This is the extended abstract of a symposium presented at the 2010 annual conference of the Australian Association for Research in Education (AARE). It was previously published at http://ocs.sfu.ca/aare/index.php/AARE/AARE_2010/paper/view/2258

Evaluating automated processes for revealing students' mathematical thinking

Symposium Extended Abstract

Kaye Stacey, Vicki Steinle, Margaret Wu, Robyn Pierce ***University of Melbourne***
Jason Giri ***University of Ballarat***

Understanding student thinking is key to planning effective teaching. Research over three decades has established good knowledge of stages of development and common misunderstandings in many school topics. However, in order for teachers to make formative assessment a normal part of their practice, mapping of student understanding needs to be available for a greater proportion of the school curriculum. Online testing of students is beginning to create substantial databases which could provide relevant information. The aim of this symposium is to present and discuss methods which move towards automation of data analysis.

The three symposium papers will evaluate different methods of analysing data on students' understanding of mathematics. SMART tests ("Specific Mathematics Assessments that Reveal Thinking" www.smartvic.com) are online diagnostic tests with feedback related to student thinking and teaching suggestions made available to the teacher. A very extensive database of student responses is emerging from the SMART test development. This presents the challenge of needing to find more efficient methods of identifying new clusters of responses and combinations of answers that will give insights into students' incorrect, incomplete and unexpected mathematical conceptions, as well as the stages of development which are typical in specific topic areas. This complex data set is therefore a useful test bed for data mining methods in educational research.
The characteristics and fitness for purpose of three methods will be discussed in this symposium:

- Item Response Theory, including Rasch modelling (a probabilistic model)
- Response Pattern Analysis (a deterministic model)
- Data mining using a modified Global k-means algorithm for a large dataset and selection of optimal parameters for clustering.

A systems approach has been taken to this investigation. The criteria for evaluating the methods include: ease of data handling; run time; usability of output; identification of known clusters; efficacy with numeric and categorical data; efficacy with increasing numbers of dimensions; and provision of results which give insight into student thinking. Subsets of the SMART test data were analysed using each of the methods listed above. The resulting output has been studied by the research team and a description and evaluation of each method will be presented. This will be followed by discussion of the respective contributions each method can make to improving test design and to revealing patterns of student thinking.

Contribution 1

Using Item Response Modelling for Revealing Students' Mathematical Thinking

Margaret Wu, University of Melbourne

This presentation demonstrates how student test data can be analysed using item response modelling to identify misconceptions and common errors in mathematical thinking. The data sets are from the online SMART (Specific Mathematics Assessments that Reveal Thinking) tests.

Item response theory (IRT) has been widely used for the construction of proficiency scales in educational assessment. In Australia, IRT has been the key methodology underpinning the construction of the literacy and numeracy scales in NAPLAN (National Assessment Program – Literacy and Numeracy). However, the application of IRT for diagnostic purposes has not been as widely used as for measurement purposes. One possible reason for this is that IRT typically models the chance that a student responds correctly or incorrectly to a question, rather than focuses on different kinds of incorrect responses.

This paper uses IRT to examine student performance by response categories of each question rather than by a simple dichotomy of correct/incorrect responses. The analyses provide a much richer set of diagnostic information in terms of identifying misconceptions. Plots of the proportion of students in each response category as a function of ability level show the relative frequencies of different errors, as well as how the error rates vary between low and high ability students. For some questions, it can be seen that, even when mathematics ability increases, the misconceptions remain for a large proportion of high achievers. For other questions, the error rates drop sharply as the overall mathematics ability increases. These analyses provide information beyond just the difficulty of a question. The analyses inform the teachers of the kind of problems that remain as constant sources of confusion, in contrast to the problems that are typical stepping-stones along the developmental continuum.

An IRT approach for providing diagnostic information works well when each question is designed to test specific skills and misconceptions. Profiles of students' mathematics proficiency can be constructed based on groups of items tapping into similar skills and/or misconceptions, provided a skills audit of test items has been carried out a priori. That is, IRT provides confirmatory analyses rather than exploratory analyses of patterns of item responses. Combining IRT analyses with the other two methodologies presented in this symposium, a comprehensive set of tools can be developed for providing useful feedback to teachers and students.

Contribution 2

Exploring student thinking using response pattern analysis

Vicki Steinle & Kaye Stacey. University of Melbourne

Patterns of responses to carefully designed items have been shown before to reveal students' thinking. Steinle and Stacey (2003) showed that a specially designed Decimal Comparison Test provided data that was amenable to response pattern analysis. Large groups of students give the same patterns of responses to items, which can be explained by their holding particular misunderstandings regarding place value and the size of decimal numbers. Stacey and Steinle (2006) demonstrated that this data is not amenable to classic approaches such as Rasch analysis, pointing out that conceptual learning in this case was better mapped than measured on a scale. In particular, a fine-grained analysis of student thinking and learning showed that students' thinking over time can move between conceptions. The different conceptions do not necessarily lead to more correct answers and a higher score, even when they are based on an improving understanding of fundamental principles. Response pattern analysis has subsequently been used to confirm and explore students' misconceptions related to interpretations of letters in algebra (Steinle et al, 2009), and other examples will be given in this paper.

Response pattern analysis is a fine-grained, student-focussed approach, aiming to identify how students think about a topic. It contrasts with approaches which aim to measure how much a student knows. In this paper, we will outline some important differences in this orientation, and in the symposium we also show how it can be used in conjunction with other techniques to identify common conceptions. Data and analysis from a large bank of student responses to online mathematics items in the SMART tests (www.smartvic.com) will be shown.

