National numeracy tests: A graphic tells a thousand words

Tom Lowrie
Charles Sturt University
Carmel M. Diezmann
Queensland University of Technology

Mandatory numeracy tests have become commonplace in many countries, heralding a new era in school assessment. New forms of accountability and an increased emphasis on national and international standards (and benchmarks) have the potential to reshape mathematics curricula. It is noteworthy that the mathematics items used in these tests are rich in graphics. Many of the items, for example, require students to have an understanding of information graphics (for example, maps, charts and graphs) in order to solve the tasks. This investigation classifies mathematics items in Australia's inaugural national numeracy tests and considers the effect such standardised testing will have on practice. It is argued that the design of mathematics items is more likely to be a reliable indication of student performance if graphical, linguistic and contextual components are considered both in isolation and in integrated ways as essential elements of task design.

New forms of accountability: The national testing agenda

In most educational settings, and particularly in schools, standardised measures of student performance are increasingly influencing (and possibly driving) practice and the day-to-day decisions that teachers make. A worldwide move towards centralised testing, which took place in the early 1990s—for example, in England (Office of Her Majesty's Chief Inspector of Schools—Ofsted) and the USA (especially the 2001 *No Child Left Behind Act*)—has dramatically increased the volume of data that teachers are either required to interpret or compile in relation to their practice (Avenell, 2006). The initial momentum was associated with public accountability that, not surprisingly, was aligned to an era of economic rationalism. Schools and other institutions promoting this model of accountability use high-stakes testing, with obvious consequences for schools, teachers and students who fail to meet the systemic benchmarks. In Australia, the new Rudd Labor federal government—as part of its commitment to implement a national curriculum—has indicated that published reporting of national testing results will occur. Even education systems that were not subjected to such benchmarking scrutiny are moving toward much more comprehensive standardised testing.

Today, benchmarking is common practice in most school systems. As Smith indicates:

> most schools are now being bombarded with information and data related to their students' performance . . . [and] such data [can] provide a significant resource to encourage school improvement—provided it's accurately interpreted and effectively devolved to the relevant teachers in the classroom. (2005, p. 12)

The sophistication and detail of the information (and data) presented to classroom teachers differs markedly from country to country, but international testing instruments (for example, Trends in International Mathematics and Science Study—TIMSS—and the Program for International Student Assessment—PISA) are creating some form of reporting consistency.

Although there are strong proponents of formalised testing (for example, Coyne & Harn, 2006), much of the data generated is limited to 'snapshots' of student performance as large cohorts (often in relation to national averages). Consequently, the information teachers receive is relatively unsophisticated, generic, and only slightly more detailed than the information given to parents (Jones & Egley, 2007). Such information usually includes graphs that place individuals or cohorts on a continuum that is divided into grades or proportional clusters, utilising similar information that compares this cohort to other groups (for example, students in other states or regions), and percentile breakdowns for individual questions or combinations of questions (for example, strands in mathematics or areas of study in literacy). There is growing concern that teachers may use testing to 'drive' their teaching—anticipating what may be included in assessment and then teaching accordingly. Indeed, a growing body of research (for example, Jones & Egley, 2007; Pedulla et al., 2003) shows that mandatory testing has been a powerful influence on what gets taught in classrooms and, to a lesser extent, on the methods of instruction. Moreover, Stecher and Barron found that 'more teachers reported increasing the amount of time spent on subjects that were tested at their grade level than on subjects that were not tested' (2001, p. 268). Interestingly, primary teachers were found to be more influenced by testing than their secondary colleagues, while a majority of teachers at each grade level indicated that state testing programs have led them to teach in ways that contradict their ideas of sound instructional practices (Pedulla et al., 2003). There is also a disquiet among the education community that teachers may be judged more by how they educate their students for testing rather than taking a student-driven approach (Hattie, 2005). Furthermore, elements of the curriculum that are not easily testable (and thus measurable), such as open-ended problem-solving, are at risk of being squeezed out of the classroom: future curriculum standards could be lowered to allow more students to perform well in standardised testing (Hattie, 2005). As McNeil (2000) argues, placing a premium on students' performance in tests has led to instruction that is focused primarily on test pre-paration, thus limiting the range of educational experiences awarded to students and potentially reducing the instructional skills of teachers.

Despite such concerns, an increasing number of countries are establishing mandatory testing across primary and secondary schooling. It is noteworthy that

Australia has among the highest numbers of mandatory tests (four tests in Years 3, 5, 7 and 9) in the world (O'Donnell & Sargent, 2008). In the following sections we argue that the lack of attention to the graphical component in tests is highly problematic because the alignment of content and tasks is particularly important in high-stakes assessment and in informing instructional practice (Kulm, Wilson & Kitchen, 2005).

The Australian context

In May 2008, approximately 1 million Australian students in Years 3, 5, 7 and 9 participated in the inaugural national numeracy testing. Previously, standardised testing was conducted at a state level—with as many as seven different numeracy tests being administered by different states to students of different ages—and thus no nationwide comparisons were available except through an equating process. The national testing agency will generate reports to various stakeholders at different levels of analysis:

> The results from these national literacy and numeracy tests will provide an important measure of how Australian schools and students are performing in the areas of reading, writing, spelling and *numeracy*. The results from the assessment program will be used for individual student reporting to parents, school reporting to their communities, and aggregate reporting by States and Territories against national standards. (Curriculum Corporation, n.d., emphasis added)

It is noteworthy that this national assessment agenda is the first step toward Australia's inaugural national curriculum. As the framework for the national curriculum is being developed, data from the national assessment instruments will be used to make comparisons about student, school, and state and territory performance. In this paper, we will highlight the problematic nature of such reporting—particularly if the teaching and learning experiences of any new curriculum are overly influenced by student results on such assessments. In addition, we will argue that the increasing role of information graphics in the construction of mathematics items should be considered. Nevertheless, these graphic representations need to be used appropriately, otherwise fallacious and misleading impressions of student performance will eventuate—which in turn will create unreliable data for decision-making.

