A COMPARISON BETWEEN TEACHERS' AND PUPILS' TENDENCY TO USE A REPRESENTATIVENESS HEURISTIC

Thekla Afantiti Lamprianou, Julian Williams and Iasonas Lamprianou

University of Manchester

This study builds on a previous research on children's probability conceptions and misconceptions due to the representativeness heuristic. Rasch measurement methodology was used to analyse fresh data collected when a 10-item instrument (described by Afantiti Lamprianou and Williams, 2002, 2003) was administered to a new sample of 754 pupils and 99 teachers. A hierarchy of responses at three levels is confirmed for the teachers' sample, but a hierarchy of two levels is constructed for the pupils' responses. Each level is characterised by the ability to overcome typical 'representativeness' effects, namely 'recency', 'random-similarity', 'base-rate frequency' and 'sample size'. Less experienced teachers had a better performance on the instrument. The educational implications of our findings are discussed.

INTRODUCTION AND BACKGROUND

This paper builds on previous work on pupils' understandings and use of the representativeness heuristic in their probabilistic thinking (Afantiti Lamprianou and Williams, 2002, 2003). One of the aims of the Afantiti Lamprianou and Williams study was to contribute to teaching by developing assessment tools which could help teachers diagnose inappropriate use of the representativeness heuristic and other modes of reasoning based on the representativeness heuristic. The misconceptions based on the representativeness heuristic are some of the most common errors in probability, i.e. pupils tend to estimate the likelihood of an event by taking into account how well it represents its parent population (how similar is the event to the population it represents) and how it appears to have been generated (whether it appears to be a random mixture).

Williams and Ryan (2000) argue that research knowledge about pupils' misconceptions and learning generally needs to be located within the curriculum and associated with relevant teaching strategies if it is to be made useful for teachers. This involves a significant transformation and development of research knowledge into pedagogical content knowledge (Shulman, 1987). Pedagogical Content Knowledge (PCK) "goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching" (Shulman, 1986, p.9). Pedagogical Content Knowledge also includes the conceptions and preconceptions that students bring with them to the learning. If those preconceptions are misconceptions, teachers need knowledge of the strategies most likely to be fruitful in reorganizing the understanding of learners. Many studies have found that teachers' subject knowledge and pedagogical content knowledge both affect classroom practice and are modified and influenced by practice (Turner-Bisset, 1999).

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2, pp. 9-16. Melbourne: PME.

Along the same lines, Norman (1993) stresses that "there is little in research literature documenting either what teachers know or the nature of their knowledge" (Norman, 1993, p.180). What is more, Hadjidemetriou and Williams (2002) found that some teachers harbour misconceptions themselves (Hadjidemetriou and Williams, 2002). Godino, Canizares and Diaz (n.d.) conclude in their research that very frequently teachers do not have the necessary preparation and training in probability or statistics in order to teach efficiently; they also concluded that student teachers may have various probabilistic misconceptions themselves and this might affect their teaching.

Bearing that in mind, the instrument that was piloted and calibrated to the pupils in our study mentioned above (Afantiti Lamprianou and Williams, 2002, 2003) was now administered to a new sample of pupils and teachers. The administration of this diagnostic instrument to the teachers aimed to investigate (a) whether teachers' probabilistic thinking was affected by the 'representativeness' heuristic and (b) whether teachers were aware of these common misconceptions or of the significance of the representativeness heuristic. This was achieved by asking the teachers not only to answer the items themselves, but also to predict the common errors and misconceptions their pupils would be likely to make on each item, in the manner of Hadjidemetriou and Williams (2002) for a similar instrument assessing graphicacy. Finally, the results of the analyses of the teachers' and pupils' responses are compared.

METHOD

Ten items were used to construct the instrument (reached at http://lamprianou.no-ip.info/pme29/). The items identify four effects of the representativeness heuristic; the *recency* effect, the *random-similarity* effect, the *base-rate frequency* effect and the *sample size* effect. Most of the items have been adopted with slight modifications of these used in previous research (Green, 1982; Kahneman, Slovic and Tversky, 1982; Shaughnessy, 1992; Konold et al, 1993; Batanero, Serrano and Garfield, 1996; Fischbein and Schnarch, 1997; Amir, Linchevski and Shefet, 1999). Other items were developed based on findings of previous research.

