Diezmann, Carmel M. and Lowrie, Tom (2009) ***An instrument for assessing primary students' knowledge of information graphics in mathematics***.
Assessment in Education : Principles, Policy and Practice, 16(2). pp. 131-147.

An Instrument for Assessing Primary Students' Knowledge of Information Graphics in Mathematics

Carmel M. Diezmann*[a] & Tom J. Lowrie[b]

[a]Queensland University of Technology, Australia

[b]Charles Sturt University, Australia

Abstract

Information graphics have become increasingly important in representing, organising and analysing information in a technological age. In classroom contexts, information graphics are typically associated with graphs, maps and number lines. However, all students need to become competent with the broad range of graphics that they will encounter in mathematical situations. This paper provides a rationale for creating a test to measure students' knowledge of graphics. This instrument can be used in mass testing and individual (in-depth) situations. Our analysis of the utility of this instrument informs policy and practice. The results provide an appreciation of the relative difficulty of different information graphics; and provide the capacity to benchmark information about students' knowledge of graphics. The implications for practice include the need to support the development of students' knowledge of graphics, the existence of gender differences, the role of cross-curriculum applications in learning about graphics, and the need to explicate the links among graphics.

Keywords: Information graphics, Mathematics assessment, Numeracy tests, Test construction, Primary students

*Corresponding author. School of Mathematics, Science and Technology Education, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland, 4001, Australia.
Email: c.diezmann@qut.edu.au

Introduction

Worldwide there has been a strong emphasis on the need to develop a mathematically proficient population who can cope with the mathematical demands of everyday life. In some countries, such as Australia, this everyday mathematical competence is referred to as "numeracy" and those who possess this capability are argued to be numerate. For example, "To be numerate is to use mathematics effectively to meet the general demands of life at home, in paid work, and for participation in community and civic life" (Department of Employment, Education, 1997, 15). In other countries, similar mathematical proficiency is referred to as "quantitative literacy" or "mathematical literacy". Here, we purposively use the terms "numerate" and "numeracy" to signal that we are discussing more than quantity as might be inferred from the term "quantitative literacy". Additionally, we avoid the term "mathematical literacy" because later we discuss literacy in relation to the linguistic demand of text items. Furthermore, the term "numeracy" is commonplace in Australia, which is the context for our study.

Although numeracy involves mathematical proficiency, it is less abstract than mathematics and has immediate relevance in the lives of students (Steen, 2001):

> Mathematics climbs the ladder of abstraction to see, from sufficient height, common patterns in seemingly different things. Abstraction is what gives mathematics its power; it is what enables methods derived in one context to be applied in others. But abstraction is not the focus of numeracy. Instead, numeracy clings to specifics, marshaling all relevant aspects of setting and context to reach conclusions. To enable students to become numerate, teachers must encourage them to see and use mathematics in everything they do. Numeracy is driven by issues that are important to people in their lives and work, not by future needs of the few who may make professional use of mathematics or statistics. (17-18)

Due to the importance of everyday relevance in numeracy, in addition to focusing on mathematical content (International Association for the Evaluation of Educational Achievement [IEA], 2003), numeracy tests typically embed mathematical content within authentic contexts. Additionally, numeracy tests at least to some extent attempt to minimise the linguistic demand of items (e.g., Shaftel, Belton-Kocher, Glasnapp, and Poggio, 2005; Shorrocks-Taylor, and Hargreaves, 2000). However, what is important but either not reported or not considered is the graphical component of numeracy items in these tests. This 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 to inform instructional practice (Kulm, Wilson, and Kitchen, 2005). Following a brief overview of mathematical graphics, we present a rationale for focusing on graphics in numeracy testing, describe the construction of a graphically-oriented test for primary students, and report preliminary results of the use of this instrument. Note, our use of the term graphics here precludes any visual representations that have a purely decorative function.

Graphical Languages in Mathematics

Information graphics is a burgeoning field which has widespread applicability for the management, communication, and analysis of information across numerous disciplines (Harris, 1996). In mathematics, graphics can be categorised into six types of graphical languages, which represent mathematical relationships among perceptual elements via particular encoding techniques (Mackinlay, 1999) (Table 1). These perceptual elements are position, length, angle, slope, area, volume, density, colour saturation, colour hue, texture, connection, containment, and shape (Cleveland and McGill, 1984). Examples of the encoding techniques that are employed with these elements include the placement of marks on an axis (e.g., a number line), the positioning of information between two axis (e.g., a bar graph), and the spatial location of marks (e.g., maps).

Insert Table 1 about here

Although examples of graphical language items vary in structure and surface detail, they all utilise a visual-spatial format. Hence, spatial ability, which is a composite of abilities and includes mental rotation, spatial perception, and spatial visualisation (Voyer, Voyer, and Bryden, 1995), is fundamental to decoding graphics. Students' spatial ability may be affected by their capacity to process visual information (e.g., Raven, 1998), any visual perception or processing problems (e.g., Zangemeister and Steihl, 1995), and their gender at particular ages on specific tasks. A meta-analysis of spatial ability and gender revealed a difference in favour of males in under 13-year-old students on mental rotation tests (Voyer et al., 1995). However, no other gender differences were reported. Thus, within a class there is likely to be some innate differences in students' ability to decode graphics.