Graphics in an information age

In what could be considered a burgeoning information age, our society has become more reliant (often from necessity) on representing information in diagrammatical and graphical forms. Such information, for which multiple representations are often provided, uses dynamic forms of spatial and visual information to manipulate images. At the same time, school curricula are becoming increasingly graphic in nature with the mathematics curriculum, in particular, moving away from predominantly word-based problems to the integration of graphical representations to convey information (Lowrie & Diezmann, 2005). In addition, the nature of graphics-based tasks has also changed, with multiple representations and increased detail embedded within graphics. As a consequence, mathematics tasks are more likely to include graphics information, and the graphics are more detailed and generally represent information with increased richness. Furthermore, the entire nature of test design has changed dramatically in recent years as graphical and visual representations become increasingly embedded within items.

A comparison of two state-based tests in Australia (over a 13-year period) revealed distinct differences between the number of graphic items included in the respective tests and the richness of the graphics presented. Figure 1 shows items from two New South Wales Basic Skills tests, 13 years apart. It illustrates a difference in graphic richness as well as a change in literary demands (for example, the worded instructions to complete the task). The first task requires the student to interpret a two-dimensional graphic that is relatively free of detail and information that could be considered distracting. By contrast, the second task presents information in more detailed and saturated ways (including both two-dimensional, bird's-eye perspectives and elevation perspectives of three-dimensional objects). We are not suggesting that one task is more challenging or worthwhile than the other.

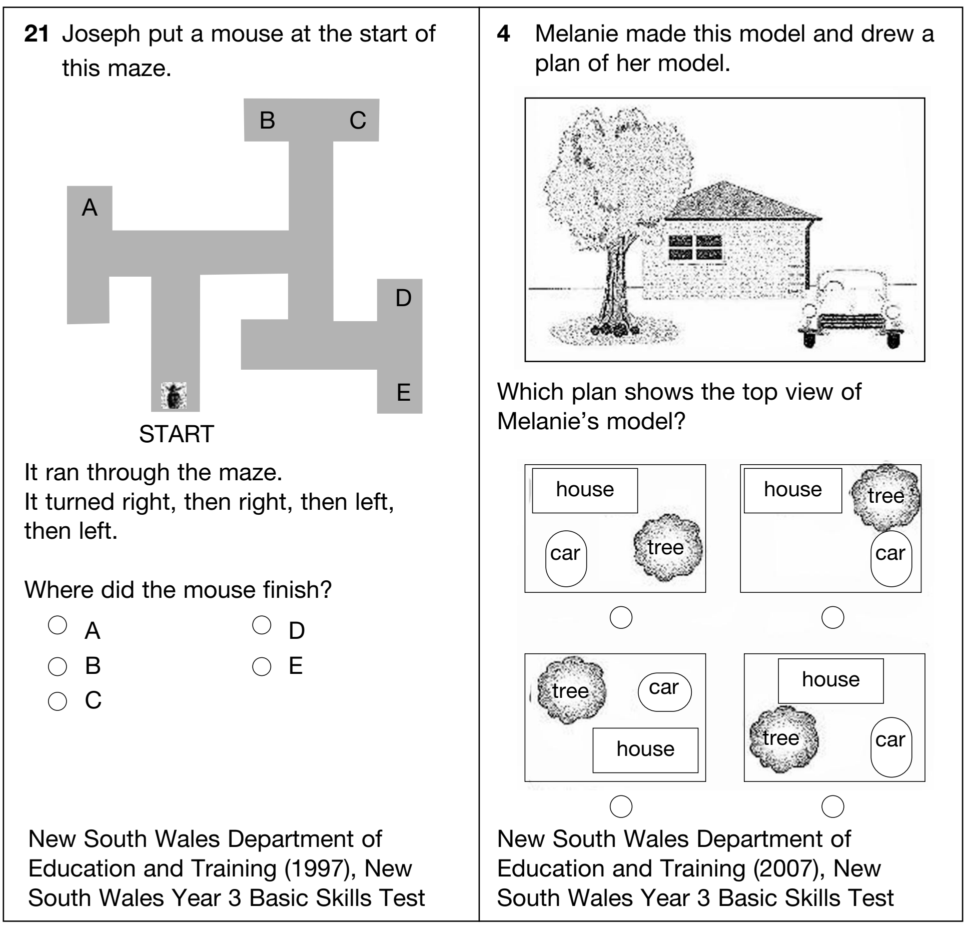


Figure 1 A comparison of test items more than 10 years apart

We do contend, however, that the mathematics understandings and problem-solving processes required to complete these are different.

Roth (2002) argues that greater attention must be given to the practices of reading, producing and understanding graphical representations. Lowrie and Diezmann (2005) maintain that the explicit teaching of such practices has to occur in order for students to effectively decode graphical information. As educational bodies place increased emphasis on the importance of graphical representations (for example, Australian Association of Mathematics Teachers, 1997; Department for Education and Employment, UK, 1998; National Council of Teachers of Mathematics, 2000), it is unsurprising that standardised testing has taken a similar course.

