Standardised Assessment in Mathematics: The Tale of Two Items

Tracy Logan
Charles Sturt University
<tlogan@csu.edu.au>

Jane Greenlees
Charles Sturt University
<jgreenlees@csu.edu.au>

This paper describes the sense making of 11- 12- year-olds as they interpret two mathematics items which include graphics. In particular, it outlines the changed behaviour (and performance) of students when solving items when slight modifications were made to the graphic or the mathematical language. The results show that performance increased when the graphic was modified but diminished when the language was modified. Implications include the need for test designers to carefully consider the graphic embedded within assessment items.

The capacity to interpret and decode graphs requires the problem solver to use spatial information to make sense of nonspatial relationships and concepts (Gattis & Holyoak, 1996). In mathematics contexts, these relationships and concepts are associated with mathematics literacy and content/context information. Although graphics are often considered "one of the simplest symbolic systems for interpreting information on the relationship between two or more sources" (Parmar & Signer, 2005, p. 250) primary students often find such representations overloaded with information and therefore difficult to decode (Lowrie & Diezmann, 2007). Moreover, graphs can become ineffective if too much information or too little information is presented (Kosslyn, 2006). If the graphics associated with testing items are not well designed, it is unlikely that results (outcomes) will be a reliable reflection of student understanding.

Research Design and Methods

This investigation builds upon the work of a three-year longitudinal study in which we monitored the development of primary students' ability to decode test items with high graphical content. This study focuses on items that were modified in relation to either the graphic or the mathematical language. The aims of this component of the study were to:

1. Ascertain student performance on graphics items and determine which elements of an item influence performance; and
2. Identify the sense-making that led to success on the items.

The Participants

Forty Grade 6 students (aged 11-12 years) from three regional NSW schools (one Government and two non-Government) took part in this study. The participants were from varying socioeconomic and academic backgrounds and were not involved in any treatment program throughout the study. These participants were accustomed to interpreting and solving items of this nature since they were part of the larger study (see Diezmann & Lowrie, this symposium).

Data Collection and Analysis

The following section describes the two phases of the project.

Phase 1. This phase formed part of a larger study which traced these participants' sense making over a 3-year period. The research team conducted structured, in-depth, interviews where students had an opportunity to verbalise and justify the processes they used to complete items from the Graphical Languages in Mathematics test (Diezmann & Lowrie, this symposium).

Phase 2. From an analysis of the interview data, students' responses were coded to ascertain the problem-solving processes students used to solve the respective items. Once these data were collated, consistent patterns in student responses were sought. These patterns were associated with students' interpretation of the item but also elements of the item which included the graphic and the mathematical language embedded within the item. As a result it was evident that these elements had a significant impact on the way in which the participants interpreted and solved the items. These items were slightly redesigned and the students were re-interviewed and asked to solve the modified items.

Results

In this paper, we focus on the two items (see Appendix) that had the highest negative change and the highest positive change between the first and second interviews. We discuss the change in the processing and sense-making students undertook as they solved the modified items.

After initial analysis of the first interview, it was evident that students employed similar (or at least consistent) strategies to make sense of the respective items. Therefore coding was developed in order to capture the patterning of incorrect responses that took place as problem solving occurred. Although the strategies identified were relatively generic it was certainly the case that some of these approaches were more likely to be employed for particular items. Of the 18 incorrect responses to Item 1 (The Whale Item), 15 of the students' responses were coded as—*did not consider* ***i****nformation on* ***g****raphic*. With respect to Item 2 (the Line Graph), 31 of the 35 incorrect responses were coded as—***o****verly influenced by* ***i****rrelevant information or pictures embedded in the graphic*.

Based on students' performance on Test A (original) and Test B (modified), these two items were identified as having the largest change in performance. Although the effect size (measured by Cohen's *d*) for the Whale Item was relatively small, it had the highest negative trend (a diminished performance on the modified test). While the Line Graph (with a large effect size) had the highest positive trend (an increased performance on the modified test). Table 1 highlights these results.

Table 1

Students' Performances across Test A and Test B

	Whale item		**Line graph item**	
	A	**B**	**A**	**B**
% Correct	60	53	22	60
Effect size (Cohen's *d*)	-.14		.83	

The Whale Item

The majority of students chose one of two "plausible" solutions in Test A (see Appendix). Students who chose the correct response (Fin whale) looked at length as an exact measurement and the mass as an approximate measurement. Those who answered the item incorrectly (Right whale) still used an effective strategy but allocated exact measurements to both the length and the mass. By highlighting the word *approximately* in Test B we hoped to bring to students' attention the second variable, mass and that it was an approximate measurement, thus eliminating one of the two options they were unsure about. However, it was found in Test B that the students who had initially solved the item correctly now became distracted by the bolded word and based their answer on the mass measurement, looking for the exact 80 tonnes and assigning the approximation to the length, the reverse of what occurred in Test A.