The Impact of Graphics on Student's Mathematical Performance

Graphics or visual representations impact on mathematical performance in two ways. Firstly, some students find various graphics difficult to interpret or decode. For example, on the National Assessment of Educational Progress [NAEP], many fourth graders had difficulty reasoning from a bar graph (National Center for Education Statistics [NCES], n.d.) and using a scale (NCES, n.d.) (Table 2). These students' success on the scale item was no better than chance accuracy (1 out of 4, 25%). Although eighth graders outperformed fourth graders on both these items, many older students also had difficulty with graphics (Table 2). The performance differences between the bar graph item and the scale item at each grade level indicate variance in the difficulty level of particular graphics. Differences in the relative difficulties of some graphics and increased success with age on a limited range of graphics have been reported by others (e.g., Wainer, 1980). Secondly, students' performance is impacted by the type of graphic that is used. Baker et al. (2001) reported substantial variance in eighth and ninth graders' ability to interpret informationally equivalent graphics with students' comparative success rates of 95% on a histogram, 56% on a scatterplot, and 17% on a stem-and-leaf plot. They argued that this performance variance was due to students' transfer of knowledge about bar graphs to the other three graphics, and that although histograms and scatterplots share surface features with bar graphs, stem-and-leaf plot vary at the surface level from bar graphs. Thus, in numeracy tests the graphics in use are likely to impact on students' performance on particular items.

Insert Table 2 about here

In Australia, we have two further reasons to explore the role of graphics in mathematics testing based on the TIMSS Results for Grade 4 International Benchmarks on items related to visual representation. Firstly, the TIMSS data revealed that in 2003 a similar percentage of Australian fourth graders achieved each of the International Benchmarks as in 1995 (Mullis, Martin, Gonzalez, and Chrostowski, 2004) (Table 3). This result is disappointing considering that numeracy has been a national priority for a many years and considerable funding has been provided to support the improvement of numeracy (Department of Education, Science and Training, n. d.). Secondly, a lower percentage of Australian fourth graders met the Advanced International and High International

Benchmarks than the international average in both 1995 and 2003. Both the lower proportion of students exhibiting high level capability and the continuation of this trend are a concern.

Insert Table 3 about here

The Rationale for a Test of Graphical Languages in Mathematics

There are two major reasons for constructing and implementing a test of graphical languages. Firstly, despite the importance of graphics, there appears to be no existing instrument to measure students' knowledge of them. The most recent study on the development of primary students' ability to decode information graphics appears to be a quantitative study by Wainer (1980). While Wainer's study provided evidence of developmental differences across the primary years and the relative difficulties of some graphics, the contemporary applicability of this study is questionable due to the limited graphics used, the sole reliance on mass-testing, and the possibility of time-period effects in the cross-sectional design (Willett, Singer, and Martin, 1998). We agree with Hiebert (1999) who argues strongly that without an adequate literature base, it is unrealistic to expect that teachers can enhance mathematics achievement because they lack the requisite wherewithal. Thus, the development of an instrument to measure graphical knowledge is a proactive and essential step towards making implicit mathematical knowledge explicit, an important contemporary research priority in mathematics education (Ball, 2004).

Secondly, knowledge of students' performance on graphic languages will inform the construct validity of numeracy tests, and the utility of trend data. For example, an analysis of a state-based Year 5 numeracy test (Queensland School Curriculum Council [QSCC], 2000) revealed that the test included items from five of the six graphic languages with the exception of the connection-languages (e.g., trees, networks). Given that connection languages include general-purpose diagrams that represent a multitude of problem structures (Diezmann, 1999; Novick, 2001), the validity of including two Opposed Position items (i.e., graphs) on the test and no Connection item is questionable. In fact, no Connection items were included in either the Year 3 or Year 7 tests in the same year (QSCC, 2000a, 2000c). Furthermore, the utility of numeracy tests for comparative purposes may be limited. As the relative difficulty of graphic language items is unknown, it cannot be determined how a change in the graphic items on a year level test from one year to the next would impact on overall test difficulty. This lack of information impacts on the interpretation of trend data in relation to a particular cohort's performance in a given year.

The Graphical Languages in Mathematics Test

The Graphical Languages in Mathematics [GLIM] Test is a 36-item multiple choice test that was designed to investigate students' performance on each of the six graphical languages. This instrument comprises six items that are graduated in difficulty for each of the six graphic languages. The test was developed from a bank of 58 graphically-oriented items. These items were selected from published state, national and international tests that have been administered to students in their final three years of primary school or to similarly aged students (e.g., Educational Testing Centre, 2000; QSCC, 2000b, 2000c). Due to the limited Connection items in existing mathematics tests, content free Connection items were sought from science tests (e.g., Educational Testing Centre, 2001). The tasks in the item bank were variously trialled with primary-aged children (N = 796) in order to select items that: (a) required substantial levels of graphical interpretation, (b) required minimal mathematics knowledge, (c) had low linguistic demand, (d) conformed to reliability and validity measures, and (e) varied in complexity. Our selection of items was also validated by two primary teachers. An example of an item from each subtest is shown in Appendix A.

