4-2012

Some drivers of test item difficulty in mathematics

Ross Turner
ACER, Ross.Turner@acer.edu.au

Some Drivers of Test Item Difficulty in Mathematics

Ross Turner, Australian Council for Educational Research, ross.turner@acer.edu.au

Introduction

This paper is one of four contributions to the symposium session at the AERA's 2012 Annual Meeting titled *Exploring Reading and Mathematics Item Difficulty: Teaching and Learning Implications of PISA Survey Data.* The paper presents a rubric used to analyse mathematics test items developed for use in the OECD's PISA survey[1]. The rubric focuses on a set of mathematical competencies that are components of mathematical literacy. The work on which this report is based suggests that demand for activation of these competencies functions as a significant driver of item difficulty, which potentially has implications for the teaching and learning of mathematics.

Background

Over the last decade or more, education theorists, practitioners and systems in many parts of the world have recognised the importance of mathematics teaching and learning outcomes that are broader than knowledge of mathematical skills and procedures related to specific curricular content. A fuller appreciation of mathematical competence extends also to various processes that are increasingly seen as central to an individual's understanding of mathematical ideas and capacity to apply his or her mathematical knowledge. Evidence of this recognition can be seen in the formal curriculum statements of various educational jurisdictions around the world.

In the USA, the NCTM (see http://www.nctm.org/standards/) now has a set of five 'process standards' (in addition to its content standards) on problem solving, reasoning and proof, connections, communication, and representation. These are informing the curriculum statements and teaching and learning practices of many US states, and have also been a key influence on the recently developed 'common core standards' for the US.

In the UK, mathematical processes including representing, analysing, interpreting and evaluating, communicating and reflecting have been built centrally into the new national curriculum (see http://www.education.gov.uk/schools/teachingandlearning).

Curriculum documents placing emphasis on similar mathematical processes can also be found in Canada, and in Australia. But these approaches are by no means restricted to the English-speaking world. Similar developments can be observed in many Asian, European and South American countries. For example, the secondary

[1] The Programme for International Student Assessment (PISA) is an international comparative survey of the Organisation for Economic Cooperation and Development (OECD) targeting sampled 15-year-olds in the 60+ participating countries.

mathematics syllabuses of the Singapore Ministry of Education (2006) includes a set of mathematical processes (reasoning, communication and connections, thinking skills and heuristics, applications and modelling) as central elements of its mathematics framework (see http://www.moe.gov.sg/education/syllabuses/sciences/).

Similarly, the new national course of study in Japan for elementary schools and for junior high schools, released in 2008 and being progressively implemented, includes a new focus on the mathematical processes of thinking and reasoning mathematically, mathematical modelling and application, and representing and communicating mathematically (see http://www.mext.go.jp/English/elsec/1303755.htm).

The OECD's PISA mathematics framework is another place where this recognition of the importance of mathematical processes can be seen. The original PISA framework (OECD, 1999) includes a set of mathematical competencies that are fundamental to its definition of mathematical literacy. These were based on work done by Mogens Niss and his colleagues (Niss, 1999; Niss and Højgaard, 2011).

A sense of dissatisfaction and some urgency is emerging in the public discourse about the outcomes of school mathematics education as rich data become available from international surveys such as the OECD's Programme for International Student Assessment. There is a growing recognition of the need to improve mathematical outcomes that will help to ensure individuals are properly equipped to handle the increasingly complex quantitative and technological demands of the workplace and of society in general in the 21st century. We hope that the following discussion might contribute to a further enriching of understanding about the role of a set of mathematical processes, and to help promote consideration of the potential merits of focussing more directly on these processes in mathematics classrooms.

The PISA mathematical competencies

The mathematics component of the PISA survey aims to generate comparative estimates of average levels of mathematical literacy among 15-year-old students in participating countries. It defines mathematical literacy as

> *… an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens (OECD, in press).*

The definition emphasises using mathematics in context. The framework describes a set of seven 'fundamental mathematical capabilities' that underpin the processes needed by individuals to effectively make use of their mathematical knowledge and skills. These capabilities are derived from the mathematical competencies described by Niss and his Danish colleagues (Niss, 1999; Niss and Højgaard, 2011). A version of these competencies has been described in the frameworks for each of the PISA surveys (e.g. OECD 1999, p.43), and they have been used to build the proficiency descriptions through which the OECD reports PISA mathematics outcomes and levels of mathematical literacy (OECD 2004, p. 47). The framework for the 2012 PISA survey labels these capabilities as communication, mathematising, representation, reasoning and argument, devising strategies for solving problems, using symbolic, formal and technical language and operations, and using mathematical tools.