The items were divided into three parts. The first part consisted of multiple-choice answers and the respondents were asked to choose an option. In the second part the respondents were asked to give a brief justification for their choice by answering the open-ended question 'Explain why'. Part three was only available in the Teacher version of the instrument and asked teachers to predict which common errors and misconceptions they would expect pupils to make on each question.

Since all items had both a multiple-choice and an open-ended question, a common item Partial Credit analysis (Wright and Stone, 1979; Wright and Masters, 1982) was run. One mark was given for the correct multiple-choice answer and another one for the correct explanation of the open-ended question for each of the ten items.

The calibrated instrument was administered to 754 pupils and 99 teachers from schools in the NW England. For purposes of comparison, the same analysis (i.e. the Rasch analysis described above) was run for the pupils' and the teachers' datasets.

RESULTS FOR THE TEACHERS' SAMPLE

The results of the Partial Credit analysis for the teachers' sample indicated that the data-model fit was appropriate. For example, Item 6 (Random Similarity Effect) had the largest Infit MNSQR (1.16) which is considered to be appropriate for all practical intents and purposes of this study. The item reliability index was 0.95 with a separation index of 4.57. Less than 5% of the respondents had fit statistics indicating poor model-data fit and this is also acceptable for empirical data. The average ability for the teachers was 0.46 (SD=1.01). The ability measures ranged from -3.12 to 2.45 logits. The average raw score was 8.8 (out of 20 maximum possible marks) with a SD of 4.1 but this is difficult to interpret because of the missing responses.

Figure 1 illustrates the ability distribution of the teachers and the difficulty of the items broken down by sub-item (e.g. 3.1 denotes the multiple choice part of item 3 and 3.2 indicates the 'Explain why' part of the same item). According to Figure 1, the test and sample can be interpreted as falling into a hierarchy of three levels. At level 1, approximately -3.0 to -0.5 logits, teachers can succeed on answering correctly questions that tested for the recency effect items (Q1, Q2 and Q3) and also the multiple-choice parts of two Random Similarity Effect items (Q4.1 and Q5.1). At level 2 (approximately from -0.5 to 1 logits), teachers attain higher performance and they can explain their answers to the Random Similarity question 4.2 and also answer correctly the Base Rate Effect questions (Q7 and Q8). Fewer teachers manage to attain level 3 by answering the hardest Random Similarity questions (Q5.2 and Q6) and the Sample Size effect questions (Q9 and Q10).

Overall, the inexperienced teachers were statistically significantly more able than the more experienced teachers in the sense that they had larger average Rasch measures. The largest difference was between the secondary inexperienced and primary experienced teachers. The secondary inexperienced teachers were, on average, at the borderline between Level 2 and Level 3. However, the primary experienced teachers were on the borderline between Level 1 and Level 2.

By averaging the ability estimates of those teachers who made an error, we are able to plot errors on the same logit scale in the table. No teachers gave responses to the multiple-choice parts of questions 1-6 (Recency and Random Similarity effects). Teachers who gave responses indicating the Base Rate (questions Q7 and Q8) misconceptions had a rather low ability. Answers indicating misconceptions based on the Sample Size effect (questions Q9 and Q10) were given by a more able group of teachers.

The teachers were not very successful in describing the most common errors and misconceptions that their pupils were likely to make (this refers to the third part of the items which asked the teachers to predict the common errors and misconceptions of the pupils on each question). Just above 50% of the teachers mentioned that they expected the answers of their pupils on questions Q1, Q2 and Q3 to be influenced by the negative recency effect (62.2% for Q1, 51.6% for Q2, 58.5% for Q3). Around 85% of the respondents expected their pupils' responses to questions Q7 and Q8 to be influenced by the Base Rate effect (83.8% in Q7 and 86.7% in Q8). Very few respondents, however, acknowledged that their pupils' thinking would be influenced by the Random Similarity effect on questions Q4 (12.1%), Q5 (9.8%) and Q6 (0%). The percentages for the Sample Size effect were a bit larger (18.9% for Q9 and 29.6% for Q10).