(Re)presenting graphics in assessment

Visual representations, such as graphs, diagrams, charts, tables, and maps are part of the emerging field of information graphics found throughout current school curricula. Such graphics are regularly used to represent mathematics content in standardised testing (Diezmann, 2008; Logan & Greenlees, 2008). It is somewhat problematic that research on the use and understanding of images and graphics is quite limited (Postigo & Pozo, 2004), despite the view that such forms of mathematics literacy are essential in today's society (Goldin, 1998; Zevenbergen, 2004). It is also noteworthy that scant consideration has been awarded to the general view of this literacy with respect to assessment. Postigo and Pozo (2004) argue that previous research conducted in this field is quite heterogeneous, since the study of maps, diagrams and numerical graphs has its own syntax and conventions. In addition, most studies have considered student performance (from a correctness perspective)—and therefore considered graphical problems—in relation to the understanding of mathematics content rather than a student's ability to make sense of the graphical component of the task. It is also the case that student performance across different types of graphics (for example, number lines and maps) is not generally strong (Lowrie & Diezmann, 2005) and that correlations between items within the same graphic type are at best moderate (Lowrie & Diezmann, 2007). These findings challenge the view that mathematics content (and thus students' understanding of mathematics concepts) is actually being assessed when mathematics items have substantial graphics attributes. New forms of (assessment) item representation, particularly those rich in graphics, thus place increased attention on students' capacity to decode and interpret the various elements that constitute the task (Diezmann & Lowrie, in press).

Decoding graphics

The decoding of graphics items and tasks requires the student to contend with multiple sources of information that may include text, keys or legends, axes and labels (Kosslyn, 2006); as well as elements of density and saturation (Bertin, 1967/1983). It is therefore necessary to consider these 'components' (which are often interrelated) in conjunction with the actual mathematics that is contained within a given task. As Hittleman (1985) indicates, student thinking can be interrupted simply by moving between the text of a question and the information in the graphic. Even with much older college students, Carpenter and Shah (1998) found that students spent the majority of time analysing information from particular regions of the graphic (for example, moving between the axes and the labels) and were unable to keep track of the information presented in its entirety. The elements used in constructing a graphic have an impact on how well students understand and interpret the task and influence their success in choosing appropriate strategies to use on the task and ultimately to complete the task. For primary-aged students, the comprehension of the graphic can be a demanding aspect of a mathematics task in its own right. The actual mathematics of a given task is not likely to be the critical aspect of reasoning and problem-solving if the student is not able to access and interpret the information effectively. Students' performance may thus be a measure of their ability to comprehend the graphical (or linguistic) components of a task rather than their knowledge of the mathematics within the task. We are concerned that mathematics items constructed for mandatory national tests do not have an adequate alignment between content and the representation of the graphic. Substantial data is obtained (and reported) on student performance on mathematics tasks but rarely do we consider whether the tasks actually assess student knowledge and numeracy understandings.

The nature of graphical composition

Kosslyn (2006) suggested that the graphical composition of a task included not only the actual graphic but also all of the information embedded within the task. Research conducted with colleagues (Diezmann & Lowrie, 2008; Logan & Greenlees, 2008) has indicated that it is difficult to separate the graphical features that are embedded in a task from other demands (including mathematical content and linguistic demands). As Brna, Cox and Good (2001) suggest, diagrammatic reasoning is influenced by the nature of the task, the semantic properties of the diagram, and the person's prior knowledge, which include skills, preferences and experiences). The actual graphical components influence task complexity since the student needs to be aware of the content domain and conventions regulating sign use to decode mathematical formulae and graphs. These structures intend to provide the spatial framework that helps to organise information and the particular conventions that represent information. As a consequence, errors may occur not because of 'misconceptions' or limited cognitive 'understandings', but rather because students are unfamiliar with the contexts and situations for which such conventions are constructed and the extent to which contextual meaning (and experiences) influence the interpretation of the graphic (Roth, 2002).

The context in which mathematical content is presented may influence a student's initial sense of a task, leading to the use of routine and highly practised responses. For example, with language, students may pay only superficial attention to the written text within a task, finding key words that may indicate important information relating to the graphic. This can hinder students' holistic understanding of the task, and hence, the rationality and correctness of their answers (Wiest, 2003). Additionally, as Boaler maintains, many tasks require students to 'suspend reality and ignore their common sense in order to get a correct answer' (1994, p. 554).

The actual literacy demands required to interpret a task also have an impact on sense making—particularly with young children, as they interrogate data and interpret the multiple meanings that often accompany their vocabulary and concept development. The multiple layering of 'meaning' is also applied to the use of language in everyday contexts and interactions (Adams, 2003). For example, the word 'flip' has both an everyday meaning and a mathematical meaning; young children need to be able to appropriately identify this term wherever it is used. Specific processes associated with 'working mathematically', including questioning, communicating and reasoning, provide opportunities for students to unpack the vocabulary embedded in tasks, and thus help reduce literacy demands. Furthermore, explicit teaching of terminology that is applied to mathematics (for example, volume and net) needs to be undertaken. We are not suggesting teachers should be teaching to the test but rather that the elements that constitute a mathematics task need to be understood.

