THE INFLUENCE OF GRAPHICS IN MATHEMATICS TEST ITEM DESIGN

Jane Greenlees[1], Tracy Logan[1,2]

[1]Charles Sturt University, Australia, [2]University of Canberra, Australia

This study investigated the performance and reasoning of 143 Australian students who completed mathematics tasks sourced from their national test. Specifically, this study examined changed student performance and reasoning on items where the graphic component was modified. The results of the study revealed significant performance differences between the original and modified items and provided insight into how these modifications influenced student reasoning.

GRAPHICS IN TEST ITEMS

The design of mathematics test items has received heightened attention recently. For example, the United States National Mathematics Advisory Panel (2008) made the following recommendation in regard to items used in both national and state achievement tests:

> More research is needed on test item design features and how they influence the measurement of the knowledge, skills, and abilities that students use when solving mathematics problems on achievement tests. (p. 61)

In light of such advice, this study aimed to outline what semi-structured interviews and survey data revealed about the influence of a graphic in the design of numeracy test items and the impact these graphics had on student performance and reasoning.

In deconstructing mathematics test items, Lowrie, Diezmann and Logan (2012) found that, typically, many assessment items consisted of three elements that organised mathematical information: text, symbols, and graphics. This paper will focus on the graphic component of test items. Bertin (1967/1983) defined graphics as visual representations for "storing, understanding and communicating essential information" (p. 2). Within the context of this study, graphics refers to any diagram, pictorial representation or graph used within a test item. The graphics in these items can be classified under two distinct categories, contextual and information. Contextual graphics are used for illustrative purposes, usually to provide a context for the written text. In contrast, an information graphic presents mathematical information in a visual-spatial form that supplements the text and symbols and is essential for task solution (Diezmann & Lowrie, 2008).

According to Lowe and Promono (2006), "test and graphic have long been combined in various ways to provide complementary sources of information and on a wide variety of topics" (p.22). This dual use of text and graphics has resulted in cognitive load theory (CLT) becoming more prominent within assessment design and how the

2014. In Oesterle, S., Liljedahl, P., Nicol, C., & Allan, D. (Eds.) Proceedings of the Joint Meeting of PME 38 and PME-NA 36,Vol. 3, pp. 209-216. Vancouver, Canada: PME.

components within an item may impact on working memory (Sweller, 1999). CLT was examined with respect to multimedia learning and the impact of inappropriate and unnecessary graphics. Within this research Mayer and Moreno (2002) identified four design principles to aid students in learning more deeply and preventing the overloading of their visual and/or verbal working memories. These included the notion of contiguity, coherence, modality, and redundancy. Of particular relevance to mathematics assessment was the coherence and redundancy principles. With regard to the coherence principle, it was found that students "learn more deeply when they do not have to process extraneous words or sounds in verbal working memory or extra pictures in visual working memory" (Mayer & Moreno, 2002, pp. 116-117). Bobis, Sweller and Cooper (1993) explored the notion of the redundancy principle, finding a possible redundancy of some graphics within an item but also a necessity for graphics in other items in regards to cognitive load.

Within the Australian context, there has been research conducted on the influence of graphics in mathematics test items on student reasoning (e.g., Diezmann & Lowrie, 2012; Logan & Greenlees, 2008). Much of the findings highlight the difficulty students have interpreting and decoding the graphic presented in the item. Lowrie and Diezmann (2009) argued that decoding graphics is a skill that is seldom taught and that primary school-aged students often find such representations overloaded with information and therefore difficult to interpret. Indeed, they found that test item design had considerable impact on how students solved tasks and that many errors:

> involved students not considering information in the graphic, being overly influenced by information (often irrelevant information) in the graphic, or not considering the connections between embedded graphical information and the textual and symbolic information (p. 153).