> I had a feeling it was the Right [whale] so I looked at mass and it said 80 tonnes. It was closer to 80 to me and the length kind of shows 25m [Tommy].

The modification of one word highlights the ease with which students can misinterpret the language in an assessment item. While the graphic itself did not change, the students' interpretation of the graphic was very closely linked to their understanding of the language and how that information applied to the graph. Adams (2003) found that students often miss important information focusing on key words without reading the entire question. We therefore envisaged that the students' performance would have improved on the modified item as we were drawing attention to an aspect of the question that was initially missed.

The Line Graph Item

On the line graph item in Test A, the performance of students was very poor (22% correct) (see Appendix). It became apparent that the appearance of dots on the line graph at various intervals along the line was being interpreted as a stopping point.

It has 6am and 7am, so that took 1 hour until she had a rest. Because it's a line graph the circle/dot is like a rest and it tells you how long she rode [Alex].

These dots were removed in Test B (see Appendix) to give the students the opportunity to actually read the line graph without being distracted by the dots. In Test B, students went from having a simplistic understanding of a line graph—being able to read the axes but being unable to interpret the line—to being able to incorporate all elements of the graphic. For example in Test A Rebecca responded, "I chose 1 hour because she started at 6am and she stopped at 7am because here it has a dot where it was a new hour". Whereas in Test B she explained, "I chose two hours because on the graph it keeps on going up until she gets from 10am to 12pm and then it just goes straight so she's not moving any distance which means she must have stopped". This change in Rebecca's thinking suggests that the visual features of the graph can affect children's interpretation of graphical items (Gattis, 2002) and the important role the format of the graph plays in students' comprehension and reasoning processes (Carpenter & Shah, 1998).

Conclusions and Implications

Information graphics have become increasingly important in representing, organising and analysing information and consequently the presence of graphics is now more evident in syllabus documents. The prevalence of such representation in curricula is in turn reflected in assessment practices—and particularly standardised instruments. This study provides insights into the impact that the graphical elements and associated language have on student understandings and performance. These elements are both influential but we found that the design of graphical items can be enhanced (and thus become a more reliable indication of performance) if more attention is paid to the design of the graphic. Such implications are particularly relevant at a time when national testing is becoming increasingly influential in mathematics education research and classroom practice.

Acknowledgments. The Graphic Languages in Mathematics project was funded by the Australian Research Council (#DP0453366). Special thanks to Tom Lowrie and Carmel Diezmann for their contributions.

References

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

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.

Gattis, M. (2002). Structure mapping in spatial reasoning. *Cognitive Development, 17*, 1157-1183.

Gattis, M., & Holyoak, K.J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 231-239.

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

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

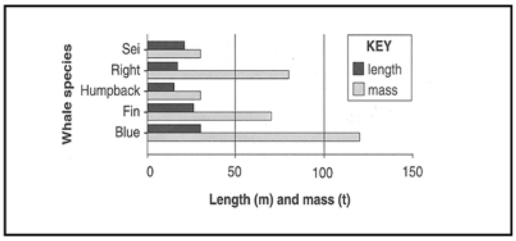
Parmar, R. S., & Signer, B.R. (2005). Sources of error in constructing and interpreting graphs: A study of fourth-and fifth-grade students with LD. *Journal of Learning Disabilities, 38*, 250-261.

Appendix: Standard and Modified Test Items

<table>
<tr><th>Test A Item</th><th>Test B (Modified) Item</th></tr>
<tr><td>

1

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?

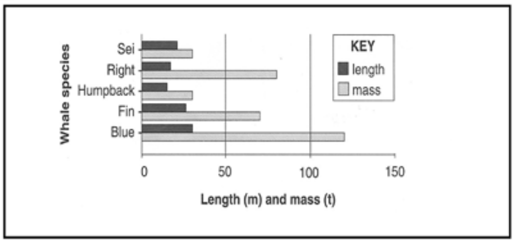
Answer

- ▢ Right whale
- ▢ Humpback whale
- ▢ Fin whale
- ▢ Blue whale

</td><td>

1

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?

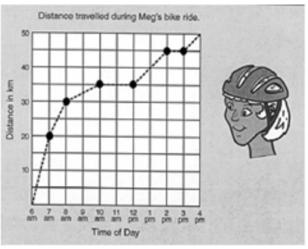
Answer

- ▢ Right whale
- ▢ Humpback whale
- ▢ Fin whale
- ▢ Blue whale

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

2

How long was Meg's first rest?



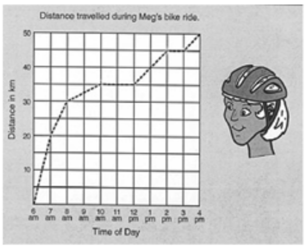
Answer

- ▢ 1 hour
- ▢ 2 hours
- ▢ 3 hours
- ▢ 4 hours

</td><td>

2

How long was Meg's first rest?



Answer

- ▢ 1 hour
- ▢ 2 hours
- ▢ 3 hours
- ▢ 4 hours

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