The role and nature of information graphics in national tests

This section presents an analysis of the mathematics items used in the inaugural Year 3 and Year 5 Australian national numeracy tests (Ministerial Council on Education, Employment, Training and Youth Affairs, 2008a; 2008b) to ascertain the role of information graphics in the tests and to review the type (category) of graphics that are used. At a functional level, graphics can be classified as either context graphics or information graphics (Diezmann, 2008). Context graphics (see Figure 2) are often used for illustrative purposes to represent objects, people or locations. They contain no mathematical information pertinent to the task and can often be misleading. By contrast, information graphics (see Figure 3) are an integral component of a task—with information embedded within the graphic needing to be decoded in order to solve the task. There are many thousands of information graphics but Mackinlay (1999) categorises them into six types that he refers to as graphical languages. These languages are 'axis', 'opposed position', 'retinal list', 'map', 'connection' and 'miscellaneous'. Like text-based languages, graphical languages have unique signs, symbols and characteristics. An overview of each graphical language is shown in Table 1.

An analysis of item representation in national numeracy tests

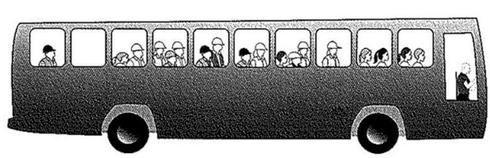
What proportion of items from the new national tests contain graphics?
From the 75 items across the two tests, a total of 64 items (85%) contained either information (n = 45) or context graphics (n = 19) (see Table 2). As a result, students' ability to make distinctions between these two types of graphics and consequently to use them appropriately will affect their performance. For example, in Figure 2 students do not need to use the image of the bus to answer the associated word problem as all relevant information is in the written text. By contrast, in Figure 3 students should use the information graphic to determine the number of

sit-ups Manu does on Wednesday. For some students, knowing when to and how to extract information embedded in graphics can be problematic. Elsewhere, Diezmann, Lowrie and Kozak (2007) have found that low-performing students tend to draw upon everyday knowledge not specifically relevant to the actual task in order to generate a solution, whereas high-performing students intuitively draw on implicit information embedded within a graphic to decode a task.

Table 1 Structure and functionality of the six graphical languages

Graphical languages	*Graphical knowledge*	*Mathematics functionality*
Axis (e.g., number line)	Relative position of a mark on an axis	Number line as a measurement model
Opposed position (e.g., graph)	Relative position of marked sets of points between two axes	Everyday use of graphs
Retinal list (e.g., mental rotation, flip)	Conventions in using colour, shape, size, saturation, texture, or orientation in representation; markings are not dependent on position	Translations, rotations, reflections, discrimination skills
Map	Model of spatial representation of locations or objects and the convention of key use	Bird's-eye view, two-dimensional and three-dimensional representations
Connection (e.g., family tree)	Conventions of structured networks with nodes, links and directionality	Everyday applications (e.g., train maps, knockout competitions)
Miscellaneous (e.g., calendar)	Conventions of additional graphical techniques (e.g., angle, containment) in representation	Various, depending on the graphic

A bus took some students to camp.
It left the school at 10:00 a.m.



The bus trip took one and a quarter hours.
What time did the bus get to the camp?

Figure 2 A context graphic (Ministerial Council on Education, Employment, Training and Youth Affairs, 2008a, Year 3 numeracy test, item 22)

	Manu's program		
	Monday	Tuesday	Wednesday
Sit-ups	10	15	20
Push-ups	6	9	12
Jumps	15	20	25

How many **sit-ups** does Manu do on Wednesday?

Figure 3 An information graphic (Ministerial Council on Education, Employment, Training and Youth Affairs, 2008a, Year 3 numeracy test, item 12)

What types of graphics are included in the two sets of test items? An analysis of the test items based on Mackinlay's (1999) six graphical languages reveals that all graphical languages were represented across both tests (Table 2). Miscellaneous (38%) and retinal list (29%) items were more commonly used, while map and axis items (both 8%) were used less frequently. Most surprising was the fact that opposed position items (which included bar and column graphs) were very much under-represented, despite the fact that they feature so predominantly in school curricula. It was noteworthy that connection items were not represented in either test, despite the fact that the interpretation of family trees and sporting draws (for example, tennis) require such processing and, in fact, helps with important mathematical skills such as proportional and logical reasoning.

Table 2 Proportions of test items that contain graphics by year and type

Year	*Total items*	*Graphics items*	*Context graphics*	*Information graphics: graphical languages*					
				*MI**	*RL*	*MA*	*AX*	*OP*	*CO*
3	35	91% (n = 32)	11	9	6	2	1	2	1
5	40	80% (n = 32)	8	8	7	3	4	1	1
Total	75	85% (n = 64)	19	17	13	5	5	3	2

Key: MI = miscellaneous; RL = retinal list; MA = map; AX = axis; OP = opposed position; CO = connection
* Note: This includes items that could be classified as visual representations embedded in the item to represent concrete understandings of ideas or symbols

The compositional structure of information graphics

Together with colleagues (Logan & Greenlees, 2008), we have investigated the influence that graphics representation has on student performance and sense making. Forty Grade 6 students from three regional schools in New South Wales, Australia took part in this study. The participants were asked to solve the six items from the Graphical Languages In Mathematics instrument (Diezmann & Lowrie, in press), as part of an ongoing analysis of their mathematics decoding performance. After analysing student responses, these items were modified with changes to either graphic or non-graphic (including context and literacy demands) elements. Thus, the six *modified* items were variations of those in the standard instrument. The Appendix presents both standard and modified items. The participants completed the modified items approximately six weeks after solving the standard items. We assumed that the modified items would provide opportunities for the students to use more efficient strategies to complete the tasks. Furthermore, we anticipated that the item modification would provide scope to consider student sense-making in situations where task representation was altered.