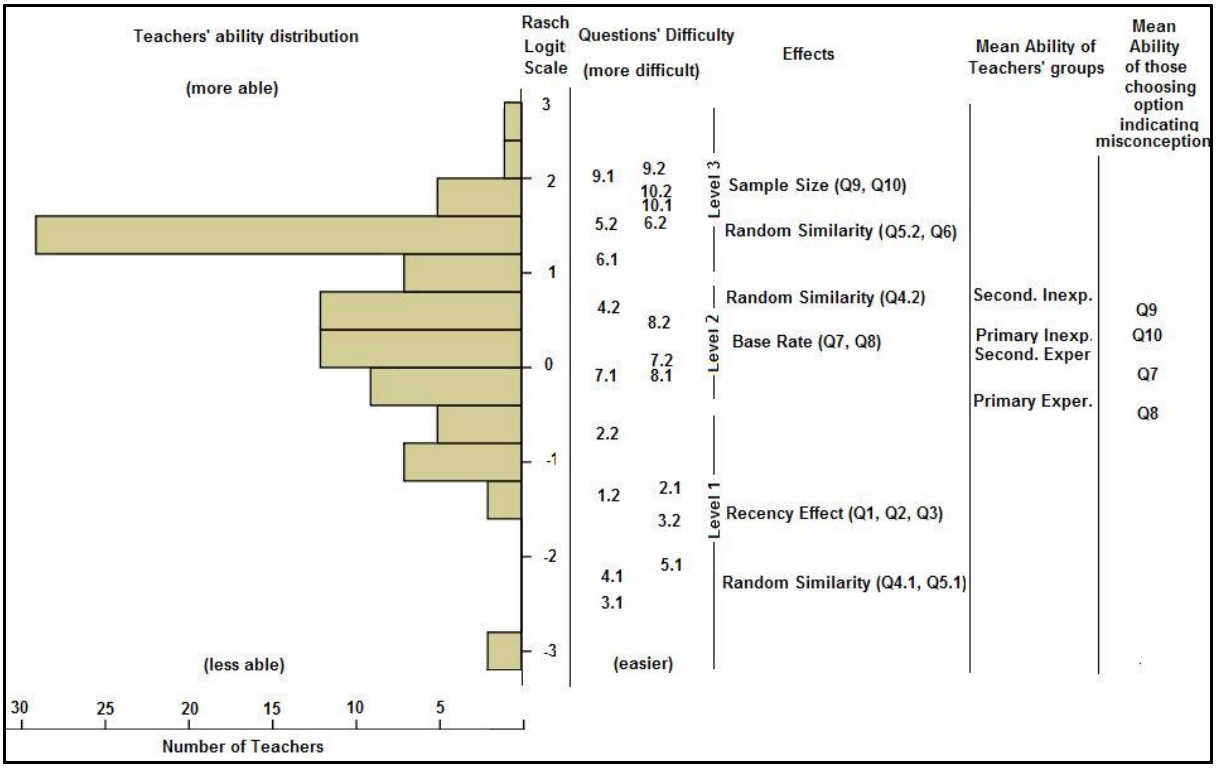

Figure 1: Teachers' ability distribution and item difficulty on the same Rasch scale

When a teacher predicted successfully the common errors and misconceptions of the pupils on a question, he/she was awarded 1 mark. For example, if a teacher predicted successfully the common errors and misconceptions of the pupils on all questions, he/she would receive 10 marks in total (one for each item). However, we could not use the raw score of the teachers across all items as an indicator of their knowledge of pupils' misconceptions because of the large percentage of missing cases. Therefore, we used the Rasch model to convert the raw score of the teachers to a linear scale bypassing the problem of the missing cases. It was found that the 68 inexperienced teachers had an average of 'predictive ability' (to predict the misconceptions of their pupils) of -0.93 logits (SD=1.28). The 31 experienced teachers had an average 'predictive ability' of -0.18 logits (SD=1.40). A t-test showed that the difference was statistically significant (t=-2.643, df=97, p=0.010) and that the experienced teachers were significantly more able to predict the common errors/misconceptions of the pupils.