In an ongoing research investigation, experts involved in PISA implementation have looked closely at PISA mathematics survey questions, and have judged the extent to which successfully answering those questions demands activation of various mathematical competencies that reflect the PISA framework. For the purpose of that investigation, six competencies have been given operational definitions, and each of these competencies has been described at four levels. It is recognised that the six chosen competencies overlap to some extent, and that they frequently operate in concert and interact with each other; nevertheless the goal has been to treat each competency as distinctly as possible.

Competency definitions and their level descriptions

In this section, the six variables used in this research are discussed and defined, and four graduated levels of operation of each are described (labelled as level 0 – level 3). The definitions and descriptions are in development, so the following versions should be regarded as a draft form.

Communication

The communication competency on which this variable is based has both 'receptive' and 'constructive' components, since it includes understanding of the communications of others, and active communication by the student. The receptive component includes understanding what is being stated and shown in the stimulus including the mathematical language used, what information is relevant, and what is the nature of the answer requested. The constructive component includes presentation of the answer and in some cases solution steps, explanation of reasoning used or a justification of the answer provided.

In written and computer-based items, communication relates to text and images, still and moving. It does not apply to formal mathematical representations such as graphs and geometric diagrams, although the need to extract information from these representations in conjunction with text and images contributes to the communication demand. Text includes informally presented mathematical expressions. In computer-based items, the instructions about navigation and other issues related to the computer environment may contribute to the general communication demand, but usually not to the mathematical communication required.

Communication does not include knowing how to approach or solve the problem, how to make use of particular information provided, or how to reason about or justify the answer obtained; rather it is the transmission and interpretation of relevant information.

Demand for this competency increases according to the complexity of material to be interpreted in understanding the task, for example with the need to link multiple information sources or to move backwards and forwards (to cycle) between information elements; and with the need to provide a detailed written solution or explanation.

Definition: Reading, decoding and ***interpreting*** statements, questions, instructions, tasks and objects; ***imagining*** and understanding the situation presented and ***making sense*** of the information provided including the mathematical terms referred to; ***presenting and explaining*** one's mathematical work or reasoning.

Level 0: Understand a short sentence or phrase relating to concepts that give immediate access to the context, where all information is directly relevant to the task, and where the order of information matches the steps of thought required to understand the task. Constructive communication involves only presentation of a single word or numeric result.

Level 1: Identify, select and directly combine relevant elements of the information provided, for example by cycling once within the text or between the text and other related representation/s. Any constructive communication required is simple, and may involve writing a short statement or calculation, or expressing an interval or a range of values.

Level 2: Select and identify elements to be combined, and use repeated cycling to understand instructions, or decode and link multiple elements of the context or task. Any constructive communication involves providing a brief description or explanation, or presenting a sequence of calculation steps.

Level 3: Recognise and interpret logically complex relations (such as conditional or nested statements) involving the combining of multiple elements and connections. Any constructive communication would involve presenting an explanation or argumentation that links multiple elements of the problem.

Devising strategies

The focus of this variable is on the strategic aspects of mathematical problem solving: selecting, constructing or activating a solution strategy and monitoring and controlling the implementation of the processes involved. 'Strategy' is used to mean any choice of actions to approach the problem in order to solve it. A strategy typically comprises one or more stages each made up of related steps.

The knowledge, technical procedures and reasoning needed to actually carry out the solution process are taken to belong to the representation or the using symbolic, formal and technical language and operations or the reasoning and argument competencies.

Demand for this competency increases with the degree of creativity and invention involved in identifying a suitable strategy, with increased complexity of the solution process (for example the number of steps or the range and complexity of the steps needed in a strategy), and with the consequential need for greater metacognitive control in the implementation of the strategy towards a solution. Identifying and extracting data, analysing those data and performing a calculation on them, would normally constitute separate strategic steps.