For three of the items (Items 1, 4 and 5) the graphic was altered while a non-graphic element was changed for the other three items (see Appendix). The graphic variations included removing pictures from above a number line (Item 1), removing plotted dots from the slope of a line graph (Item 4) and shading the background on a retinal task (Item 5). Non-graphic changes included adding numerals to a line graph (Item 2), bolding a word (Item 3) and changing the context of a task (Item 6). Table 3 provides a description of student success across the standard and modified items. Each of the sets of items have been classified with respect to changes that were made to either the graphic, the wording or the context of the task—specifically the addition or removal of graphical or literacy elements.

Table 3 Student performance across standard and modified tests

Question	*Test*		*Effect size*	*Change*
	A (%) Correct	*B (%) Correct*	*size d*	
1	44	53	.18	*Graphic removed* Emphasis taken away
2	33	35	.04	*Context changed* Data added
3	60	53	−.14	*Wording changed* Emphasis added
4	22	60	.83	*Graphic removed* Emphasis taken away
5	51	65	.29	*Graphic added* Emphasis added
6	24	18	−.14	*Context changed* Emphasis taken away

Effect sizes (measured by Cohen's *d*) revealed the degree of change in performance across the two tests. For four of the six questions effect size was small with such results indicating minor differences between the performance of students across the standard and modified test items. The students' performances increased on four of the six items with the three largest effect sizes being associated with a change to the graphic. By contrast, modification to the literacy or context resulted in only minor improvements or decreases in performance (see Appendix for examples of the items).

The two most significant changes involved an aspect of the graphic being removed (Question 4) or added (Question 5). In Question 4, dots were removed from the slope of the graph with dramatic performance increases occurring (from 22% correct to 60% correct). The removal of the dots allowed children to focus on the movement of the line (and thus interpreting a rest as a plateau of the line) rather than focusing on points along the line (which many students interpreted as a rest). As Logan and Greenlees (2008) explain, many students saw the dots as a pause based on the analogy of a full stop in a sentence. In Question 5, shading was applied to the background of the graphic in an attempt to give definition to the vacant puzzle piece. The addition of the shaded background allowed students to see in their 'mind's eye' that they had to fit the puzzle piece into the vacant space, rather than sliding the piece into the side of one of the options.

Changes to the literacy or context aspects of the items resulted in a negative change in student performances on two of the items. For Question 3, a change in the literacy aspect actually confused students. Many students' incorrect responses to the standard item focused on both variables (i.e., length and weight) as being exact measurements. It was envisaged that by bolding the word 'approximately', those students would assign an approximation to the weight and an exact measurement to the length. In fact, it had the opposite effect with many students who correctly solved the standard item changing their answer in the modified form. The exaggeration of the word 'approximately' drew students' attention to that variable (weight) and this became the measure from which they chose their answer.

In Question 6, students' understanding of the context of the task (that is, knowledge of the food chain) unduly influenced (and thus hindered) their interpretation of the task, with many students' incorrect responses relying on their prior knowledge to answer the question. By changing the food chain to something nonsensical, it was anticipated that students would use the key and the arrows of the graphic to work out a solution. Again, this proved not to be the case, with a number of students who correctly solved the standard item distracted by the unfamiliar context of the modified item. It was still apparent that a majority of the students had trouble applying the key and interpreting the directionality of the arrows in the graphic regardless of the context of the task.

This study highlighted the extent to which a word, a phrase or an element of a graphic could influence students' capacity to decode information. It was evident that variations in graphics had a significant (and generally positive) effect on student performance. Moreover, when the graphical elements of items were modified, many students who had incorrectly solved tasks were able to reason—in sophisticated ways—about the nature and content of the tasks. By contrast, changes in mathematical literacy or context had only a small effect on student performance and sense making (Logan & Greenlees, 2008). It was somewhat disconcerting that item construction had such a large impact on performance rather than student knowledge of concepts. In most cases, the errors involved students not considering information in the graphic, being overly influenced by information (often irrelevant) in the graphic, or not considering the connections between embedded graphical information and the textual and symbolic information.

Since graphics are processed in both verbal- and imaged-based processing systems (Paivio, 1971), it is not surprising that even variations in graphics will change the way an item is represented and thus understood. Younger students are generally more influenced by a graph's structure and content than older students, who possess more sophisticated skills to decode information (Shah & Hoeffner, 2002). As Kirby (1994) maintained, the processing of spatial tasks is quite complex; as a result, students need to receive instruction on how to process such information from the early years of school. It has also been found that changes to context and written information had some impact on student performance, but it was the modifications to the graphic that most supported students' sense-making (Logan & Greenlees, 2008). Since the intent of standardised testing is to provide opportunities for students to show what they understand about a concept, it is essential that the graphics embedded in mathematics items are well designed.

Conclusions and implications

Across many aspects of our day-to-day experiences (and indeed in most areas of the school curricula), information graphics have become increasingly necessary in representing, organising and analysing information. It is not surprising that assessment practices are aligned to such societal and curriculum changes—and it seems to be particularly salient with respect to standardised instruments (Diezmann, 2008; Logan & Greenlees, 2008). The design of mathematics items, particularly those used in standardised instruments, is more likely to be a reliable indication of student performance if graphical, linguistic and contextual components are considered separately and collectively in task design.