The Implementation of the GLIM Test

The GLIM test was designed to be administered in both whole class and individual situations. The complete test (i.e., the 36 items) can be undertaken through mass testing procedures. This test can also be broken up into three sub instruments of varying difficulty which each contain two items of each graphical language type. In the present study, the GLIM test was implemented in both whole class and interview situations with the interview data comprising the students' responses to the easiest sub instrument. Firstly, a total of 217 9-to 10-year-old students completed the GLIM test in a mass testing situation. Students in this cohort were drawn from five schools in a large regional city. Here, we report on one implementation of the complete GLIM test. Secondly, an additional 67 9-to 10-year-old students from two metropolitan schools participated in individual interviews that contained the twelve easiest items from the instrument. We report on the pair of Axis items from these interviews. We argue that the preliminary analyses of the results of the mass testing and interviews provide an opportunity to analyse the utility of the instrument. Moreover, it provides opportunities to interrogate the data from the perspective of the students' sense making. In particular the extent to which the students' understanding of the task was related to content knowledge, numeracy, and graphicacy.

The GLIM Test and Mass Testing

Our analysis of the implementation of mass testing of the GLIM test with 217 students (9– 10-year-olds) yielded three outcomes. First, the analysis revealed that students were more successful in completing the Miscellaneous [= 4.24, S.D.= 1.39] and Map [= 4.06, S.D.= 1.29] item languages than the other four categories of graphics (Table 4). Recall, there were six examples of every type of graphic, hence, "6" is the maximum score. Although we would not suggest that such results indicate a hierarchy of complexity in relation to language type, the results are noteworthy since all items were derived from a "pool" of standardised questions from highly-reputable testing agencies. This instrument provides additional information to that usually gleamed from such standardised instruments because of the frequency of specific items (categorised in graphical languages) participants are required to solve. In these types of tests, it would be unusual to have more than two items from most of these languages. Consequently, patterns of performance cannot be analysed. We, therefore, would argue that our results are revealing. Interestingly, miscellaneous graphics and maps are explicitly addressed in other key learning areas as well as mathematics. By contrast, the mean score for the Opposed position category was considerably lower (15% less than the Miscellaneous category) despite the concentration of activities involving line charts, bar graphs and histograms in the curriculum and in numeracy tests. These results indicate differences in the difficulty level of various graphical languages for students of this age. Although some of this variance could be attributed to content, the particular graphical languages used are likely to be a major factor influencing students' performance (Baker et al., 2001).

Insert Table 4 about here

Second, the analysis highlighted some gender differences in students' decoding performance in relation to the six graphical languages. The mean scores for the male students were higher than that of the female students in all six categories. T-tests were conducted to determine whether there were statistically significant differences between the performances of males and females across the six graphical language categories. There were no statistically significant differences across the gender variable for five out of the six categories [Opposed position ($t = .001$, $p = .98$); Retinal list ($t = .84$, $p = .36$); Map ($t = 2.71$, $p = .10$); Connection ($t = 1.28$, $p = .26$); and Miscellaneous ($t = .15$, $p = .69$)].

There was, however, a statistically significant difference between the mean scores of the male and the female students in relation to the Axis graphical languages [Axis ($t = 12.2$, $p \leq .001$)]. A gender difference for Axis languages was unanticipated because no mental rotation was required and the students were under 13 (Voyer et al., 1995). However, these results are consistent with Hannula's (2003) findings of gender differences on a number line task (i.e., Axis language) in favour of boys for Finnish fifth (n = 1154) and seventh graders (n = 1525). Hannula's explanation that gender differences appeared to occur on tasks that were more difficult for students is inadequate for this study because Axis items were the third easiest of the six graphical languages (Table 3). However, gender differences with graphic items could be explained by children's experiences in out-of-school contexts (Diezmann, 2005; Lowrie and Diezmann, 2007).

Third, the analyses showed that the six graphical languages were all positively correlated with each other (Table 5). With the exception of the Axis-Opposed position correlation ($r = .15$, $p \leq .05$), all correlations were statistically significant at a $p \leq .01$ level. Nevertheless, even the strongest relationships [Connections-Maps ($r = .39$, $p \leq .01$) and Miscellaneous-Connections ($r = .41$, $p \leq .01$)] were only moderately correlated with each other despite the strong statistical significance. The Connections-Map and Miscellaneous-Connections correlations accounted for approximately 16% of the variance. The Miscellaneous and Connection categories had the strongest correlations with the other graphical languages (with all correlations stronger than $r = 0.27$). By contrast, the correlations between the Retinal list category and other language categories were somewhat weaker (most correlations less than 0.30). In most cases, the Retinal list items required the participants to consider graphical features including shape, size, saturation, texture, and orientation. Thus, decoding these graphics required an understanding of the use of perceptual elements to convey mathematical information. The weakest correlation was between the Axis and Opposed position languages. Although both of these languages use axes, they differ substantially at a structural level with information encoded in only one dimension in Axis items and in two dimensions in Opposed position items (Mackinlay, 1999). Although all items employed graphics and the mathematical content and linguistic demands of these items were minimised, it is noteworthy that the correlations between items within each graphical language were relatively weak. One plausible explanation for this result is the structural differences between each of the graphical languages (McKinlay, 1999).

Insert Table 5 about here

The GLIM Test and Individual Interviews

Our preliminary analyses revealed the relevance of the GLIM test for in-depth investigation of students' thinking. For example, in interviews with 67 students (aged 9-10 years) from two schools (Overton, n = 24; and Stanley, n = 43) on the two easiest axis items yielded information about the proportion of students who were successful on these tasks, the knowledge that led to successful use of the number line, and errors that resulted in unsuccessful use of the number line. The two selected Axis items are similar in that they focussed on the identification of unnumbered positions on a number line and dissimilar in that Item 1 and 2 focussed on whole numbers and decimals respectively. The students completed these two items during an individual interview and then explained their thinking. The interview data comprised students' multiple choice selections and the reasons they gave for their answers. The interviews were video- and audio-taped to facilitate analysis. These data were analysed to determine the proportion of successful students and the reasons for students' success or failure on these items.