Schnotz (2002) also suggested that students will often pay only brief attention to the graphic, thinking that their general knowledge will suffice. However, students need to have a "schema-driven analysis" (p. 116) activated through explicit teaching before they can engage with the graphic content sufficiently. Therefore, it is critical that further research is undertaken on test item design in order to better understand the construction of these items and the processes required to decode them.

THEORETICAL FRAMEWORK

Utilising Lowrie, Diezmann and Logan's (2012) framework, items in the study were deconstructed according to the elements of text, symbols and graphics. This framework has been identified as a useful method to recognise the necessary elements that make up the composition of a graphical task and how the individual elements of a graphical task influenced students' reasoning. By recognising these three elements holistically and individually, Lowrie, Diezmann and Logan identified the influential role the graphic element played in students' sense making. Following on from these findings, this study focussed on the graphic element and specifically, the principles of coherence and redundancy in relation to contextual and information graphics.

RESEARCH DESIGN AND METHODS

This study is part of larger investigation that focused on the impact of test item design on student's performance and sense making. These design elements included the use of text, symbols and graphics. The 15 items used were sourced from the 2010 Year 5 National assessment and were selected based on their relevance to the particular design elements. An item modification process was undertaken, which resulted in the 15 items being modified according to the text, symbols or graphics. We ensured that variations to the respective elements of the standard items would achieve the same grade-level content whilst still allowing for variations in task design (Kettler, Elliott, & Beddow. 2009). The resulting item design modification produced three items with graphical modifications, which became the focus of this paper.

Due to the test items with modified graphics being administered only one week after the original test, we made minor modifications to the wording or contexts of the items. It was anticipated that students may remember answers they gave from the first interview and ignore the changes made to the items within the second interview. Therefore, one of the limitations of this process was that modified items could not be an exact match to the original items. For this paper, the focus was on modification to the graphic component of test items, both contextual and information graphics. The research questions for this paper were:

1. Does modification to the graphic component of test items impact on students' performance?
2. How does the graphic component of test items influence student's reasoning?

The Participants

The inquiry took place in four Australian primary schools in the state of New South Wales. The schools were diocesan Catholic primary schools for children aged 5-12 years. The students were in Year 5, and were aged 10 or 11 years. Altogether, 143 students were involved over a three-year period: 106 in the testing cohort and 37 in the interview cohort.

Data Collection and Analysis

The following section describes the four phases of the study over a four-week period.

Phase 1. The original 15-item test was administered to 106 students. It is important to note that the numeracy items from the national test were used as 'representative' mathematics items suitable for students in primary schools. (week 1)

Phase 2. Using the original 15-item test, one-to-one interviews were conducted with 37 students purposively selected by the teachers to be representative of the classes. The interviews provided the students with the opportunity to solve the 15 items and describe their thinking strategies and solutions. (week 2)

Phase 3. Using the interview data obtained in phase 2, the 15 test items were modified (see Figure 1 for item modification) according to three design elements that were particularly influential in the student’s solution process. The modified test was then administered to the 106 students. The cohort’s performance between the two tests was then analysed. (week 3)

Phase 4. The modified test was administered to the 37 interview participants to identify any changes in mathematical reasoning. Their responses were viewed as representative of the larger cohort. (week 4)