Implications for the classroom

We believe that current practice in national numeracy testing is likely to underestimate students' understanding of mathematical concepts unless specific attention is paid to the graphical languages used in the tests. In particular:

- classroom teachers should be conscious of explicitly teaching the various graphical languages in order to support the development of students' ability to decode information graphics. This explicit teaching will remain a challenge if teachers are unable to identify the important attributes (and differences) among items from each of the graphical languages.
- learning opportunities should be broad and include graphical languages that are typically used outside formal mathematics contexts (maps, miscellaneous) in addition to those explicitly incorporated into the mathematics curricula (axis, opposed position).
- specific language or terminology needs to be talked about in a variety of ways because one word (for example, flip) can have a marked influence on responses.

- teachers need to be aware of the fact that knowledge transfer across and within graphical languages is not overly high.
- all graphical elements—such as text, keys or legends, axes and labels—need to be considered when children are learning to decode information graphics.
- teachers should be conscious of the limitations to understanding that can eventuate if graphical representations are restricted to particular prototypes since questions in national tests tend to display graphics from a broad spectrum of sources.

Implications for test designers

For the same reasons, national numeracy testing will yield more authentic data if test developers take account of these issues in their practice:

- the construction of mathematics test items should be developed from a 'holistic design' perspective. Graphics need to be carefully chosen to ensure the integrity (and meaning) of the item is maintained (Diezmann, 2008).
- when creating test items, designers need to not only consider the mathematics experiences students bring to a task but also the assessment experiences students have acquired. The abundance of graphics in mandatory testing is a relatively new phenomenon and consequently even slight changes in graphic representation influence performance. It is essential that poorly constructed graphics do not impact on performance.
- mathematics test items created within a real world or authentic context can often be misinterpreted by students (Boaler, 1993). Designers need to be aware that such representations may not provide a clear indication about student understanding with respect to the mathematical intent of the task.

Keywords

mathematics tests	benchmarking	numeracy
mathematics achievement	student assessment	pedagogy

References

Adams, T. L. (2003). Reading mathematics: More than words can say. *The Reading Teacher, 56*(8), 786–795.

Australian Association of Mathematics Teachers (1997, October). *Numeracy = everyone's business.* Report for the Numeracy Education Strategy Development Conference: Adelaide, Australia.

Avenell, K. (2006). Standardised tests: Losing sight of our true purpose. *Australian Educational Leader, 28*(1), 34–35, 47.

Bertin, J. (1983). *Sociology of graphics* (trans. W. J. Berg). Madison, WI: University of Wisconsin Press. (Original work published 1967.)

Boaler, J. (1993). The role of contexts in the mathematics classroom: Do they make mathematics more 'real'? *For the Learning of Mathematics, 13*(2), 12–17.

Boaler, J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to 'realistic' mathematic contexts. *British Educational Research Journal, 20*, 551–564.

Brna, P., Cox, R., & Good, J. (2001). Learning to think and communicate with diagrams: 14 questions to consider. *Artificial Intelligence Review, 15*, 115–134.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*, 75–100.

Coyne, M., & Harn, B. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools, 43*(1), 33–43.

Curriculum Corporation. (n.d.). *National literacy and numeracy testing 2008.* Retrieved February 6, 2008, from http://www.curriculum.edu.au/ccsite/cc_nap,19875.html

Department for Education and Employment, UK. (1998). *The implementation of the national numeracy strategy. The final report of the numeracy taskforce.* London: Author.

Diezmann, C. M. (2008). Graphics and the national numeracy tests. In M. Goos, R. Brown & K. Makar (Eds.), *Navigating currents and charting directions: Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia, 2*, 659–662. Sydney: Mathematics Education Research Group of Australasia.

Diezmann, C. M. & Lowrie, T. (2008). The role of information graphics in mathematical proficiency. In M. Goos, R. Brown & K. Makar (Eds.), *Navigating currents and charting directions: Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia, 2*, 647–650. Sydney: Mathematics Education Research Group of Australasia.

Diezmann, C. M., & Lowrie, T. (in press). An instrument for assessing primary students' knowledge of information graphics in mathematics. *Assessment in Education: Principles, Policy and Practice.*

Diezmann, C.M., Lowrie, T., & Kozak, N. (2007). Essential differences between high and low performers' thinking about graphically-oriented numeracy items. In J. Watson & K. Beswick (Eds.), *Mathematics: Essential research, essential practice: Proceedings of the 30th annual conference of the Mathematics Education Research Group of Australasia, 1*, 226–235. Adelaide: Mathematics Education Research Group of Australasia.

Educational Testing Centre. (2002). *Year 6 Australian Schools Science Competition.* Sydney: Educational Assessment Australia, The University of New South Wales. [Item 3 in the Appendix]

Educational Testing Centre. (2001). *Year 5 Australian Schools Science Competition.* Sydney: Educational Assessment Australia, The University of New South Wales. [Item 1 in the Appendix]

Educational Testing Centre. (1995). *Year 7 Australian Schools Science Competition.* Sydney: Educational Assessment Australia, The University of New South Wales. [Item 6 in the Appendix]

Goldin, G. A. (1998). Representational systems, learning and problem solving in mathematics. *Journal of Mathematical Behaviour, 17*(2), 137–165.

Hattie, J. (2005). What is the nature of evidence that makes a difference to learning? In *Using data to support learning: Proceedings of the 10th national research conference of the Australian Council for Educational Research* (pp. 11–21). Melbourne: ACER.