RESULTS FOR THE PUPILS' SAMPLE

The results of the Partial Credit analysis for the pupils' sample indicated that the data-model fit was appropriate. The fit of Item 6 (Random Similarity Effect) had the largest Infit MNSQR (1.26) which is considered to be appropriate for all practical intents and purposes of this study. All other items had even better Infit MNSQR statistics (between 0.75 and 1.08). The item reliability index was 0.99 with a separation index of 21.65 which is an indication of a very reliable separation of the item difficulties. Just above 5% of the respondents had fit statistics indicating poor model-data fit and this is also acceptable for empirical data. The average ability for the pupils was -0.83 logits (SD=1.12). The ability measures ranged from -3.93 to 3.64 logits. The average raw score was 7.5 (out of 20 maximum possible marks) with a SD of 2.6 but this is difficult to interpret because of the large number of missing responses or not administered items.

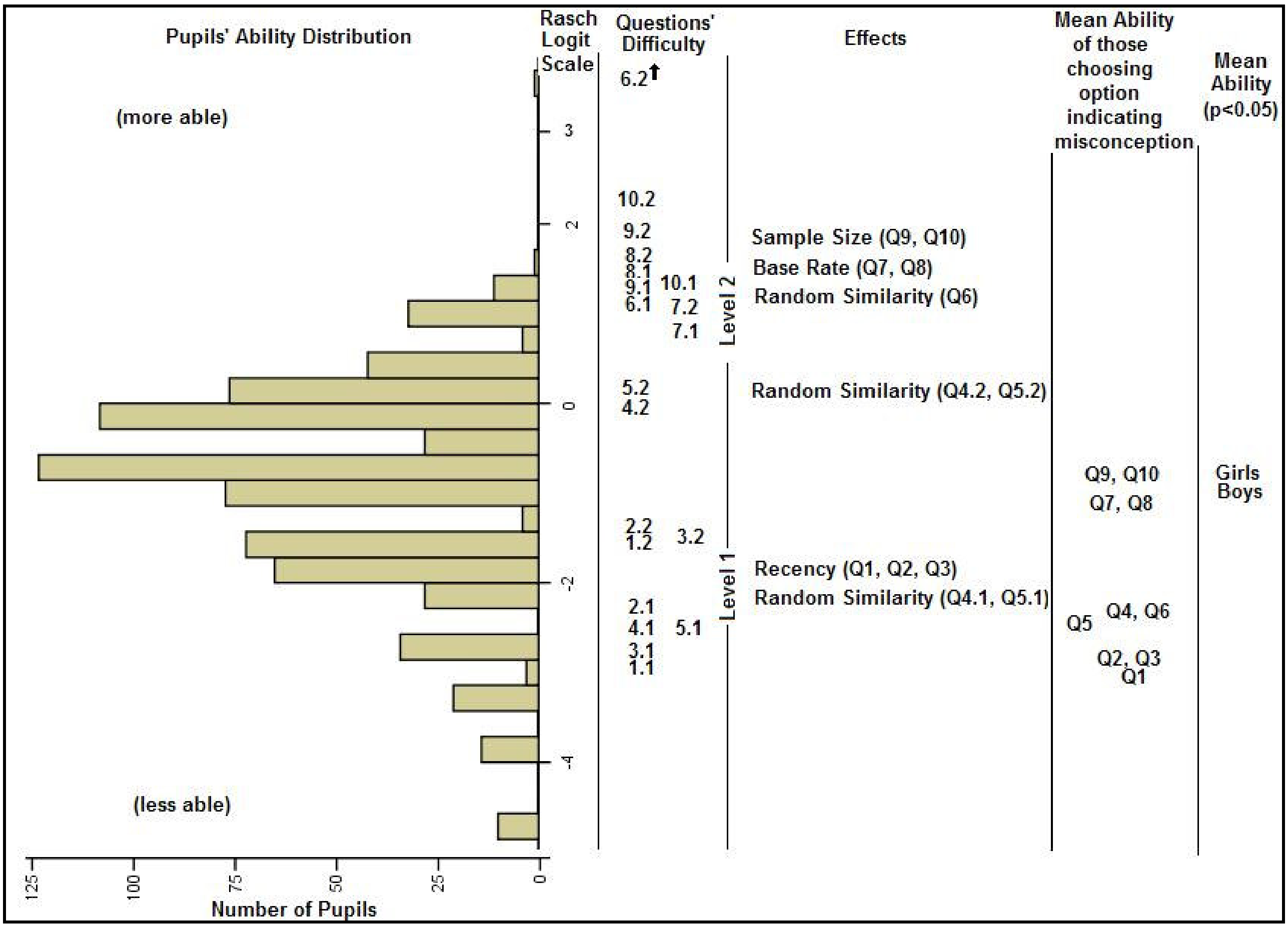

Figure 2: Pupils' ability distribution and item difficulty on the same Rasch scale

According to Figure 2, the test and sample can be interpreted as falling into a hierarchy of two levels. At level 1, approximately -4.0 to -0.5 logits, pupils can succeed on answering correctly questions that tested for the recency effect items (Q1, Q2 and Q3) and also the multiple-choice parts of two Random Similarity Effect items (Q4.1 and Q5.1). At level 2 (approximately from -0.5 to 4 logits), pupils attain higher performance and they can answer the multiple choice of question Q6, explain their answers to the Random Similarity question Q4.2 and Q5.2 and they can also answer correctly the Base Rate Effect questions (Q7 and Q8). Fewer pupils manage to attain the top of level 2 by answering the hardest Sample Size effect questions (Q9 and Q10). Almost nobody managed to give a correct response to question Q6.2.

By averaging the ability estimates of those pupils who made an error, we are able to plot errors on the same logit scale in the figure. Most of the pupils gave responses to the multiple-choice parts of questions 1to 6 (Recency and Random Similarity effects) which indicated that their probabilistic thinking was affected by the representativeness heuristic. The average ability of those pupils for items 1 to 6 was around -2.5 logits (Q1:-3.07 to Q6:-2.36 logits) which is well below the mean ability of the whole sample (-0.83 logits). However, the pupils who gave responses indicating the Base Rate (questions Q7 and Q8) and the Sample Size (questions Q9 and Q10) misconceptions had a mean ability in the area of -1 logit (Q7:-1.10 to Q9:-0.84 logits) which is near the mean ability of the sample.