Insert Figure 1 about here

On average between 80% and 91% of students were successful across tasks and schools with means higher for Item 1 than Item 2. The former focuses on whole number and the latter on decimals. There were two points of interest. First, the extent of the gender difference was unexpected. With the exception of Item 2 for Overton, there was between 9% and 13% performance difference in favour of males (Table 6). This result is consistent with the results of the larger cohort (N = 217) from the mass testing across the six Axis items (Lowrie and Diezmann, 2005). In that analysis, the Axis items were the only set of graphical languages that revealed a statistically significant difference between the mean scores of male and female students across the total set of any of the graphic languages [Axis ($t = 12.2$, $p \leq .001$)]. Second, there was a substantial performance difference across schools with Stanley (n = 43) outperforming Overton (n = 24) on both items. It is important to note that statistically significant differences between the performance of students across schools was limited to specific items within a language rather than across any “collective” (ie., the six items) language. Elsewhere (see Lowrie & Diezmann, 2005) we report that there was no differences between the performance of students’, on any of the six languages, across states—despite the fact that the syllabus documents in these states are quite different. In the present investigation, a plausible explanation for the higher performance of one school over another is the perceptual variability of the curriculum (e.g., Dienes, 1964; Moss and Case, 1999). Thus, at one school the classroom teachers may have used multiple representations of the same ideas to evoke more sophisticated understandings of the concept. Such results point to the need to have an instrument that includes a comprehensive range of items within each language since explicit teaching (and thus the timing of the teaching in relation to standardised testing) will have an impact on student performance.

Insert Table 6 about here

Successful students gave seven reasons for their selection of responses for either Item 1 or 2 (Table 7). These reasons can be grouped into two categories. The Measurement category consists of those reasons that indicate an understanding of the number line as a measurement model through explanations that refer to distance, proximity or reference points (i.e., CI, EP, LR, RS, RP). The Inappropriate category comprises explanations that focus solely on counting (CO) or guessing (GU). On Item 1 (whole numbers), all successful students from Overton and Stanley gave reasons from the Measurement category. However, on Item 2, there were substantial differences between Overton and Stanley. All successful students from Overton gave Measurement reasons. However, at Stanley, some students gave Measurement reasons (81.5%) and others gave Inappropriate reasons (18.5%). One interpretation of these results is that as the difficulty of the number line item increases from whole numbers to decimals some students were less able to provide an appropriate reason for their selection of response.

Insert Table 7 about here

Unsuccessful students made a range of errors in their use of the number line (Table 8). These errors are of two major types. Solution errors comprised difficulties with distance (FD), position (IP), counting (IC) or misreading the diagram (MD). The predominant Solution error across items and schools was inappropriate counting (IC). This error supports Fuson's (1984) concern that measurement foundation of the number line is overlooked by students and teachers. Explanation errors consist of guessing (GU) and vague answers (VA). Explanation is one of six fundamental cross-cultural mathematical practices (Bishop, 1991). Hence, to guess or give a vague answer is a general error because it indicates a lack of understanding of an important mathematical practice. Students need to be encouraged to provide adequate explanations because the communication of an explanation provides them with an opportunity to review and, if necessary, refine their mathematical thinking. For example, Shaun (a Stanley student) realised his error on Item 2 during his explanation for selecting D: “Because it’s the closest to number one and D might be 5 [meaning 1.5]. Oh because um the one is there and I *accidentally* put it down near this area and it should be one point three [1.3]” (emphasis added).

Insert Table 8 about here

Conclusions and Implications

The development and implementation of the GLIM instrument was an attempt to explore students’ knowledge of graphically-oriented items that are typically employed in numeracy tests. Our conclusions to date are that the GLIM test is beneficial for the following four reasons.

First, the analysis of the GLIM test can provide information about the nature and relative difficulty of the graphics. This information is useful for informing instruction, when designing classroom tests or interpreting numeracy test results containing graphics. It was apparent that the students found the Miscellaneous and Map items to be the easiest. By contrast, the Connection and Retinal list items were significantly more difficult for the students to solve. Interestingly, this challenges the very nature of the traditional teaching of mathematics since the items most successfully solved are often applicable across the curriculum rather than solely within the discipline. Consequently, the notion of numeracy across discipline areas should be encouraged and indeed fostered as the complexities within graphical languages can be explored and understood within authentic contexts. Theoretically, this mathematical variability enables a more comprehensive understanding of a particular idea after it has been encountered in various contexts (Dienes, 1964). Students of this age are exposed to Map languages in science (e.g., solar system), studies of society (e.g., population maps) and physical education (e.g., orienteering), and thus, through a range of activities experience maps across different representations and processing modes.

Second, the GLIM test provides benchmark information about students' knowledge of graphics at particular ages. The data presented here are drawn from 9- to 10-year-olds. However, there can be significant improvement in primary students' knowledge of graphics across a 12 month period (Lowrie and Diezmann, 2007).

Third, the results of the mass testing with GLIM and the use of the GLIM test items in individual interviews can inform educational practice in the following ways.