Definition: Selecting or devising, as well as controlling the implementation of, a mathematical strategy to solve problems arising from the task or context.

Level 0: Take direct actions, where the strategy needed is explicitly stated or obvious.

Level 1: Find a straight-forward strategy (usually of a single stage) that combines the relevant given information to reach a result or conclusions.

Level 2: Devise a straight-forward multi-stage strategy, or use an identified strategy repeatedly, where using the strategy requires targeted and controlled processing, in order to transform given information to reach a conclusion.

Level 3: Devise a multi-stage strategy, where using the strategy involves substantial monitoring and control of the solution process in order to find a conclusion; or evaluate or compare strategies.

Mathematising

The focus of this variable is on those aspects of the modelling cycle (mathematisation, de-mathematisation) that link an extra-mathematical context with some mathematical domain. A situation outside mathematics may require transformation into a form amenable to mathematical treatment, or a mathematical object or result may need to be interpreted and validated in relation to some related situation or context.

The intra-mathematical treatment of ensuing issues and problems within the mathematical domain is dealt with under the devising strategies competency, the representation competency and the using symbolic, formal and technical language and operations competency. Hence, while this variable deals with representing extra-mathematical contexts by means of mathematical phenomena and entities, the representation of mathematical entities by other entities (mathematical or extra-mathematical) is dealt with under the representation competency.

Demand for activation of this competency increases with the degree of creativity, insight and knowledge needed to translate between the context elements and the mathematical structures of the problem.

Definition: ***Mathematising*** an extra-mathematical situation (which includes structuring, idealising, making assumptions, building a model), or ***making use*** of a given or constructed model by ***interpreting*** or validating it in relation to the context.

Level 0: Either the situation is purely intra-mathematical, or the relationship between the extra-mathematical situation and the model is not relevant to the problem.

Level 1: Make an inference about the situation directly from a given model; translate directly from a situation into mathematics where the structure, variables and relationships are given.

Level 2: Modify or use a given model to satisfy changed conditions or interpret inferred relationships; or identify and use a familiar model within limited and clearly articulated constraints; or create a model where the required variables, relationships and constraints are clear.

Level 3: Link, compare, evaluate or choose between different given models; or create a model in a situation where the assumptions, variables, relationships and constraints are to be identified or defined, and check that the model satisfies the requirements of the task.

Representation

PISA items are generally presented in text form, often with some graphic (including diagrammatic or tabular) material that helps set the context, and sometimes with graphic material that carries a representation of some key mathematical element of the problem. The problem solver may need to devise such a representation, or to link different representations of mathematical objects in order to make progress towards a solution. By 'mathematical representation' we understand a concrete expression (mapping) of a mathematical entity – a concept, object, relationship, process or action. It can be physical, verbal, symbolic, graphical, tabular, diagrammatic or figurative. The existence of simple text instructions and photographs by themselves do not generally involve activation of representation competency. The act of extracting relevant mathematical information, from text or from a graph or table, is where this variable commences to apply.

While this variable deals with representing mathematical entities by means of other entities (mathematical or extra-mathematical), the representation of extra-mathematical contexts by mathematical entities is dealt with under the mathematising competency.

Demand for this competency increases with added complexity of interpretations of representations, with the need to integrate information from multiple representations, and with the need to devise representations rather than to use given representations.

Demand for this competency increases with added complexity of interpretations of representations, with the need to integrate information from multiple representations, and with the need to devise or analyse interpretations.

Definition: ***Interpreting***, translating between, and ***making use*** of given mathematical representations; ***selecting*** or ***devising*** representations to capture the situation or to present one's work. The representations referred to are depictions of mathematical objects or relationships, which include symbolic or verbal equations or formulae, graphs, tables, diagrams.

Level 0: Directly operate on a given representation where minimal interpretation is required in relation to the situation, for example going directly from text to numbers, reading a value directly from a graph or table.

Level 1: Explore or use a given standard representation in relation to a mathematical situation, for example to compare data, to depict or interpret trends or relationships.

Level 2: Understand and use a representation that requires substantial decoding and interpretation; or translate between and use different standard representations of a mathematical situation, including modifying a representation; or construct a representation of a mathematical situation.