Due to the success of the response pattern analysis to provide insights into student thinking, a pattern recognition script was written to automate this process. At its simplest, the pattern recognition script identifies students with the same set of data on a defined set of items within a spreadsheet. The data can be accuracy data (correct, incorrect or blank) or response data (e.g. free numerical response, free algebraic response, the selection for a multiple-choice item, or position coordinates of "drag & drop" responses in one or two dimensions). When using this script, the user can set a tolerance which allows the user to identify groups of similar response patterns, rather than only identifying identical response patterns.

A challenge with this technique is that the user needs to define which items are to be included, and the possibilities for responses multiply as the number of items increases. Consequently, this technique requires the user to be a 'content expert' with prior reasons for selecting the combinations of items and the form of the response to investigate. For example, accuracy data, having only three options (correct, incorrect or blank) provides a coarser analysis than response data. Once groups of students with the same/similar response patterns are identified, then the next step is to determine whether the learning is best described as a gradual movement up a hierarchy towards expertise or a movement between different types of (correct and/or incorrect) thinking. Behind all of this investigation is the fact that there is a duality between questions asked and understanding revealed. It is only by asking the right questions that we uncover how students think, and only by knowing how students think can we know which questions are worth asking.

Contribution 3

Looking for clusters in n-dimensions: mining mathematics education data

Jason Giri, *University of Ballarat* and **Robyn Pierce**, *University of Melbourne*

Online trials of “Specific Mathematics Assessments that Reveal Thinking” (SMART tests) have resulted in an extensive database of results. In other fields as diverse as aircraft engineering and cancer research, extensive databases are mined in search of associations that have not previously been identified. Such methods have not yet been applied to mathematics education data. This paper reports on the findings, to date, of applying a modified Global k-means algorithm (Bagirov, 2008) to data collected from the SMART test trials.

Based on previous research on students’ understanding of the various topics across the junior secondary curriculum, SMART tests are designed to diagnose student thinking and provide feedback for teachers that will assist them to personalise the teaching of each student. The volume of data resulting from this process presents not only the challenge of needing to find more efficient methods of automating the diagnostic process but also an opportunity to identify new clusters of responses and combinations of answers that will give insights into students’ incorrect, incomplete and unexpected mathematical conceptions.

In the past identifying patterns in students’ test responses and following up by interviewing the appropriate students has proven successful for identifying incorrect student thinking. A key example of this work has been Steinle and Stacey’s investigation of students’ understanding of decimal numbers (Stacey, 2005, Steinle, 2004). Incomplete and incorrect understandings are identified by looking at the combinations of students’ responses, not just single items. Preliminary exploration of SMART test data suggests students’ numerical misconceptions in algebra may be revealed by examination of combinations of responses from two test items (Steinle et al 2009). However, when using traditional methods, working with more than two or three items becomes too cumbersome.

The modified global k-means algorithm may be simultaneously applied to the results of a large number of items. Working with n-dimensional vectors, this clustering technique uses non-smooth optimization to assign data points to clusters. While based on the minimum ‘distance’ between points in n-dimensional space these clusters are irregular in shape. Following the success of the application of the Global k-means algorithm to other databases it seems likely that it may be able to efficiently identify clusters (of students) already expected on the basis of previously identified patterns and perhaps reveal previously unrecognized patterns of student thinking.

To date the method has been applied to the numeric results in the database with some success. Clustering that can be interpreted on the basis of previously recognized patterns have been identified and items worth further investigation noted. Missing data and the typical mixture of categorical and numerical responses generated by mathematics testing present some challenges for application of the algorithm. A plan has been proposed for incorporating coded categorical responses into the vectors by first transforming the assigned codes to ensure normed interval spacing. This type of analysis has not previously been undertaken.

Further research using these new methods is warranted for exploring data to confirm previous hypotheses related to patterns of student thinking, suggest new hypotheses worth further investigation, and offer more efficient diagnostic techniques than those currently in use.

References

Bagirov, A.M. (2008). Modified global k-means algorithm for minimum sum-of-squares clustering problems, *Pattern Recognition* (41), No. 10, October 2008, pp. 3192-3199. https://doi.org/10.1016/j.patcog.2008.04.004

Stacey, K. (2005) Travelling the road to expertise: A longitudinal study of learning. In. H. Chick & J. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 19-36). University of Melbourne: PME. http://www.igpme.org/wp-content/uploads/2019/05/PME29-2005-Melbourne.zip

Stacey, K., & Steinle, V. (2006). A case of the inapplicability of the Rasch Model to mapping conceptual learning. *Mathematics Education Research Journal*, 18(2), 77-92. https://link.springer.com/article/10.1007/BF03217437

Steinle, V. (2004). *Changes with Age in Students' Misconceptions of Decimal Numbers.* Doctor of Philosophy Dissertation, University of Melbourne. https://minerva-access.unimelb.edu.au/handle/11343/39024

Steinle, V., & Stacey, K. (2003). Grade-related trends in the prevalence and persistence of decimal misconceptions. In N.A. Pateman, B.J. Dougherty & J. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education*, 4, 259 - 266. Honolulu: PME. http://www.igpme.org/wp-content/uploads/2019/05/PME27-2003-Honolulu.zip

Steinle, V., Gvozdenko, E., Price, B., Stacey, K., & Pierce, R. (2009) Investigating students' numerical misconceptions in algebra. In R. Hunter, B. Bicknell, T. Burgess (Eds). *Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia,* 2, 491 – 498. Wellington: MERGA. https://www.merga.net.au/Public/Public/Publications/Conference_Proceedings.aspx