- Teachers need to be proactive in supporting the development of students' ability to decode information graphics. However, the provision of appropriate support may be challenging for teachers because they have difficulty identifying which types of graphics are easier or harder for students (Baker et al., 2001).
- Due to gender differences in performances on the Retinal list items, girls should receive strategic support in activities incorporating this language. This language incorporates graphics such as saturation (e.g., in graphical representations of population density).
- Learning opportunities should be broad and include graphical languages that are typically used outside formal mathematics contexts (i.e., Maps, Connections, Miscellaneous, Retinal

list) in addition to those explicitly incorporated into the mathematics syllabus (Axis, Opposed position) (see Diezmann, 2005 for a discussion of informal contexts).

- Where appropriate, explicit links between graphical languages should be addressed to support cognitive transfer. For example, the perceptual element of position is relevant to Axis, Opposed position and Map languages (Mackinlay, 1999).
- The informational content of graphics used for instructional purposes should be explicated to ensure that all students have access to embedded information. For example, in an Opposed position item, the X and Y axis are built on understandings of a single axis in the Axis language.
- The successful and unsuccessful students' responses indicate the importance of students' appreciation that the number line is a measurement rather than a counting model (Fuson, 1984). Thus, instruction needs to emphasise the linearity rather than cardinality of the model.
- The role of explanation as a metacognitive process needs to be highlighted. Explanation provides opportunities for students to share their thinking and through this process to reflect upon and refine their thinking.

Finally, the GLIM test itself can be used to assist teachers to analyze test items in terms of the graphical demands of test items as well as the mathematical and linguistic demands. This is a promising avenue for teacher professional learning because a degree of expertise is needed with graphical representation (e.g., Dreyfus and Eisenberg, 1990).

The utility of the GLIM instrument is far reaching, however, the categorization of the graphical languages remains problematic. We have concerns that require further investigation. By its very nature the Miscellaneous category comprises items that cannot be classified within the other homogeneous language categories. Interestingly, these miscellaneous items share some of the features of other graphical languages however they incorporate additional graphical techniques (e.g., the use of angles). As a consequence, the set of miscellaneous items may well be structurally dissimilar (e.g., calendar and pie chart). Nevertheless, we maintain that such “common” mathematical representations warrant inclusion in any graphics instrument.

Since student performance across these graphical languages is only weakly or moderately correlated it is important for teachers to explicitly address the specific attributes that are embedded within each language. For example, students need to understand the use of scale and keys on a map which is quite different from understanding the use of points on a number line (Axis graphical language). These and other conventions should be taught to enable the students to full access the embedded mathematical information. Hence, learners should be exposed to various items within a graphical category rather than concentrating on replications of many graphically similar items (e.g., Figure 1). In other words, it is important that learners appreciate the graphical encoding technique that is common to all items within each graphical language. Rather than using a teaching approach that presents a range of items from different graphical languages, it would be more advantageous to focus on the relevant conventions within a language to enable students to construct a transferable understanding of each language type. Novick (1990) argues that representational transfer has cognitive benefits for individuals because they can draw on their knowledge of a particular information graphic, to derive the solution to an item that uses the same type of graphic.

This instrument has provided an understanding of how students' make sense of graphically-oriented items. The findings suggest that the graphical component of many items used in numeracy tests impacts substantially on performance. The GLIM test, therefore, serves an important role in benchmarking students' graphics performance and assisting teachers to monitor students' progress and identify their strengths and weaknesses. If we are to achieve a numerate population, we need to be proactive and to educate students about the various types of graphics. Additionally, if we are to rely on performance data from numeracy tests to inform policy and practice in mathematics education, we need to understand the role that graphics play in students' performance on these tests.

Acknowledgements.

This research is supported by Australian Research Council funding (#DP0453366). Earlier versions of this paper were presented at the Mathematics Education Group of Australasia annual conference (Lowrie and Diezmann, 2005) and the Psychology of Mathematics Education annual conference (Diezmann and Lowrie, 2006). Special thanks to Lindy Sugars, Melissa Hannaford and Tracy Logan who assisted with the collection and interpretation of data for this project.

References

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2001) Toward a model of learning data representations, in: *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (Mahwah, NJ, Erlbaum), 45-50.

Ball, D. L. (Chair) (2004) *Mathematical proficiency for all students: toward a strategic research and development program in Mathematics Education. RAND Mathematics Study Panel.* http://rand.org/pubs/monograph_reports/MR1643/MR1643.pref.pdf (4 October 2007).

Bishop, A. J. (1991) *Mathematical enculturation: a cultural perspective on mathematics education* (Dordrecht, The Netherlands, Kluwer).

Blackwell, A. (Ed.), (2001) *Thinking with diagrams* (Dordrecht, The Netherlands, Kluwer).

Cleveland, W. S., & McGill, R. (1984) Graphical perception: theory, experimentation, and application to the development of graphical methods, *Journal of the American Statistical Association*, 79(387), 531-554.

Department of Employment, Education, Training and Youth Affairs (1997) *Numeracy =Everyone's Business.* The Report of the Numeracy Education Strategy Development Conference May 1997 (Adelaide, Australian Association of Mathematics Teachers).

Department of Education, Science and Training (n.d.) *National literacy and numeracy plan.* <http://www.dest.gov.au/sectors/school_education/policy_initiatives_reviews/key_issues/literacy_numeracy/national_literacy_and_numeracy_plan.htm> (20 August 2006).

