CONSTRUCTING AND VALIDATING AN INSTRUMENT TO MEASURE STUDENTS' ATTITUDES AND BELIEFS ABOUT LEARNING MATHEMATICS

Sinead Breen

CASTeL & Department of Mathematics, St Patrick's College Drumcondra

Joan Cleary

Department of Mathematics, Institute of Technology, Tralee.

Ann O'Shea

Department of Mathematics, NUI Maynooth.

When embarking on a study of attitudes and opinions, it is important to design an instrument of inquiry that will provide valid, reliable and interpretable information that addresses the specific question of interest. Reported here are attempts made to evaluate the reliability and validity of a survey instrument using Rasch analysis, factor analysis, and Cronbach's alpha. The instrument aimed to measure students' confidence and anxiety in relation to learning mathematics, and to examine their goal orientation and beliefs on the nature of intelligence, by recording their responses to a number of statements using a Likert-style format. A previous study conducted by the authors failed to endorse Dweck's (1986) claims that a student's theory of intelligence and confidence in his/her present ability combine to determine the student's behaviour when presented with an unfamiliar and/or challenging task. The present study was initiated in order to provide the opportunity to investigate this claim more comprehensively. The techniques (of Rasch and factor analysis) used to investigate the validity of the instrument designed, also facilitate the computation of a quantitative measure of the level of each trait possessed by a student.

INTRODUCTION

Motivation for Study

Dweck (1986) maintains that a student's theory of intelligence (as to whether intelligence is fixed or malleable) and confidence in his/her present ability combine to influence the student's behaviour when presented with an unfamiliar task. As part of a study carried out in 2006 and 2007, first year students in three third level institutions in Ireland were asked whether they believed that mathematical ability could be improved and were asked to rate their confidence in approaching mathematics. However, in that study it was found that an individual's level of confidence but not his/her view of the nature of intelligence played an important role in how he/she approached, persevered with and performed on unfamiliar tasks (Breen, Cleary & O'Shea, 2007). A more comprehensive survey on third-level students' beliefs about the nature of mathematical intelligence and the level of confidence felt in approaching mathematical tasks and courses was undertaken to investigate these issues further. Explicitly included in the questionnaire were items addressing goal orientation, tendency to seek out or avoid challenges, and persistence in the face of difficulty (or unfamiliarity) as Dweck maintains that a child's theory of intelligence orients him towards a certain class of goal (learning or performance) and that this is turn influences his pattern of behaviour (adaptive or maladaptive) which can be identified by the child's attitude towards challenges and the levels of persistence or perseverance he displays (Dweck, 1986; Elliot & Dweck, 1988).

Traditionally, factor analysis has been used to validate rating-scale instruments by seeking to establish the underlying factorial structure of the data, uncovering the dimensions inherent in the dataset. For instance, in an investigation of the process by which theories of intelligence and performance affect student learning, Stipek and Gralinski (1996) used factor analysis to identify the dimensions underlying their items assessing beliefs, and verified the reliability of the two factors found by computation of the associated alpha coefficients. Mulhern and Rae (1998) evaluated the reliability of the Fennema-Sherman Mathematics Attitudes scales for a sample of 196 secondary schoolchildren using Cronbach's alphas and then also employed factor analysis to investigate the validity of these scales. Their findings lead them to develop a shorter instrument involving only 51 of the original 108 items, while retaining a comparable level of reliability.

In recent times, Rasch analysis has also been used for the purpose of instrument validation. A systematic evaluation of the impact of technology on the teaching and learning of mathematics was reported by Galbraith and Haines (1998). They used Rasch analysis to judge the effectiveness of the rating-scales they developed, including reliability and goodness-of-fit. Apparent associations between their attitude scales (e.g. mathematics confidence, computer confidence) were further explored using factor analysis. Bradley, Sampson & Royal (2006) emphasised the fundamental role the quality of a measurement tool plays in the analysis of the data it seeks to produce and lamented the lack of attention sometimes suffered by this phase of a research study. They employed Rasch analysis to address this concern for their own survey instrument. Meanwhile, Grimbeek and Nisbet (2006) explained how they had previously used factor analysis to examine a dataset of primary teachers' beliefs about compulsory numeracy testing in Queensland, but decided to improve on the outcome by utilising a combination of Rasch analysis and an iterative sequence of factor analyses. They reported this approach to have produced an instrument with a factor structure statistically and conceptually more elegant than their initial model. The fact that Rasch analysis explicitly takes into account the categorical and ordinal nature of Likert-scale items motivated its use by Grimbeek and Nisbet.