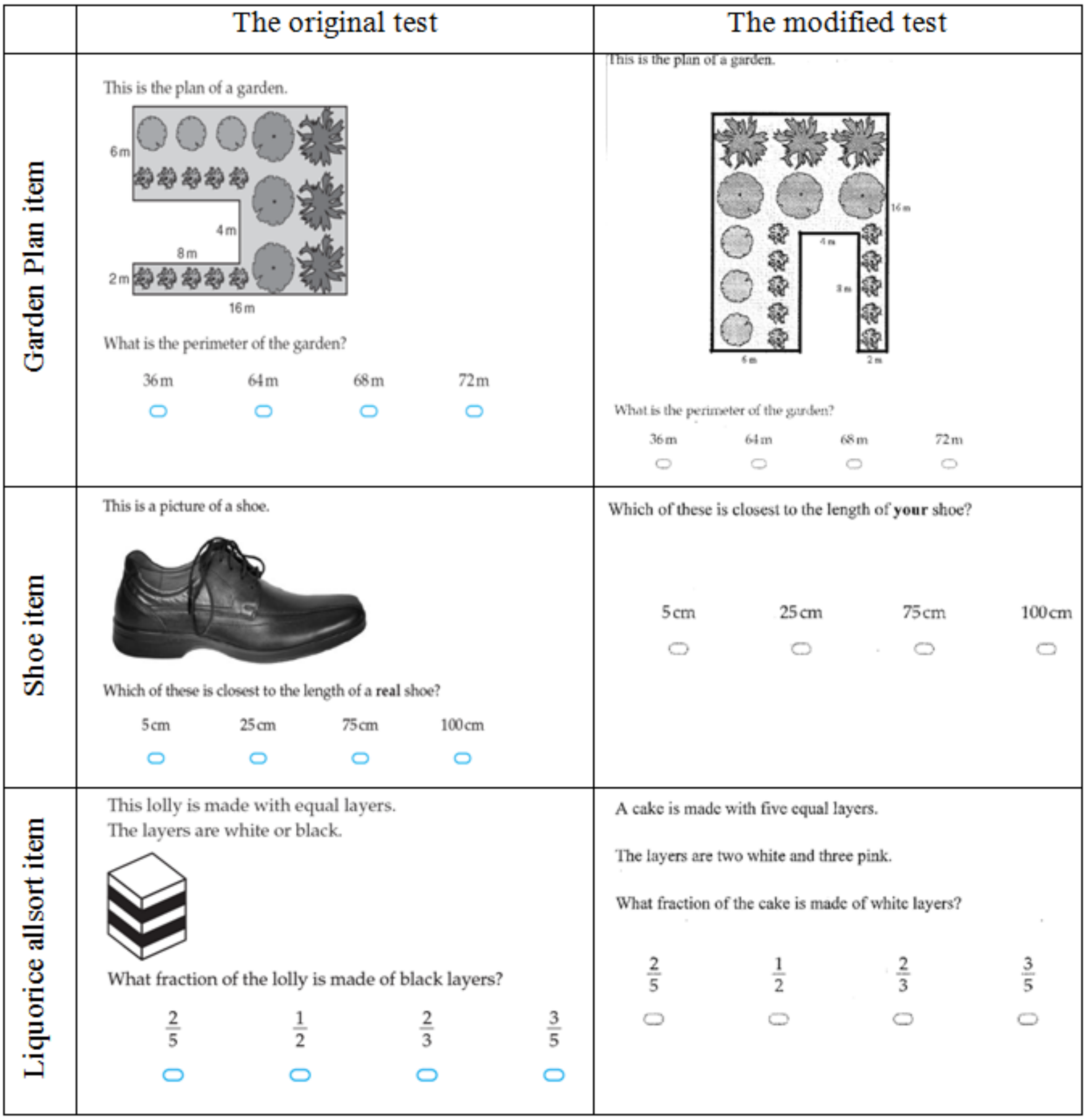

Figure 1: The original and modified tasks

RESULTS AND DISCUSSION

The Impact of Modifying Graphics on Students’ Performance

In this paper, we identified three items where the graphics were highly influential on students’ reasoning in the first phase of interviews. The analysis of variance indicated that there was a statistically significant difference for items when a graphic was repositioned as well as when a contextual or an information graphic was taken away (see Table 1). For the Garden plan and the Shoe items, student performance increased on the modified item. However, for the Liquorice allsort item, performance declined from 92% to 83% on the modified item. Noteworthy is the observation that performance decreased when the information graphic was removed. To better understand why these changes might have occurred, the interview data was analysed.