Dienes, Z. P. (1964) *Building up mathematics* (2nd edn.) (London, Hutchison Educational).

Diezmann, C. M. (2005) The impact of primary students' informal experiences on their diagrammatic knowledge, in: M. Goos, C. Kanes, & R. Brown (Eds) *Proceedings of the Mathematics Education and Society* [MES] *Conference* (MES, Gold Coast), 144-154.

Diezmann, C. M., & Lowrie, T. (2006) Primary students' knowledge of and errors on number lines, in: P. Grootenboer, R. Zevenbergen, & M. Chinnapan (Eds) *Proceedings of the 29th Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1) (Sydney, Mathematics Education Group of Australasia), 171-178.

Dreyfus, T., & Eisenberg, T. (1990) On difficulties with diagrams: Theoretical issues, in: G. Booker, P. Cobb & T. N. de Mendicuti (Eds), *Proceedings of the 14th Annual Conference of the*

International Group for the Psychology of Mathematics Education [PME](Vol. 1) (Oaxtepec, Mexico, PME), 27-33.

Educational Testing Centre (2001) *Australian Schools Science Competition Year 6* (Sydney, University of New South Wales).

Eisenberg, T., & Dreyfus, T. (1986) On visual versus analytic thinking, in: *Proceedings of the 10th Annual Conference of the International Group for the Psychology of Mathematics Education Conference* (London, UK, PME), 153-158.

Fuson, K. (1984) More complexities in subtraction, *Journal for Research in Mathematics Education*, 15(3), 214-225.

Glasgow, J., Narayanan, N. H., & Karan, B. C. (Eds) (1995) *Diagrammatic reasoning* (Menlo Park, CA, AAI Press).

Hannula, M. S. (2003) Locating fractions on a number line, in: N A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds) *Proceedings of the 27th Annual Conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th Annual Conference of PME-NA* (Vol. 3) (Honolulu, USA, PME), 3-24.

Harris, R. L. (1996) *Information graphics: A comprehensive illustrated reference* (Atlanta, GA, Management Graphics).

Hiebert, J. (1999) Relationships between research and the NCTM Standards, *Journal for Research in Mathematics Education*, 30(1), 3-19.

International Association for the Evaluation of Educational Achievement (2003), *TIMSS Assessment Frameworks and Specifications (2nd Ed)* <http://timss.bc.edu/timss2003i/PDF/t03_af_book.pdf> (24 August 2006).

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

Larkin, J. H., & Simon, H. A. (1987) Why a diagram is (sometimes) worth ten thousand words, *Cognitive Science*, 11, 65-99.

Lowrie, R., & Diezmann, C. (2005) Fourth-grade students' performance on graphical languages in mathematics, in: H. L. Chick & J. L. Vincent (Eds) *Proceedings of the 30th Annual Conference of the International Group for the Psychology of Mathematics Education* (Vol 3) (Melbourne, PME), 265-272.

Lowrie, T., & Diezmann, C. M. (2007) Solving graphics problems: student performance in the junior grades. *The 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*, (San Francisco, CA, Morgan Kaufmann), 66-81.

Moss, J., & Case, R. (1999) Developing children's understanding of the rational numbers: A new model and an experimental curriculum, *Journal for Research in Mathematics Education*, 30(2), 122-147.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004) *TIMSS 2003 International Mathematics Report. Chapter 2.* <http://timss.bc.edu/PDF/t03_download/T03_M_Chap2.pdf> (accessed 13 February 2005).

National Center for Educational Statistics (n.d.) *NAEP Questions.* <http://nces.ed.gov/nationsreportcard/itmrls/search.asp> (10 January 2003)

Novick, L. R. (1990) Representational transfer in problem solving, *Psychological Science*, 1(2), 128-132.

Queensland School Curriculum Council (2000a) *2000 Queensland Year 3 test: Aspects of Numeracy* (Victoria, Australian Council for Educational Research).

Queensland School Curriculum Council (2000b) *2000 Queensland Year 5 test: Aspects of Numeracy* (Victoria, Australian Council for Educational Research).

Queensland School Curriculum Council (2000c) *2000 Queensland Year 7 test: Aspects of Numeracy* (Victoria, Australian Council for Educational Research).

Queensland School Curriculum Council (2001a) *2001 Queensland Year 3 test: Aspects of Numeracy* (Victoria, Australian Council for Educational Research).

Queensland School Curriculum Council (2001b) *2001 Queensland Year 5 test: Aspects of Numeracy* (Victoria, Australian Council for Educational Research).

Queensland School Curriculum Council (2002) *2002 Queensland Year 3 test: Aspects of Numeracy* (Victoria, Australian Council for Educational Research).

Raven, J. C. (1998) *Raven's progressive matrices* (London, Oxford Psychologists).

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2005) The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities, *Educational Assessment*, 10(4), 333–356.

Shorrocks-Taylor, D. & Hargreaves, M. (2000) Measuring the language demands of mathematics tests: the case of the statutory tests for 11-year-olds in England and Wales, *Assessment in Education: Principles, Policy & Practice*, 7(1), 39-60. <http://www.metapress.com.ezp01.library.qut.edu.au/media/acecbdavyr3tvw4u8g8h/contributions/g/x/x/q/gxxql29p0t2llg39.pdf> (24 August 2006).