METHOD

Description of Study - Design & Administration of Questionnaire

The questionnaire used collected personal information (including gender, age, level of mathematics achievement at post-primary school) from the participants as well as recording responses to sets of rating-scale items relating to Confidence, Anxiety, Theory of Intelligence, Goal Orientation (Learning/Mastery and Performance) and Persistence. All rating-scale items were presented using a 5-point Likert scale where 1 represented 'disagree strongly', 2 'disagree', 3 'not sure', 4 'agree' and 5 'agree strongly'. In developing the questionnaire, the authors were mindful of including sufficiently many items to enable a statistically significant analysis of the data collected, while taking into account the time it would take respondents to complete it. The Confidence and Anxiety items used appear in tables 1 and 2 in the Results section, while the remaining rating-scale items are included in the appendix. (Note that negatively worded questions were coded in reverse order before analysis.)

Most of the items on the rating-scales exploring Goal Orientation (Learning/Performance Goals) and Theory of Intelligence were those used by Stipek & Gralinski (1996), reworded slightly to use terms appropriate to third-level students in Ireland. Theory of Intelligence 2 was modified from Schoenfeld (1989). Items on the Confidence and Anxiety rating scales were similarly adapted from the Fennema-Sherman scales (see Mulhern & Rae, 1998) and those used in the PISA 2003 student questionnaire (OECD, 2003), with the exception of Confidence 4 which was taken from Pietsch, Walker and Chapman (2003). The remainder of the items were constructed for this study, including the Persistence items which were based on hypotheses put forward by Dweck and Elliott as to the behaviour of, and strategies employed by, students with differing goal orientations (Dweck, 1986; Dweck & Elliott, 1988).

186 students from three third level institutions voluntarily participated in this study that was conducted in the second semester of the 2007/2008 academic year. All participants were in the first year of their respective programmes, namely BEd or BA (Humanities) at St Patrick's College, Drumcondra, BA or BA (Finance) at the National University of Ireland, Maynooth and Higher Certificate in Engineering or BSc at Institute of Technology, Tralee. Participants were given approximately 20 minutes to respond to the questionnaire.

Introduction to Rasch Analysis

Rasch analysis is a means of constructing an objective fundamental measurement scale from a set of observations of ordered categorical responses (to assessment items or rating-scale items). In the case of rating-scale items, some will be easier to endorse than others and this is reflected in the measurement scale produced so that the items that are easiest to endorse appear at the bottom of the scale while those that are most difficult to endorse are at the top. For instance, it seems reasonable that the item 'I am so afraid of maths that I avoid going to maths classes' indicates a higher level of math-anxiety than 'I am sometimes afraid that I will make mistakes when doing maths problems', and so the former should appear at a higher level on the resulting measurement scale than the latter. The scale produced is an interval one, and so, a one-unit difference between the positions of a pair of items (that is, between the item difficulties) at any point on the scale reflects the same level of difference in the underlying trait being measured. Moreover, the Rasch model places the respondents on the same scale as the items and relates the two via a probabilistic function of the difference between their positions. Following the assumption that useful measurement involves the consideration of a single trait or construct at a time, the Rasch model incorporates a quality control mechanism, by means of error estimates and fit statistics, to verify that all items are contributing to the measurement of the trait of interest.

RESULTS

Rasch Analysis

In order to usefully employ measures of the affective traits that we mean to study in this project, we need to be confident that the measures are constructed in a scientifically valid manner. To this end, we used Rasch analysis to validate our scales. Item fit statistics as well as reliability statistics were used to check that all items on a scale were addressing the same construct and that the measures were reliable.

We used two specialist software packages to analyse our data. They were Conquest and Winsteps (Linacre, 2009). Both packages gave similar results and we will report the results from the Winsteps analysis here. As the items to be analysed were of Likert-scale type (with responses 1-5, and higher numbers representing higher levels of agreement), it was appropriate to apply a Partial Credit Model (Masters, 1982) to the data. The software estimates the items' difficulty levels and error estimates in logits. It also computes both weighted and un-weighted mean-squared residuals for each item. The residuals represent the differences between the model's theoretical expectation of item performance and the performance actually encountered. Each raw residual is standardised using its variance before inclusion in the un-weighted mean square. In the weighted mean-square the standardised residuals are also 'information-weighted' using the variance of the observed performance: this reflects the fact that better targeted observations provide more information for the analysis of the data. Therefore, the un-weighted mean-square is more sensitive to outliers. The un-weighted mean-square is usually called the outfit statistic and the weighted mean-square is called the infit statistic. The mean squares are chi-squared statistics divided by their degrees of freedom and thus have expected values of 1. Bond and Fox (2007, p.243) report that a reasonable range of infit and outfit statistics for Likert scale items is 0.6 to 1.4. Any items whose infit or outfit does not lie in this range should be investigated.