Hittleman, D. R. (1985). A picture is worth a thousand words . . . if you know the words. *Childhood Education,* 61–62, 32–36.

Jones, B. D., & Egley, R. J. (2007). Learning to take tests or learning for understanding? Teachers' beliefs about test-based accountability. *The Educational Forum, 71*, 232–248.

Kirby, J. R. (1994). Comprehending and using maps: Are there two modes of map processing? In W. Schnotz & W. Kulhavy (Eds.), *Comprehension of graphics* (pp. 63–77). Amsterdam: Elsevier Science B. V.

Kosslyn, S. M. (2006). *Graph design for the eye and mind.* New York: Oxford University Press.

Kulm, G., Wilson, L. D., & Kitchen, R. (2005). Alignment of content and effectiveness of mathematics assessment items. *Educational Assessment, 10*(4), 333–356.

Logan, T., & Greenlees, J. (2008). Standardised assessment in mathematics: The tale of two items. In M. Goos, R. Brown & K. Makar (Eds.), *Navigating currents and charting directions: Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia, 2*, 655–658. Brisbane: Mathematics Education Research Group of Australasia.

Lowrie, T., & Diezmann, C. M. (2005). Fourth-grade students' performance on graphical languages in mathematics. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th annual conference of the International Group for the Psychology of Mathematics Education, 3*, 265–272). Melbourne: Psychology of Mathematics Education.

Lowrie, T., & Diezmann, C. M. (2007). Solving graphics problems: Student performance in the junior grades. *Journal of Educational Research, 100*(6), 369–377.

Mackinlay, J. (1999) Automating the design of graphical presentations of relational information. In S. K. Card, J. D. Mackinlay, & B. Schneiderman (Eds.), *Readings in information visualization: Using vision to think* (pp. 66–81). San Francisco, CA: Morgan Kaufmann.

McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2008a). *National assessment program: Literacy and numeracy. Year 3 numeracy test.* Melbourne: Curriculum Corporation.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2008b). *National assessment program: Literacy and numeracy. Year 5 numeracy test.* Melbourne: Curriculum Corporation.

National Center for Educational Statistics, National Assessment of Educational Progress. (2003). *Year 4 mathematics.* Retrieved September 9, 2004 from http://nces.ed.gov/nationsreportcard/ITMRLS/search.asp?picksubj=Mathematics [Item 2 in the Appendix]

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

New South Wales Department of Education and Training. (1997). *Basic skills test: Aspects of numeracy*. Sydney: Author.

New South Wales Department of Education and Training. (2007). *Basic skills test: Aspects of numeracy*. Sydney: Author.

O'Donnell, S., & Sargent, C. (with Gray, J.) (2008). *International review of curriculum and assessment frameworks internet archive comparative tables*. Retrieved October 28, 2008, from International review of curriculum and assessment frameworks Internet archive http://www.inca.org.uk/INCA_comparative_tables_July_2008.pdf

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Pedulla, J. J., Abrams, L. M., Madaus, M. K., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers.* Boston, MA: National Board on Educational Testing and Public Policy. Retrieved March 15, 2009, from http://www.bc.edu/research/nbetpp/statements/nbr2.pdf

Postigo, Y., & Pozo, J. (2004). On the road to graphicacy: The learning of graphical representation systems. *Educational Psychology, 24*(5), 623–644.

Queensland School Curriculum Council. (2001). *2001 Queensland Year 3 test: Aspects of numeracy*. Brisbane: Author. [Item 5 in the Appendix]

Queensland Studies Authority. (2002). *2002 Queensland Year 5 test: Aspects of numeracy.* Brisbane: Author. [Item 4 in the Appendix]

Roth, W.-M. (2002). Reading graphs: Contributions to an integrative concept of literacy. *Journal of Curriculum Studies, 34*(2), 1–24.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review, 14*, 47–69.

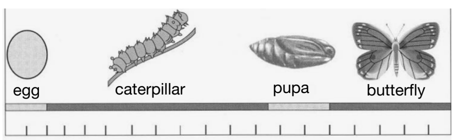
Smith, V. (2005). School-based research: a new approach to school improvement. *Professional Educator, 4*(4), 8–12.

Stecher, B. M., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in 'milepost' grades. *Educational Assessment,* 7(4), 259–281.

Wiest, L. R. (2003, May). Comprehension of mathematical texts. *Philosophy of Mathematics Education Journal, 17*. Retrieved September 29, 2006, from http://www.people.ex.ac.uk/PErnest/pome17/pdf/lwiest.pdf

Zevenbergen, R. (2004). Reconceptualising numeracy for New Times. *Curriculum Perspectives, 24*(3) 1–7.

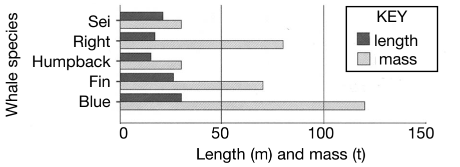
Appendix: The standard and modified versions of items

Standard item	Modified item
1 The following graph shows the length of time taken for the four stages in the life of a butterfly.  How many days are there in the caterpillar stage? Answer ☐ 4 ☐ 19 ☐ 23 ☐ 38	**1** The following graph shows the length of time taken for the four stages in the life of a butterfly.  How many days are there in the caterpillar stage? Answer ☐ 4 ☐ 19 ☐ 23 ☐ 38

2 Bay City Exton Yardville  On the road shown above, the distance from Bay City to Exton is 60 kilometres. What is the distance from Bay City to Yardville? Answer ☐ 45 kilometres ☐ 75 kilometres ☐ 90 kilometres ☐ 105 kilometres	**2** Bay City Exton Yardville  On the road shown above, the distance from Bay City to Exton is 60 kilometres. What is the distance from Bay City to Yardville? Answer ☐ 45 kilometres ☐ 75 kilometres ☐ 90 kilometres ☐ 105 kilometres

Standard item

3 The graph compares the maximum length and mass to which some whales grow.



A fisherman reported that a whale 25 metres long and weighing approximately 80 tonnes had beached itself.
Which species of whale could this be?