Item Name	Modification	Original test N = 143	Modified test N = 143	F (df 1,285) * $p \leq 0.05$
Garden plan	Repositioning of graphic	33%	47%	5.91* p = .016
Shoe	Contextual graphic removed	78%	88%	4.96* p = .027
Liquorice allsort	Information graphic removed	92%	83%	4.62* p = .032

Table 1: Modifications, Percentage Correct, and Univariate Analysis of Graphics Items

The influence on student reasoning

Repositioning graphic within the item: In the Garden plan item, the presentation and layout of the question was modified by centering and rotating the graphic to a more prominent position (see Figure 1). The analysis of the interview data revealed that the students seemed to overlook incorporating all sides into calculating the perimeter on the original item, possibly due to its close positioning to the question stem and answers. This was in spite of the fact that when asked the definition of a perimeter all students correctly described it as the distance all around the shape. An example of this can be seen in Elise's response to the original item when questioned on how she got the Answer A (36m).

Elise: I added them all together and then I did partners to 10. So I did 16 and 4 and 8 and 2 and that equalled 30 and I added the 6.

Int: What does perimeter mean?

Elise: The outside of an object.

By moving the graphic away from the clutter in the modified item, the interview carried out one week later revealed a heightened awareness to include all sides in the equation as evident in Elise's response (72cm):

Elise: I added all the ones and the ones where there wasn't any numbers like I knew that would be 8 because that's the same as that one and that would be 16 and for that one there was a gap down here so I put 6 and 4 together and that's 10 and then 2 for 12.

A number of students, who answered the item incorrectly in phase 2, reported similar processing on the modified item. This indicated that the format of the original item impacted on the visibility of the information necessary to solve the task. By placing the graphic too close to the question, the students may have failed to notice the requirement to include those sides that had no numerical value attributed to them. However, when all sides of the graphic were clearly visible, they could effectively determine their value and include them within the perimeter. Hence, the location of the graphic within an assessment item affected the coherence of the graphic in terms of its readability and impacted on students' ability to process and utilise the information contained within the question. We also acknowledge the potential benefits of rotating the graphic into a vertical position, given current research has suggested this is a preferred orientation (Giannouli, 2013).

Contextual graphic removed: The Shoe item was modified by taking away a contextual graphic (see Figure 1), meaning that the graphic did not contain information needed to solve the question. This modification was made because nearly one-third of the interview students made reference to the picture of the shoe despite its irrelevance to obtaining the answer in phase 2. It appeared that the picture of the shoe was distracting the children from analysing the question logically. The original item that students could not relate to directly, required them to estimate the length of a 'real' shoe by providing a picture of a shoe that was not to scale and ambiguous in nature. The modified item also required the students to estimate the length of a 'real' shoe but this time providing a more meaningful context by directing them to look at their own shoe.

For example, it could be perceived from her response to the original item that Mikayla did not have a sound understanding of measurement by considering a shoe to be close to 75cm. However, when investigated further it was revealed that there were aspects of the question that were negatively impacting on her mathematical reasoning.

Mikayla: Well 5cm is too small for a real shoe and 25cm is sort of a bit small too. 75cm would probably be the size of a real shoe and 100cm would be too big.

Int: So did the picture of the shoe help you work out your answer?

Mikayla: Yeah.

Int: How did it help you?

Mikayla: Well if it had been a picture of a baby shoe it would have been 5cm but because it was a picture of an adult shoe it was 75cm.

In contrast, the modified item revealed that Mikayla's measurement knowledge was in fact quite acceptable once the distraction was removed and a more accurate context created:

Mikayla: Because 5cm is too small to be my shoe and 75cm is too big and same as 100cm and 25cm was just about the right size.

Often contextual graphics are added as cues to the context or as possible forms of motivation and elaboration (Shimada & Kitajima, 2006). However, in this instance, the graphic was redundant for the purposes of problem solving and not necessary. Despite this, students were inappropriately attempting to include this in their problem solving strategies. For this reason test designers need to re-evaluate their good intentions of including such a graphic and the necessity of its inclusion.