The Rasch model also provides person and item reliability indices. The item reliability index indicates how stable the item estimates are, that is it tells us how likely it is that the estimates would remain the same if the items were given to a different group of the same size and similar behaviour patterns. The person reliability index tells us how likely it is that the person ordering would remain the same if the sample were given a new test with similar items. Rasch person reliability measures are comparable to traditional Cronbach's alphas with the proviso that the former tend to underestimate the true reliability while the latter tend to overestimate it (Linacre, 2009). Thus, the same criteria recommended for Cronbach's alphas will be employed here for Rasch person reliabilities: Kline (1993, p.11) recommends that ideally Cronbach's alphas should be high, in the region of 0.9, and that they certainly shouldn't fall beneath 0.7. The Cronbach's alpha statistic for all 38 questions was 0.918, comfortably satisfying Kline's criterion. Cronbach's alphas were also computed for the individual scales and are reported below. Low item reliability indicates that the sample size may be too small for stable comparison of items (Linacre, 2009): in what follows, it can be seen that the item reliabilities were all greater than 0.9 and also always greater than the associated person reliabilities, and so they were deemed to indicate acceptable reliability of items.

Confidence Scale

Table 1 contains the measures and infit and outfit item statistics (as calculated using Winsteps) for the Confidence items. It is clear that all items are behaving well in this scale. The person reliability index is 0.82 and the item reliability index is 0.94. Thus we can be confident that the items are measuring the same trait, and that our person measures will be reliable. The Cronbach's alpha value for this scale was 0.855.

Item	Estimate	Error	Infit MNSQ	Outfit MNSQ
Confidence 1: I learn mathematics quickly.	0.46	0.11	1.01	0.98
Confidence 2: I feel confident in approaching mathematics.	0.65	0.11	0.81	0.73
Confidence 3: I can get good marks in mathematics.	-0.18	0.13	1.11	1.10
Confidence 4: I have trouble understanding anything with mathematics in it.	-0.77	0.12	1.06	1.00
Confidence 5: Mathematics is one of my worst subjects.	0.06	0.10	0.83	0.91
Confidence 6: I am just not good at mathematics.	-0.23	0.11	1.15	1.28

Table 1: Item fit for the confidence scale

The item appearing at the bottom of the scale (with estimate –0.77) is Confidence 4, 'I have trouble understanding anything with mathematics in it'. Taking into account the negative wording, this indicates that the participants were able to disagree with this item most easily. At the top of the scale (with estimate 0.65) is Confidence 2, 'I feel confident in approaching mathematics', indicating that this item was the most difficult one for the participants to endorse.

Anxiety Scale

There were six anxiety questions in our questionnaire. We can see from Table 2 (overleaf) that one of the anxiety items does not perform well. Anxiety 5 has very large infit and outfit values. On reflection, this item, 'I like contributing and answering questions in maths class', seemed to have a slightly different focus to the other scale items. As a result, we decided to drop this question from the scale. When we did this all remaining items had infit and outfit values in the allowable range and the person reliability index (0.81) and item reliability index (0.98) for the reduced scale were good. The Cronbach's alpha value for the full scale was 0.825 but this increased to 0.847 when Anxiety 5 was deleted.

Item	Estimate	Error	Infit MNSQ	Outfit MNSQ
Anxiety 1: I get very nervous during maths classes.	-1.38	0.10	1.13	1.04
Anxiety 2: I often worry that it will be difficult for me in maths classes.	0.42	0.09	0.70	0.67
Anxiety 3: I often feel helpless when doing a maths problem.	0.38	0.09	1.09	1.17
Anxiety 4: Mathematics makes me feel uneasy and confused.	-0.54	0.10	0.71	0.66
Anxiety 5: I like contributing and answering questions in maths class.	0.97	0.10	1.64	1.76
Anxiety 6: I usually feel at ease doing mathematics problems.	0.16	0.10	0.76	0.75

Table 2: Item fit for the anxiety scale

Theory of Intelligence Scale

There were seven items on the theory of intelligence scale. An examination of the fit statistics computed for this scale showed Theory of Intelligence 6 to have a high outfit value (1.41) and so it was decided to drop item 6 from this scale. The new scale behaved well when the analysis was run again. The new scale has person reliability index of 0.72 and item reliability index of 0.98. The original scale had a Cronbach's alpha of 0.756 and this increased slightly to 0.757 for the reduced scale.

