Goodchild, S. (Ed.) Proceedings of the British Society for Research into Learning Mathematics 22(1&2) Feb/May 2002

A SCALE FOR ASSESSING PROBABILISTIC THINKING AND THE REPRESENTATIVENESS TENDENCY

Thekla Afantiti-Lamprianou and Julian Williams

University of Manchester

We report a study of children's probability conceptions and misconceptions using a diagnostic instrument developed from the literature on the representativeness heuristic. Rasch measurement methodology was used to develop the 13-item open response instrument with a sample (N=116) of 12-15 year olds. The result is that a hierarchy of responses at two levels is confirmed for this sample, and a third level is hypothesised. Each level is characterised by the ability to overcome typical 'representativeness' effects, namely 'recency', 'random-similarity' (at level 1), 'base-rate frequency' and 'sample size' (at level 2-3). The validity of our interpretations was tested and some anomalies were identified through clinical interviews with children making the errors (n=8). Another Rasch ability measure, which we named the 'representativeness tendency', is constructed from 11 multiple-choice errors.

INTRODUCTION

This study builds on previous work on children's understandings, intuitions, use of heuristics and misconceptions in their probabilistic thinking (Fischbein, 1975, 1997; Kapadia and Borovcnik, 1991; Shaughnessy, 1992) and especially the significance of the representativeness heuristic (Green, 1982; Kahneman, Slovic and Tversky, 1982; Amir and Williams, 1999; Amir, Linchevski and Shefet, 1999). The misconceptions based on the representativeness heuristic are some of the most common errors in probability: children tend to estimate the likelihood of an event by taking into account how well it represents its parent population and how it appears to have been generated.

In this study (reported also at greater length in Afantiti-Lamprianou and Williams, 2002, under review) we aim to contribute to teaching by developing an assessment tool which can help teachers diagnose inappropriate use of the representativeness heuristic in responses to questions relevant to the probability curriculum. Thirteen items were used to construct the instrument (the instrument can be seen in full on the web at http://www.education.man.ac.uk/lta/tal). 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 those used in previous research (Green, 1982; Kahneman et al., 1982; Batanero, Serrano and Garfield, 1996; Fischbein and Schnarch, 1997; Amir et al., 1999). Other items were developed based on findings of previous research.

Items called *recency 1, 2 and 3* tested for the negative recency effect or the gambler's fallacy. According to this effect, a long sequence of one outcome must be followed by the other outcome in order to equilibrate the proportions. Items called *random-similarity 4, 5, 6 and 8* tested for the effect which expects a sample to appear similar

© BSRLM and others

in proportion to the parent population and apparently randomly-generated.

Items called *base-rate 10, 11 and 12* were written to examine the effect of prior probability or the base-rate frequency of the outcomes in contexts appropriate to this age group. According to this effect, prior probabilities are effectively ignored when misleading irrelevant but stereotypical information is introduced. Finally, items called *sample size 7, 9, 13* tested for the tendency to neglect sample size in estimating probability. These items examined the belief that the probability of a certain proportion in a sample is independent of the sample size, contradicting the central limit theorem, i.e. the probability of getting a certain empirical result tends to approach the theoretical prediction as the sample gets larger.

METHOD

In order to be able to administer more items to the same sample of pupils two separate test-forms with common linking items were constructed. Test A, designed to be easier, consisted of eight items - items 1 to 6, 9, 10 - and Test B, intended to be more difficult, consisted of ten items - items 3, 4, 6 to 13. Five of the items were included in both tests. Test A was administered to pupils in Year 7 and Test B was administered to pupils in Year 8 and 9.

The tests were administered to 116 pupils from schools in the North West of the United Kingdom. Before administering the tests to the pupils, their teachers were asked to read and comment on the suitability of the tests for their classes. They found the wording of the items acceptable for the pupils' age, but they commented on the degree of difficulty of question 13 (*sample size 13*).

Since all items had both a multiple-choice and an open-ended question, a common item Partial Credit analysis (Wright and Matsers, 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 13 items. The result is a single scale consisting of a 'difficulty' estimate for each scored point and an 'ability' estimate for each child consistent with the Rasch measurement assumptions. Item 13 fell outside a model infit statistic value of 1.3 (see Wright & Stone, 1979) reflecting the difficulty of this item for the sample.

In addition to the test analyses, eight structured clinical interviews were conducted using the test items to gain insight into the cause of the effects of the representativeness heuristic, to confirm the literature, validate the items and identify anomalies.

RESULTS

The result is a scale indicating a hierarchy of responses. The test and sample can be interpreted as falling into a hierarchy of three levels (see table below). At level 1 (-3.0 to -0.5 logits), children can succeed on questions that tested for the *recency* effect and easy questions that examined the *base-rate frequency* and the *random-similarity*

© BSRLM and others

effect. At level 2 (-0.5 to 2 logits), children attain higher performance and they can explain their answers to the easier questions that tested for the *random-similarity* effect, they can manage harder *base-rate* and *random-similarity* questions and they are beginning to answer some *sample size* questions correctly. Very few children manage to attain level 3 by answering the hardest questions on *sample size* or explaining their answers to the harder questions that tested for the *random-similarity* effect. In order to establish level 3, it is suggested that a more able sample would be required.