CONCLUSIONS AND DISCUSSION

Having collected a fresh dataset of responses of pupils and teachers to the instrument which we developed in a previous study (Afantiti Lamprianou and Williams, 2002, 2003), we used Rasch analysis to investigate (a) the degree to which the probabilistic thinking of pupils and teachers suffers from the representativeness heuristic, (b) whether the item hierarchy resulting from the Rasch analysis for pupils and teachers would be similar, and (c) whether the teachers were aware of the common pupils' errors and misconceptions on the items of the instrument.

The analysis of the pupils' data showed that there is a hierarchy of two levels to characterise their probabilistic thinking and this is in agreement with Afantiti Lamprianou and Williams (2002, 2003). Indeed the item hierarchy was found to be the same as the one found by Afantiti Lamprianou and Williams, although the samples were from different schools and were collected two years apart. Pupils' probabilistic thinking was found to be affected by the representativeness heuristic to a great extent in the sense that few pupils managed to reach level 2 (to answer correctly the Base Rate and the Sample Size items). The pupils found the 'Explain why' parts of the Base Rate and the Sample Size items extremely difficult and very few succeeded in answering these correctly.

The analysis of the Teachers' responses showed that the probabilistic thinking of a large number of respondents is influenced by the representativeness heuristic. Few teachers were in a position to answer correctly the most difficult items testing the Sample Size effect (Q9 and Q10).

The item hierarchy resulting from the Rasch analysis of the Teachers' and Pupils' data is not the same. This may be seen by comparing Figures 1 and 2. The rank order of item difficulties does not remain the same when the two figures are compared (although, in absolute numbers, the differences are almost always within the 95% error of measurement). The two hierarchies seem to be qualitatively different in the sense that the Base Rate items were found by the teachers to be substantially easier in comparison to the Sample Size and the Random Similarity items.