Persistence Scale

The persistence scale contained seven items. The items mostly concern students' strategies when faced with difficult problems. In this study, persistence will also be measured using the students' performance on the difficult items on a mathematical literacy test. The fit statistics computed in this case showed all items to be behaving well. The person reliability index is 0.69 and the item reliability index is 0.97. Thus no adjustments need to be made to this scale. The Cronbach's alpha value for this scale is 0.737.

Learning Goal Scale

There were five questions on the learning goal scale. Once again all items were seen to have infit and outfit values in the acceptable range. The person reliability index was 0.8 and the item reliability index was 0.96. Thus we can be confident that our scale is uni-dimensional and the measures derived from it can be trusted. The Cronbach's alpha value for this scale was 0.829.

Performance Goal Scale

This scale had seven items. Again all items had acceptable infit and outfit values. The person reliability index for this scale was 0.67, the item reliability index was 0.96 and Cronbach's alpha statistic was 0.752. Thus this scale is behaving well.

Factor Analysis

In order to further confirm that the questionnaire administered was measuring the traits intended, the entire dataset was subjected to an exploratory factor analysis, with the number of factors sought set to six following an examination of the scree plot (see Kline, 1994; Thompson, 2004). Using Principal Components Analysis and oblique rotation, items were identified with factors according to the principle of allowing as many items as possible to be identified with one factor while minimizing the number identified with more than one (minimum saliency criterion of |0.51|) (following Mulhern & Rae, 1998, p.299). In general, the factors thus obtained matched the traits under investigation with the first factor combining Confidence and Anxiety, the second factor representing Performance Goals, the third predominantly Learning Goals and the fourth Theory of Intelligence. The final two factors contained Persistence items 2, 4, 5, 6, 7. There was further agreement with the Rasch analysis results in terms of identifying items that did not behave as expected: for instance, Theory of Intelligence 6 was not identified with any factor, while Anxiety 5 appeared with the Learning Goals in the third factor – Rasch analysis indicated that neither of these items fit well with their respective scales. Although, this factor analysis also suggested a small number of other anomalies in the dataset (for instance, it failed to associate Persistence 1 or Performance Goal 2 with any of the six factors), it strongly corroborated the results of the Rasch analysis. Further factor analyses using different methods of factor extraction and rotation and/or various criteria for the retention of factors produced very similar results.

Measures for the Confidence Scale

Following validation of the scales, we computed a measure of each student's confidence using Winsteps. The measures ranged from -5.39 logits to 5.45 logits. For instance, a student with the pattern of responses 5,5,5,1,1,2 (where 1 represents 'disagree strongly' and 5 represents 'agree strongly') to Confidence 1-6 in order was allocated 5.45 on this scale, whereas another with responses 2,2,2,2,4,2 obtained –0.92. The mean confidence measure was 0.9381 logits. Moreover, using Winsteps' facility to produce an item-person map, which shows the positions of the items (in terms of their difficulties) and the participants (in terms of their levels of confidence) on the same scale, indicated that the Confidence items used were well matched to the participants. That is, there was a suitable range of Confidence items to cater for the range of levels of confidence possessed by the participants.

Using these measures of confidence we investigated some hypotheses. We found a significant difference between the mean confidence measures of students who had taken Higher Level mathematics at secondary school and those who had taken Ordinary Level (t-test, $p<0.001$). Higher Level students had much higher confidence measures than the Ordinary Level students. When we looked at the group as a whole we did not find a significant difference between the confidence levels of males and females. For the Higher Level students we again found no significant gender difference, however when the Ordinary Level students were considered, there was a significant difference between the male and female students (t-test, p=0.034), with the male students displaying higher levels of confidence.