Information graphic removed: The Liquorice allsort item was modified by excluding the information graphic and replacing it with written data (see Figure 1). The interview data suggested that the decrease in student performance on the modified item was due to their inability to visualise the cake using the information given. Many of the students were focusing on the numbers included in the question and were not focusing on the part-whole relationship of the fraction. Misunderstandings of the requirements of the question became more apparent in the modified version. This was particularly evident in Kyle's responses to both items:

Int: [original item] How did you get your answer?

Kyle: Well at first I thought it was 2 out of 3 because there's 2 black layers but then I looked at the lolly and saw there was 5 layers not 3 so then I chose 2 out of 5.

Int: [modified item] How did you get your answer?

Kyle: I chose 2 out of 3 because there's 2 white layers and 3 pink and that's how I worked it out.

In his response to the original item, the information graphic actually prompted Kyle to think about the part-whole relationship. It was therefore more effective to represent this type of information to students as a graphic rather than a word problem. Removing the lolly graphic could have resulted in heightening the cognitive load placed on the students during the problem solving process, as they now needed to visualise what the cake looked like. Because of this, the information graphic was not redundant but rather a necessary component of the task. However, we acknowledge that students of this age should be able to generate their own representations also.

CONCLUSION

The research findings reinforced the need for further analysis and investigation into the different components of mathematics assessment items, in particular, the graphics. This includes a more comprehensive understanding about the differentiation between redundant graphics that are unnecessary to students and those that are not. It may be the case that contextual graphics have no role to play in high-stakes testing since such graphics are by definition contextual and therefore not necessary to the task at hand. However, the use of information graphics actually lightened the cognitive load for the students. Another consideration is the placement and layout of the graphic within an item and the impact this may have on the coherence and visibility of information available to the students. It could be the case that the inclusion of more "white space" and orientation of the graphic is influential in student performance and reasoning. These findings highlight the effect the slight change in test item design can have on students' understanding and reasoning.

References

Bertin, J. (1983). *Semiology of graphics*. (W. J. Ber, Trans.). Madison, WI: The University of Wisconsin Press. (Original work published 1967)

Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary-school geometry task. *Learning & Instruction, 3*, 1-21.

Diezmann, C. M., & Lowrie, T. (2012). Learning to think spatially: What do students 'see' in numeracy test items? *International Journal of Science and Mathematics Education*, *10*, 1469-1490.

Giannouli, V. (2013). Visual symmetry perception. *Encephalos*, *50*, 31-42.

Kettler, R. J., Elliott, S. N., & Beddow, P. A. (2009). Modifying achievement test items: A theory-guided and data-based approach for better measurement of what students with disabilities know. *Peabody Journal of Education*, *84*(4), 529-551.

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* (Vol. 2, pp. 655-658). Brisbane, QLD: MERGA.

Lowe, R. K., & Promono, H. (2006). Using graphics to support comprehension of dynamic information in texts. *Information Design Journal (IDJ*), *14*(1), 22-34.

Lowrie, T., & Diezmann, C. M. (2009). National numeracy tests: A graphic tells a thousand words. *Australian Journal of Education*, *53*(2), 141-158.

Lowrie, T., Diezmann, C. M., & Logan, T. (2012). A framework for mathematics graphical tasks: The influence of the graphic element on student sense making. *Mathematics Education Research Journal*, *24*(2), 169-187.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, *12*, 107-119.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: US department of Education. Retrieved from http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, *14*(1), 101-120.

Shimada, H., & Kitajima. (2006). Why do illustrations promote text comprehension? Motivation effect and elaboration effect. In *Proceedings of 5th International Conference of the Cognitive Science* (pp. 185-186). Red Hook, NY: Curran Associates Inc.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, VIC: The Australian Council for Educational Research.