Steen, L. A. (2001). The case for quantitative literacy, in *Mathematics and democracy: the case for quantitative literacy* (Princeton, NJ, National Council on Education and the Disciplines), 1-22.

Voyer D, Voyer S, & Bryden, M. P. (1995) Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables, *Psychological Bulletin*, 117(2), 250-270.

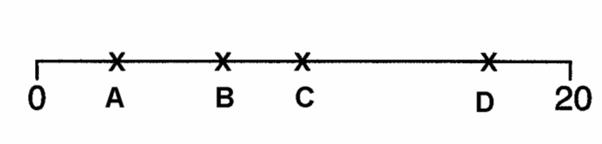
Wainer, H. (1980) A test of graphicacy in children, *Applied Psychological Measurement*, 4, 331-340.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998) The design and analysis of longitudinal studies of development and psychopathology in context: statistics models and methodological recommendations, *Development and psychopathology,* 10(2), 395-426.

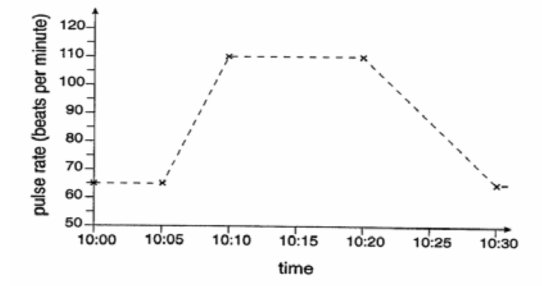
Zangemeister, W. H., & Steihl, H. S. (1995) *Visual attention and cognition* (Amsterdam, The Netherlands, Elsevier).

Appendix A[b]. Information Graphics: Six Graphical Languages Items

Estimate where you think 17 should go on this number line.



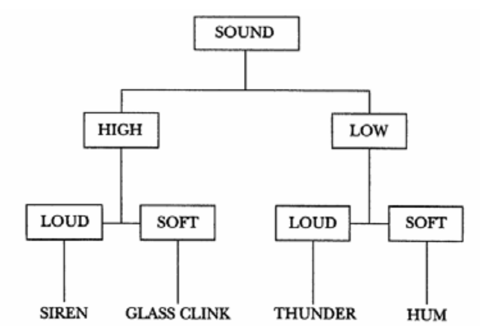
Axis Item (Adapted from QSCC, 2000a, 11)



This graph shows how Sam's pulse rate changed while she exercised. What is the difference between Sam's lowest and highest pulse rate in beats per minute?

Opposed position Item (Educational Testing Centre, 1999, 8)

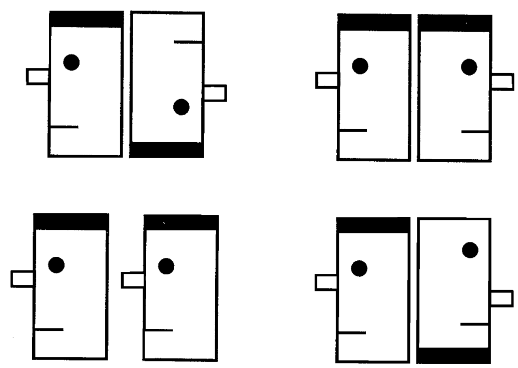
This flowchart shows a way to describe sounds.



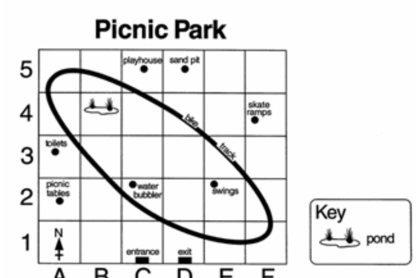
Which of the following describes a 'hum'?

Connection Item (Educational Testing Centre, 2001, 2)

Which two faces show a flip?



Retinal list Item (Adapted from QSCC, 2001a, 13)



Deb rides her bike along the bike track. What part of the park won't she ride through?

Map Item (QSCC, 2001b, 16)

Which date is 3 weeks before 29 May?

MAY 2003

Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Miscellaneous Item (QSCC, 2002, 9)

[b] The full GLIM test and the administration protocols are obtainable from the authors.

List of Tables

List of Figures

Table 1. An Overview of Graphical Languages

Graphical Languages	Examples	Encoding Technique
Axis Languages	Horizontal and vertical axes	A single-position encodes information by the placement of a mark on an axis.
Opposed position Languages	Line chart, bar chart, plot chart	Information is encoded by a marked set that is positioned between two axes.
Retinal list Languages	Graphics featuring colour, shape, size, saturation, texture, orientation	Retinal properties are used to encode information. These marks are not dependent on position.
Map Languages	Road map, topographic map	Information is encoded through the spatial location of the marks.
Connection Languages	Tree, acyclic graph, network	Information is encoded by a set of node objects with a set of link objects.
Miscellaneous Languages	Pie chart, Venn diagram	Information is encoded with a variety of additional graphical techniques (e.g., angle, containment).

