researcher*, 19(4), 15-20.
- Horn, J. L., & Morrison, W. L. (1965). Dimensions of teacher attitudes. *Journal of Educational Psychology*, 56(3), 118-125.
- Hoy, W. K., & Woolfolk, A. E. (1990). Socialization of student teachers. *American Educational Research Journal*, 27(2), 279-300.
- Marland, P., & Osborne, B. (1990). Classroom theory, thinking and action. *Teaching and Teacher Education*, 6(1), 93-109.
- Tognolini, J., & Andrich, D. (in press). Differential performance and the problems of selection. *Assessment and Evaluation in Higher Education*, 20(2).
- Verloop, N. (1988). Investigating teacher cognitions: A report of the third ISATT conference. *Journal of Curriculum Studies*, 20(1), 81-85.
- Wang, M. C., & Walberg, H. J. (1991). Teaching and educational effectiveness: Research synthesis and consensus from the field. In H. C. Waxman & H. J. Walberg (Eds.), *Effective Teaching: Current Research* (pp. 81-104). Berkeley: McCutchan Publishing Company.
- Weinstein, C. S. (1990). Prospective elementary teachers' beliefs about teaching: Implications for teacher education. *Teaching and Teacher Education*, 6(3), 279-290.
- Wright, B. D. (1967). Sample-free test calibration and person measurement. In *Invitational Conference on Testing Problems*, (pp. 85-101). Educational Testing Service.
- Wright, B. D. (1984). Despair and hope in educational measurement.

Contemporary Education Review, 3(1), 281-288.

NETTLE

AARE 1994