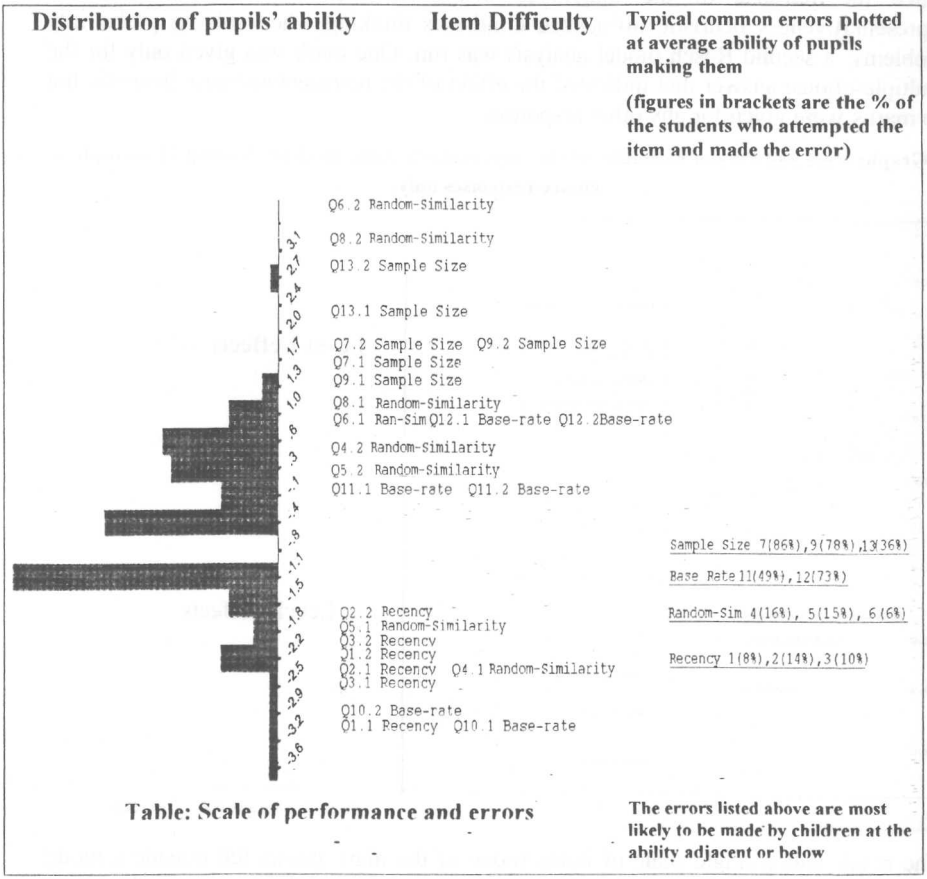


Table: Scale of performance and errors

The errors listed above are most likely to be made by children at the ability adjacent or below

By averaging the ability estimates of those children who made an error, we are able to plot errors on the same logit scale in the table. Pupils who gave responses indicating the *recency* misconceptions had a rather low *ability*. Answers indicating misconceptions based on the *random-similarity* effect were given by a broader range

© BSRLM and others

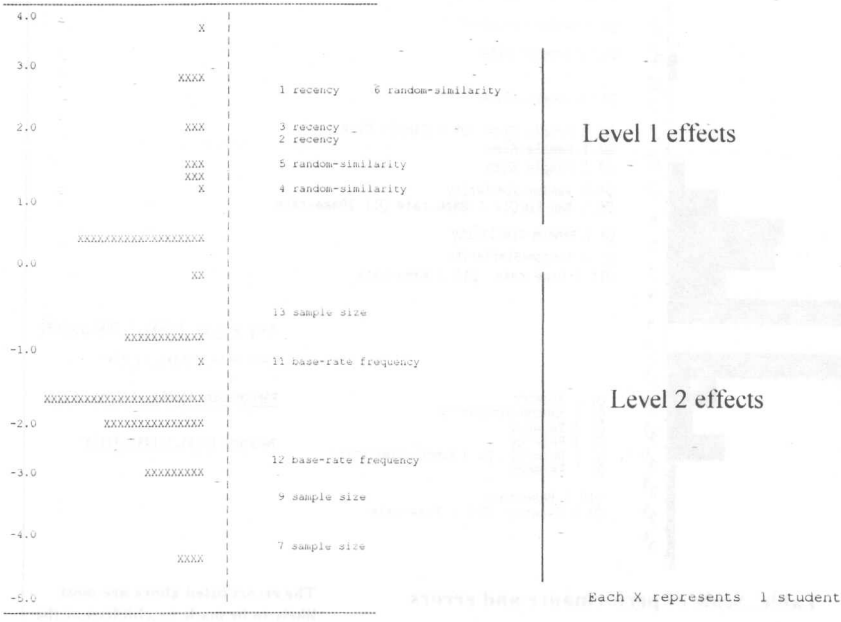
of *ability* pupils (averages ranged from -1.95 to -1.25 logits). On the other hand, the mean *ability* of the pupils who gave responses based on the *sample size* and *base- rate* effect was near the average *ability* of the sample, reflecting the fact that these errors were made by so many children (36%, 49%, 73%, 78% and 86%).