(adapted from Mackinlay, 1999)

Table 2. Two Graphic Items from NAEPs Assessment for Grades 4 and 8

NAEP Item Description	Graphical Languages	Grade 4	Grade 8	Performance Difference
Reason from a bar graph	Opposed position	49% (n = 1622)	62% (n = 1759)	13%
Use a scale to find a distance between two points	Axis	24% (n = 36764)	39% (n = 30578)	15%

Table 3. Australian TIMSS Results for Grade 4 International Benchmarks on Visual Representation

Visual Representation Component	Advanced Benchmark: (Students) can organize, interpret, and represent data to solve problems.		High Benchmark: (Students) can interpret and use data in tables and graphs to solve problems.		Intermediate Benchmark: (Students) can read and interpret different representations of the same data.		Low Benchmark: (Students) can read information from simple bar graphs.	
Years	1995	2003	1995	2003	1995	2003	1995	2003
Australian result	6%	5%	27%	26%	61%	64%	86%	88%
International average	10%	10%	33%	36%	63%	69%	85%	88%

(Mullis et al., 2004)

Table 4. Means (and Standard Deviations) of Students' Performance on Examples of the Six Graphical Languages.

Graphical Languages	Total	Male (n = 115)	Female (n = 102)
Miscellaneous	4.24 (1.39)	4.27 (1.39)	4.20 (1.38)
Map	4.06 (1.29)	4.20 (1.28)	3.91 (1.29)
Axis	3.79 (1.19)	4.04 (1.13)	3.50 (1.20)
Opposed position	3.26 (1.71)	3.27 (1.30)	3.26 (1.02)
Connection	3.13 (1.28)	3.32 (1.32)	3.03 (1.23)
Retinal list	2.95 (1.38)	3.03 (1.41)	2.86 (1.34)

Table 5. Correlations between Student Performance on Examples of the Six Graphical Languages

Graphical Languages	Axis	Opposed position	Retinal list	Map	Connection
Axis					
Opposed position	.15*				
Retinal list	.26**	.19**			
Map	.33**	.24**	.29**		
Connection	.32**	.31**	.27**	.39**	
Miscellaneous	.37**	.34**	.32**	.31**	.41**

* p = .05 level ** p = .01 level

Table 6. Overview of Successful Student Performance on Axis Items

	Item 1			Item 2		
	Male	Female	Total	Male	Female	Total
Overton	90%	78.6%	83.3%	80%	78.6%	79.2%
Stanley	95.2%	86.3%	90.6%	95.2%	81.8%	88.3%

Table 7. Frequency of Successful Performance on Items 1 and 2

Reason	Example	Overton Item 1 (n = 20)	Overton Item 2 (n = 19)	Stanley Item 1 (n= 39)	Stanley Item 2 (n = 38)
	MEASUREMENT				
closest to an item (e.g., number) [CI]	*I chose D because it's closest to 20 and C is too far away.	75.0%	47.4%	66.7%	28.9%
estimating position [EP]	*I chose D because B is right, a bit far away from 20 and C is in the middle and I thought that would be about 10 and A would be too close to the 0 to be 17.	15.0%	10.6%	10.3%	21.1%
using a letter or number as a reference point [LR]	*I think it would be D because if half of that number line is number C and you would imagine the number line in the middle and then you can just look further up to 20 …	5.0%	21.1%	10.3%	21.1%
relative amount of space [RS]	*Because of the amount of space between each letter and the amount of space between D and 20.	5.0%	5.3%	2.6%	00.0%
to right or past or after 1 [RP]	* I chose C which is a bit to the right of the 1 and I thought it would be there 'cause it's closer to 1 than D is cause I think D would be 1.5.	0%	15.8%	0%	10.6%
	INAPPROPRIATE				
counting on or back [CO]	*I think it should go there (D) because it's next to 20 and it goes 19, 18 then 17.	00.0%	00.0%	10.3%	10.6%
Guessing [GU]	#Well, it was a toss up between C & D and I chose C.	0%	0%	0%	7.9%

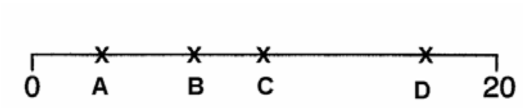
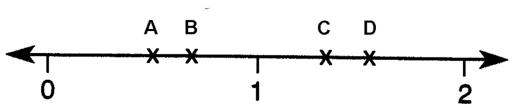
Key: *= Item 1 reason; # = Item 2 reason

Table 8. Frequency of Unsuccessful Performance on Items 1 and 2

Error	Example	Overton Item 1 (n = 4)	Overton Item 2 (n = 5)	Stanley Item 1 (n = 4)	Stanley Item 2 (n = 5)
SOLUTION					
focusing on distance (between, too far, too close) [FD]	*I chose C because that would be too close.	0%	0%	25%	0%
inappropriate counting back with each letter as a number [IC]	*I think it should be C because I reckon 19 would be about there. That would be 18 on D.	100%	0%	50%	20%
inaccurate position [IP]	#I did D because it's sort of closer to 1 than 2 and it's sort of in the middle as well to put 1.3.	0%	60.0%	0%	0%
misreading the diagram [MD]	#I said A because it's kind of half way in between the zero and the 1 and the B is a bit more like 4 so I just said A cause it's about half way.	0%	40.0%	0%	40%
EXPLANATION					
guessing [GU]	*I just guessed because I didn't really get it.	0%	0%	25.0%	0%
vague answer [VA]	#…because that would be one less and I thought that would be that too so I thought that would be good.	0%	0%	0%	40%

Key: *= Item 1 reason; # = Item 2 reason

Figure 1. Axis items.

1.[c]Estimate where you think 17 should go on this number line. 
2. [d]Estimate where you think 1.3 should go on this number line. 

[c] QSCC (2000a, p. 11)
[d] QSCC (2000b, p. 8)