CONCLUSION

The survey instrument used, following removal of just two of the original items, was found to be both valid and reliable by Rasch analysis. Factor analysis and the computation of Cronbach's alphas provided support for these findings. Despite the self-selecting nature of the sample of participants, the range of trait levels observed matched well with the range of participants (as evidenced in the Item-Person maps produced by Winsteps), further endorsing the appropriateness of the instrument for the sample. This provides a strong foundation on which to lay further exploration of Dweck's assertion of the relationships between confidence, goal orientation and beliefs about the nature of intelligence, through the construction of numerical measures of the levels of these traits possessed by the participants. The existence of such measures on the same measurement scale as the difficulty of the items themselves (an attractive feature of the Rasch model) enhances this exploration. Some work has been undertaken on the construction of estimates of a participant's confidence and its relationship to gender and prior mathematical achievement. A similar approach will be taken in the measurement of students' goal orientation and beliefs in relation to the nature of intelligence, before investigating the interplay between all three.

REFERENCES

Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model – Fudamental measurement in the human sciences* (Second edition). Lawrence Erlbaum Associates, New Jersey.

Bradley, K., Sampson, S. & Royal K. (2006). Applying the Rasch rating scale model to gain insights into students' conceptualisation of quality mathematics instruction. *Mathematics Education Research Journal*, 18(2), 11-26.

Breen, S., Cleary, J. & O'Shea, A. (2007). A study of third level students' beliefs about mathematics. In S. Close, T. Dooley, D. Corcoran (Ed.s), *Proceedings of the Second Conference on Research in Mathematics Education*, Dublin.

Dweck, C., (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040-1048.

Elliott, E. & Dweck, C. (1988). Goals: an approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54 (1), 50-12.

Galbraith, P. & Haines, C. (1998). Disentangling the nexus: Attitudes to mathematics and technology in a computer learning environment. *Educational Studies in Mathematics*, 36, 275-290.

Grimbeek, P. & Nisbet, S. (2006). Surveying primary teachers about compulsory numeracy testing: Combining factor analysis with Rasch analysis. *Mathematics Education Research Journal*, 18(2), 27-39.

Kline, P. (1993). *Handbook of psychological testing*. Routledge, London.

Kline, P. (1994). *An easy guide to factor analysis.* Routledge, London (& New York).

Linacre, J. M. (2009). *Winsteps (Version 3.68)* [Computer Software]. Chicago: Winsteps.com.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Mulhern F. & Rae, G. (1998). Development of a shortened form of the Fennema-Sherman mathematics attitudes scales. *Educational and Psychological Measurement*, 58 (2), 295-306.

OECD (2003). *PISA 2003 Student Questionnaire*. Retrieved June 26, 2009

http://www.pisa.oecd.org/dataoecd/34/7/37617728.pdf

Pietsch, J., Walker, R. & Chapman, E. (2003) The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology*, 95 (3), 589-603.

Schoenfeld, A. (1989) Explorations of students' mathematical beliefs and behaviour. *Journal for Research in Mathematics Education*, 20 (4), 338-355.

Stipek, D. & Gralinski, J. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology*, 88 (3), 397-407.

Thompson, B. (2004), *Exploratory and confirmatory factor analysis*. American Psychological Association, Washington.

APPENDIX

Rating Scale Items from Questionnaire
Theory of Intelligence 1. You have to be smart to do well in maths. 2. People are either good at maths or they are not. 3. Some people will never do well in maths no matter how hard they try. 4. You can succeed at anything if you put your mind to it. 5. You can succeed at maths if you put your mind to it. 6. It is possible to improve your mathematical skills. 7. Everyone can do well in maths if they work at it.

Persistence 1. I will risk showing that I don't know something in order to acquire new mathematical knowledge. 2. I am most proud of my mathematical performance when I feel I have done my best. 3. When presented with a choice of mathematical tasks, my preference is for a challenging task. 4. When presented with a mathematical task I cannot immediately complete, I increase my efforts. 5. When presented with a mathematical task I cannot immediately complete, I persist by changing strategy. 6. When presented with a mathematical task I cannot immediately complete, I give up. 7. When presented with a choice of tasks, my preference is for one I know I can complete.
Learning Goals 1. I work at maths because I like finding new ways of doing things. 2. I work at maths because I like learning new things. 3. I work at maths because I like figuring things out. 4. I work at maths because I want to learn as much as possible. 5. I work at maths because it is important for me that I understand the ideas.
Performance Goals 1. I work at maths because I want other people to think I'm clever. 2. I work at maths because it is important to me that the lecturer/tutor thinks I do a good job. 3. I work at maths because I don't want people to think I'm stupid. 4. I work at maths because it is important for me to do better than the other students. 5. I work at maths because I don't want to do worse than the other students in the class. 6. I will sacrifice acquiring new mathematical knowledge in order to avoid looking stupid. 7. I am most proud of my mathematical performance when I feel my performance made me look good.