This encouraged us to consider building a 'representativeness tendency' measure from those errors which we can authentically attribute to this heuristic, as a diagnostic measure of tendency to inappropriately apply this heuristic.

CONSTRUCTION OF A REPRESENTATIVENESS MEASURE

Since the purpose of the diagnostic instrument was to assess whether the representativeness heuristic influenced children's thinking when solving probability problems, a second Rasch model analysis was run. One mark was given only for the multiple-choice answer that indicated the effect of the representativeness heuristic but no marks were given for any other responses.

Graph: Item and Person Estimates of the 'representativeness tendency', using 11 multiple-choice responses only.



The result was a single scale of items (none of the mark points fell outside a model infit value of 1.3), indicating that the *sample size* effect and the *base-rate* effect were very frequent among the pupils. The *random-similarity* and the *recency* effect influenced a small number of pupils of this sample (see graph above). *Random-*

© BSRLM and others

similarity 8 and *base-rate 10* were removed from the Rasch analysis because all pupils gave different responses to the expected representativeness effects.

Children higher up the scale are more likely to make representative-effect related errors, and items higher up the scale are less commonly occurring, i.e. only made by those with a strong 'representativeness tendency'. Note that these fall into the two levels of questions identified previously in the table, with *recency* and *random-similarity* effects generally occurring at level 1 and *base-rate frequency* and *sample size* effects at level 2.

PUPILS' INTERVIEWS

The main purpose of the pupils' interviews was to validate the items of the test, in particular our interpretation that the errors in the test are symptomatic of the representativeness effects discussed in the literature. From the interviews, some items proved problematic. In *random-similarity* 6 and 8, the majority of the pupils chose a different incorrect answer to the expected *random-similarity*. This was based on the conceptual error which reflected children's failure to discriminate between sequences and classes of sequences (Amir et al., 1999). We therefore came to the conclusion that this misconception was distracting from the *random-similarity* effect, which was only given by 6% of pupils in *random-similarity* 6 and fell to 0% for *random-similarity* 8. Worse still, children chose the correct option but for the wrong reason, arguing from a representativeness perspective, that '6 Heads and 6 Tails' would be most likely, because:

> Child 33: Because, em, there's two sides on the coin and you get, em... even chances of getting six heads and six tails, because if you divide by two... added to two sides of the coin there's six on each, like heads and tails.

This also explains why so few children could give a correct explanation for their answers to these items. We suggest these items may need therefore to be deleted, redesigned or developed and re-scored.

CONCLUSIONS AND DISCUSSION

We have managed to develop two scales measuring children's responses to the instrument which is especially revealing about their inappropriate use of the representativeness heuristic in responding to test questions which are relevant to their curriculum. We have further identified some previously unknown interpretations of children's responses. What is more, the development, validation and calibration of the measures around this heuristic for 12-15 year old children are new.

Having collected responses of some teachers to the instrument, we are doubtful that teachers are aware of these common misconceptions or of the significance of the representativeness heuristic, and we suggest that many teachers might benefit from using such an instrument in their assessment and teaching. The knowledge that teachers would collect from these scales, might enrich teachers' mental models of their

© BSRLM and others

learners and help them improve their classroom practice (as discussed in Williams and Ryan, 2001). We will be studying this aspect in the next stage of the work.

REFERENCES

Afantiti-Lamprianou, T. and Williams, J.: 2002, 'A scale for assessing probabilistic thinking and the representativeness tendency'. In *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education (PME)*, (under review). England: Norwich.

Amir, G., Linchevski, L. and Shefet, M.: 1999, 'The probabilistic thinking of 11-12 year old children.' In O. Zaslavsky (ed.), *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education (PME)* (pp. 25-32). Haifa: Israel Institute of Technology.

Amir, G. and Williams, J.: 1999, 'Cultural Influences on Children's Probabilistic Thinking.' *Journal of Mathematical Behaviour*, 18(1), 85-107.

Batanero, C., Serrano, L. and Garfield, J.: 1996, 'Heuristics and Biases in secondary school students' reasoning about probability.' In L. Puig and A. Gutierrez (eds.), *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education (PME)* (pp. 51-58). Valencia: Universitat de Valencia.

Fischbein, E.: 1975, *The Intuitive Sources of Probabilistic Thinking in Children*. Dordrecht: Reidel.

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

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

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

Kapadia, R. and Borovcnik, M. (eds.): 1991, *Chance Encounters: Probability in Education*. Dordrecht: Kluwer.

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

Williams, J. and Ryan, J.: 2001, 'Charting argumentation space in conceptual locales.' In M. van den Heuvel-Panhuizen (ed.), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education (PME)* (pp. 423-430). Utrecht University.

Wright, B. and Masters, G.: 1982, *Rating Scale Analysis*. Chicago: MESA Press.

Wright, B. and Stone, M.: 1979, *Best test design*. Chicago: MESA Press.

© BSRLM and others