Level 3: Understand and use multiple representations that require substantial decoding and interpretation; or compare or evaluate representations; or link representations of different mathematical entities; or devise a representation that captures a complex mathematical situation.

Using symbolic, formal and technical language and operations

This variable reflects competency with activating mathematical content knowledge, such as mathematical definitions, results (facts), rules, algorithms and procedures, recalling and using symbolic expressions, understanding and manipulating formulae or functional relationships or other algebraic expressions. This variable also includes using mathematical concepts definable by means of symbolic expressions (e.g. ‘speed’, or ‘average’) and the formal rules of operations involved in manipulating symbolic expressions (e.g. arithmetic calculations or solving equations).

Note that this variable is abbreviated to symbols and formalism for convenience.

Demand for this competency increases with the increased complexity and sophistication of the mathematical content and procedural knowledge required.

Definition: Understanding and ***implementing*** mathematical procedures and language (including symbolic expressions and arithmetic operations), governed by mathematical ***conventions and rules***; understanding and ***utilising constructs*** based on definitions, results, rules and ***formal systems***.

Level 0: Activate only elementary mathematical facts, rules, terms, symbolic expressions or definitions (for example, arithmetic calculations are few and involve only easily tractable numbers).

Level 1: Make direct use of a simple formally expressed mathematical relationship (for example, familiar linear relationships); use formal mathematical symbols (for example, by direct substitution or sustained arithmetic calculations involving fractions and decimals); use repeated or sustained calculations from level 0; or activate and directly use a formal mathematical definition, fact, convention or symbolic concept.

Level 2: Use and manipulate symbols (for example, by algebraically rearranging a formula); activate and use formally expressed mathematical relationships having multiple components; employ rules, definitions, results, conventions, procedures or formulae using a combination of multiple relationships or symbolic concepts; use repeated or sustained calculations from level 1.

Level 3: Apply multi-step formal mathematical procedures; work flexibly with functional or involved algebraic relationships; use both mathematical technique and knowledge to produce results; use repeated or sustained calculations from level 2.

Reasoning and argument

This variable combines elements of the ‘thinking and reasoning’ and ‘argumentation’ competencies from Niss (2011). It relates to both the **internal mental processing of information** involved in drawing the inferences needed to answer the question, and the **expression of mental processing** needed to explain, justify or prove a result.

The mental processing and reflections needed to choose or devise an approach to solve a problem are dealt with under the devising strategies competency.

The length and complexity of a chain of reasoning or chain of argument needed would be important contributors to increased demand for this competency.

Definition: Logically rooted thought processes that explore and connect problem elements so as to ***make inferences*** from them, or to ***check a justification that is given*** or ***provide a justification*** of statements.

Level 0: Make direct inferences from the information and instructions given.

Level 1: Join information in order to make inferences, (for example to link separate components present in the problem, or to use direct reasoning within one aspect of the problem).

Level 2: Analyse information (for example to connect several variables) to follow or create a multi-step argument; reason from linked information sources.

Level 3: Synthesise and evaluate, use or create chains of reasoning to check or justify inferences or to make generalisations, drawing on and combining multiple elements of information in a sustained and directed way.

Methodology of the investigation

This investigation has been iterative. Commencing with an earlier version of the set of competency definitions and level descriptions, members of the research team independently scrutinised a subset of PISA mathematics test questions that had been used in the PISA 2003 survey, and rated each question on a four-point scale according to the extent to which answering the question called for activation of each of the six competencies.

The ratings data were collected and analysed to evaluate outcomes such as the degree of consistency of ratings applied by the different experts to the same questions, and the variability of ratings applied by individual experts to different questions.

The ratings were also analysed together with the empirical item difficulties of the rated items. The main purpose of that analysis was to investigate the relationship between the ratings and item difficulties across items.

At the conclusion of the rating exercise, the experts met to discuss their ratings, focussing on particular items and also on particular rating categories. The purpose was to establish improved consensus on the interpretation of category definitions, to identify instances where the wording of category definitions and level descriptions could be improved, and to refine those definitions and descriptions.