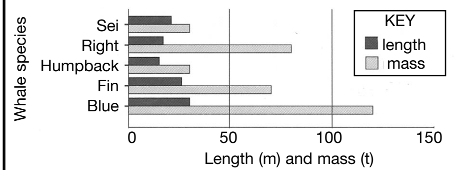
Answer

☐ Right whale ☐ Humpback whale

☐ Fin whale ☐ Blue whale

Modified item

3 The graph compares the maximum length and mass to which some whales grow.



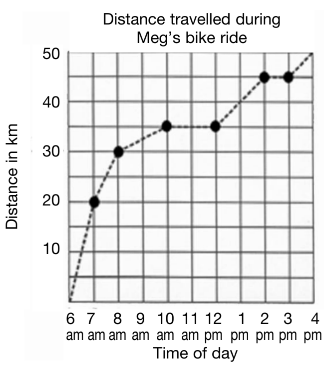
A fisherman reported that a whale 25 metres long and weighing **approximately** 80 tonnes had beached itself.
Which species of whale could this be?

Answer

☐ Right whale ☐ Humpback whale

☐ Fin whale ☐ Blue whale

4 How long was Meg's first rest?

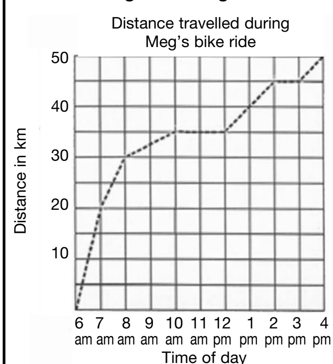


Answer

☐ 1 hour ☐ 2 hours

☐ 3 hours ☐ 4 hours

4 How long was Meg's first rest?



Answer

☐ 1 hour ☐ 2 hours

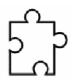
☐ 3 hours ☐ 4 hours

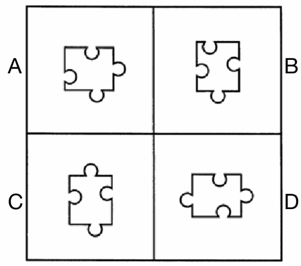
<table>
<tr><th>Standard item</th><th>Modified item</th></tr>
<tr><td>

5

Luke was using a puzzle.

Where would this part fit?





Answer

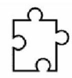
☐ A ☐ B ☐ C ☐ D

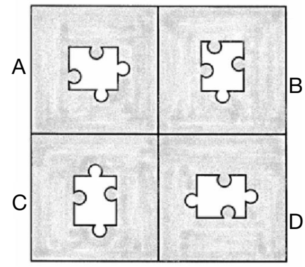
</td><td>

5

Luke was using a puzzle.

Where would this part fit?



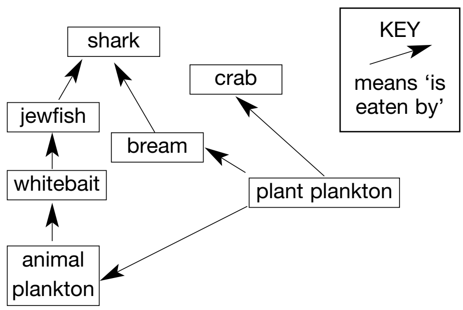


Answer

☐ A ☐ B ☐ C ☐ D

</td></tr>
<tr><td>

6 A SIMPLE FOOD WEB



The animals in this food web only eat what is shown.
If all the animal plankton die which of the following will also die?

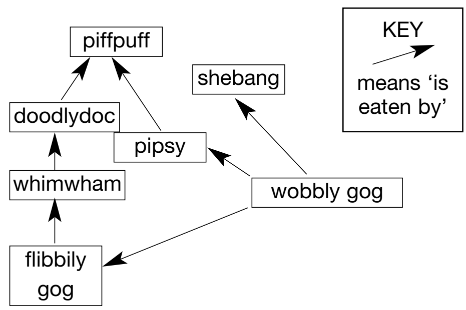
Answer

☐ crabs ☐ jewfish

☐ sharks ☐ plant plankton

</td><td>

6 A SIMPLE FOOD WEB



The animals in this food web only eat what is shown.
If all the flibbily gogs die which of the following will also die?

Answer

☐ shebangs ☐ doodlydocs

☐ piffpuffs ☐ wobbly gogs

</td></tr>
</table>

Acknowledgement

An earlier version of this paper was presented at the 10th annual Korean Institute of Curriculum and Evaluation (KICE) conference; Seoul, Korea, September 2008.

Authors

Professor Tom Lowrie is the Director of the Research Institute for Professional Practice, Learning and Education (RIPPLE) at Charles Sturt University.
Email: tlowrie@csu.edu.au

Professor Carmel Diezmann is the Assistant Dean Research in the Faculty of Education at the Queensland University of Technology, and is responsible for the leadership and management of the Centre for Learning Innovation.