One of the most striking findings, however, was the fact that the more experienced teachers were found to have a significantly poorer performance on the instrument compared to the younger and less experienced teachers. One possible explanation could be that the younger and less experienced teachers had the opportunity to receive preparation and training on probabilities and statistics because these topics became more widely available in the relevant teacher training courses in Universities. This finding is in line with the suggestion of Godino, Canizares and Diaz (n.d.) who suggested the need to increase the training opportunities for serving teachers on issues like statistics and probabilities (Godino, Batanero and Roa, 1994; Godino, Canizares and Diaz, n.d.).

This is notably in contrast to the other main result, i.e. that the experienced teachers' pedagogical knowledge was superior (i.e. that the more experienced teachers were in better position to predict the common errors and misconceptions of the pupils): this is in the direction expected, and suggests that the methodology adopted affords the making of nice distinctions between teachers' subject-content and pedagogical-content knowledge. This result reinforces the pilot work in this regard of Hadjidemtetriou and Williams (2004).

References

Afantiti-Lamprianou, T., & Williams, J. (2002). A developmental scale for assessing probabilistic thinking and the tendency to use a representativeness heuristic. In A. D. Cockburn, & E. Nardi (Eds.), *Proc. 26th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 9-16). Norwich, UK: PME.

Afantiti-Lamprianou, T., & Williams, J. (2003). A Scale for Assessing Probabilistic Thinking and the Representativeness Tendency. *Research in Mathematics Education*, 5, 173 - 196.

Amir, G., Linchevski, L., & Shefet, M. (1999). The probabilistic thinking of 11-12 year old children. In O. Zaslavsky (Ed.), *Proc. 23rd Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 25 – 32). Haifa, Israel: *PME.*

Batanero, C., Serrano, L., & Garfield, J. B. (1996). Heuristics and Biases in secondary school students' reasoning about probability. In L. Puig, & A. Gutierrez (Eds.), *Proc. 20th Conf. of the Int. Group for the Psychology of Mathematics Education*. Valencia: *PME.*

Fischbein, E., & Schnarch, D. (1997). The Evolution with age of probabilistic, intuitively based misconceptions. *Journal of Research in Mathematics Education,* 28(1), 96-105.

Godino, J. D., Batanero, C., & Roa, R. (1994). Training Teachers to teach Probability. *Third International Conference on Teaching Statistics.* Retrieved January 2005 from www.stat.auckland.ac.nz.

Godino, J. D., Canizares, M. J., & Diaz, C. (n.d.). *Teaching probability to pre-service primary teachers through simulation.* Retrieved January 2005 from www.stat.auckland.ac.nz.

Green, D. R. (1982). *Probability Concepts in 11-16 Year Old Pupils. Report of Research, CAMET.* Loughborough, England: University of Technology.

Hadjidemetriou, C., & Williams, J. (2002). Teachers' Pedagogical Content Knowledge: Graphs, from a Cognitivist to a Situated Perspective. In A. D. Cockburn, & E. Nardi (Eds.), *Proc. 26th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 3, pp. 57-64). Norwich, UK: *PME.*

Hadjidemetriou, C., & Williams, J. (2004). Using Rasch Models to Reveal Contours of Teachers' Knowledge. *Journal of Applied Measurement,* 5(3), 243 - 257.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in Students' reasoning about probability. *Journal for Research in Mathematics Education*, 24(5), 392-414.

Norman, F. A. (1993). Integrating Research on Teacher's Knowledge of Functions and their Graphs. In Romberg, A. T., Fenemma, E., & Carpender, P. T. (Eds.). *Integrating Research on the Graphical Representation of Functions,* Hillsdale, NJ: Erlbaum.

Shaughnessy, J. M. (1992). Research in Probability and Statistics: Reflections and Directions. In D.A. Grouws (Ed.). *Handbook of Research on Mathematics Teaching and Learning*. Reston: NCTM.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform, Harvard. *Educational Review*, 57(1), 1-22.

Turner-Bisset, R. (1999). The Knowledge Bases of the Expert Teacher. *British Educational Research Journal,* 25(1), 39–55.

Williams, J. S., & Ryan, J. T. (2000). National Testing and the improvement of Classroom Teaching: can they coexist? *British Educational Research Journal*, 26(1), 49-73.

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.