The process was then repeated using a different set of items from the PISA 2003 item pool, and using the refined category definitions and level descriptions that resulted from the initial phase of the work. Again the rating data were analysed. The outcomes of that analysis were reported to a PISA conference in Kiel in 2009 (Turner, Dossey, Blum and Niss, in press).

A third iteration of the same process occurred following the development of a new pool of items for the PISA 2012 survey, but prior to generation of empirical item difficulties. Experts used the refined definitions and descriptions to rate a subset of new items, the ratings were analysed, and subsequently the relationship between the ratings and the difficulty of the items was analysed. The results of that analysis are presented in the paper by Ray Adams that is part of this symposium.

Results of the investigation

A detailed presentation of the results of the analyses conducted at different stages of this investigation can be found in Turner, Dossey, Blum and Niss (in press) and in Ray Adams's contribution to this symposium.

In summary, we have found that the category definitions and rating level descriptions can be used in a reasonably consistent way by trained experts to evaluate sets of PISA mathematics survey items. The degree of consistency within and across different raters is quite high. The six variables used appear to capture different aspects of the cognitive demand of survey items. Perhaps most interestingly, the ratings contributed by a small group of experts can be used to predict a substantial proportion of the variation in item difficulty.

It appears that the scheme is not yet sufficiently developed to support highly reliable item ratings by a single rater, but using average ratings across a small group of raters generates highly reliable data. It is expected that further refinement of the scheme, including development of additional annotated examples designed to establish more clearly the meaning of each category and the intended interpretation of the level descriptions, will increase the likelihood that the scheme can be used by individuals for any relevant purpose desired.

Conclusions and further observations

As further evidence accumulates about the relationship between the difficulty of mathematics test items, and the demand for activation of mathematical competencies such as those discussed here, the more convinced we become that these competencies should be targeted more directly in school mathematics. While it is clear that other factors bear on what drives mathematical difficulty, if a set of six competencies can be used to predict 60-70% of the variability in item difficulty, this is an important finding.

One area for possible teacher action lies in the kind of problems used as vehicles for teaching and practice: these must include items that allow for treatment of mathematics in context. These items should allow for some decision-making (strategic thinking) about approach, as opposed to problems for which the solution path is specified or obvious; and for the modelling process to be used – both mathematising aspects of the context, and de-mathematising the result (interpreting a mathematical result in relation to the context), as opposed to simply replicating learned procedures and algorithms. Teachers should familiarise themselves with the available materials designed to promote mathematical modelling. The tasks used should also promote the use of representations of the situation, including multiple representations where appropriate, and preferably tasks should provide the opportunity for students to manipulate the representations as part of the solution process.

A second area for teacher action lies in the kind of conversation that is promoted in the classroom, and the nature of the mathematical communications that are expected of students. Students need to practise expressing their ideas and conclusions through both speaking and writing. They need to be given the opportunity to consider the mathematical communications of others, and to engage with the mathematical ideas being expressed. Communication skills must be developed consciously, including the ability to explain and argue a case mathematically. But expressing oneself mathematically also helps to develop and cement mathematical thinking and reasoning processes.

March 2012

References

Niss, M. (1999). Kompetencer og uddannelsesbeskrivelse (Competencies and subject description), *Uddannelse*, 9, 21-29.

Niss, M.A. (2003). Mathematical Competencies and the Learning of Mathematics: The Danish KOM Project. In Gagatsis, A. & Papastavridis, S. (Eds), *3rd Mediterranean Conference on Mathematical Education*. Athens: The Hellenic Mathematical Society, 110-124.

Niss, M. A. and Højgaard, T. (Eds,) (2011). *Competencies and Mathematical Learning: Ideas and inspiration for the development of teaching and learning in Denmark* (IMFUFA tekst). Roskilde: Roskilde University.

OECD (1999). Measuring Student Knowledge and Skills: A New Framework for Assessment. Paris: Directorate for Education, OECD.

OECD (2004). Learning for Tomorrow's World – First Results from PISA 2003. Paris: Directorate for Education, OECD.

Turner, R., Dossey, J., Blum, W., and Niss, M. A. (in press). Using mathematical competencies to predict item difficulty in PISA. In Prenzel, M., Kobarg, M., Schöps, K., and Rönnebeck, S. (Eds), *Research in the context of the Programme for International Student Assessment*.