Item Context Factors Affecting Students' Performance on Mathematics Items

Felipe Almuna Salgado
The University of Melbourne
<felipea@student.unimelb.edu.au>

Kaye Stacey
The University of Melbourne
< k.stacey@unimelb.edu.au >

This paper reports how the context in which a mathematics item is embedded impacts on students' performance. The performance of Year 10 students on four PISA items was compared with performance on variants with more familiar contexts. Performance was not better when they solved items with more familiar contexts but there was some evidence that items requiring the second-order use of context were more influenced by an alteration of context than items of first-order use. Recommendations for further study are included.

In the 1980s and 1990s advances in cognition and measurement theories led researchers to consider how assessment could be developed to ensure that meaningful aspects of students´ learning and thinking are captured. In this manner, different curricula and standard documents around the globe started to develop new forms of connectedness of mathematical content by focusing on mathematical modelling, contextualisation and problem solving (Stillman, 2000). The use of mathematical tasks in context was hoped to foster motivation for the learning of mathematics, and demonstrate and teach students to deal with the links and transfers between mathematics and the real world.

Mathematical tasks in context are particularly relevant in PISA, the Programme for International Student Assessment conducted since 2000 by the Organisation for Economic Co-operation and Development (OECD). Blum (2002) notes that since 2001 (the first time PISA results were published), researchers have debated the role of item context in students' performance; some researchers believe it makes items simpler but others consider that including contextual elements makes items significantly more challenging for students (de Lange, 2007). As a result, this paper investigates specifically how a systematic alteration of the context in which a PISA item is embedded impacts students' performance.

Key Terminology

Context of Mathematical Items and Tasks

There are several names in the mathematics education literature for what this paper (along with many other researchers) calls context, such as *cover history*, *thematic content*, and *setting*. In order to have a more refined understanding of the term context, Clarke and Helme (1998, p. 132) define *figurative context* as "the scenario where the task is encountered". Busse and Kaiser (2003) further refine figurative context into the *objective figurative context* and the *subjective figurative context*. According to these authors, the *objective figurative context* is related to the "description of the scenario given in the task" (p. 4) whereas the *subjective figurative context* refers to the "individual representation of the objective figurative context" (p. 4). The *objective figurative context* is "often implicitly meant by researchers when referring to the context" (Busse & Kaiser, 2003, p. 4). For this reason, the term *objective figurative context* is referred to as (item) *context* in this study except when otherwise stated. The context also needs to be distinguished from the

2014. In J. Anderson, M. Cavanagh & A. Prescott (Eds.). Curriculum in focus: Research guided practice *(Proceedings of the 37th annual conference of the Mathematics Education Research Group of Australasia)* pp. 55–62. Sydney: MERGA.

stimulus—the actual material about the item context that is presented by the item to the student (e.g., in words, pictures and graphs).

Mathematical Core

The phrase *mathematical core* is used to describe the underlying mathematical model and solving challenges inherent in a problem. In this study, items are described as having the same *mathematical core* if they have the same or very similar underlying mathematical model, similar calculational demands, etc.

Levels of Context Use

The very first distinction among levels of context use can be traced back to de Lange (1979). This was refined by OECD (2009) to distinguish three levels of context use. Two of these levels are common in PISA items. These levels were borrowed for this research.

The *zero-order use of context* is when the item generally involves just mathematical terms, shapes, data and the translation of textually packaged mathematical problems (e.g., about triangles). Alternatively, some real-world terms might be included just to camouflage or add a little interest to the mathematical operations required. An early years' example is 'four more than three apples'. PISA has used almost no zero-order context items. *First-order use of context* takes place when the context is "needed for solving the problem and judging the answer" (OECD, 2009, p. 31). Finally, the *second-order use of context* takes place when "students need to move backwards and forwards between the mathematical problem and its context in order to solve the problem or to reflect on the answer within the context to judge the correctness of the answer" (OECD, 2009, p. 31). Information drawn from knowledge of the context (not just the problem statement or known mathematical facts) is required to solve the problem. A similar classification of the levels of context use is provided by Stillman (1998). According to her, context can function as a border (zero-order/first-order), a wrapper (first-order) or tapestry (second-order and more) for a problem. As discussed below, these definitions of levels need clarification in further work.

Research into Context

In school mathematics, when formulating mathematical representations of tasks embedded in contexts, factors such as (a) activation of real-world knowledge, (b) students' engagement, and (c) students' familiarity with the task context are found to be important elements affecting the task accessibility to students and performance on these tasks, although the effects and their exact causes are not yet well known. Studies related to the factors above will be now briefly discussed.

Context and Activation of Real-World Knowledge

Two seminal studies in this area are by Carraher, Carraher and Schliemann (1985, 1987). In Carraher at al. (1985), the performance of five young Brazilian street sellers (aged 9 to 15 years old) on mathematical tasks presented in real-world contexts (involving selling, giving change) was better than on school-type mathematical tasks and on context-free symbolic tasks involving identical numbers and operations. The researchers explained the better performance by differences in the way students approached the real-world versions because in the simulated store contexts students mentally worked with money (a concrete real-world construct), which changed the arithmetic demand of the tasks. Carraher et al. (1987) conducted a follow-up study with 16 Brazilian third graders. These authors found that students showed significant differences in performance when they solved simulated store contexts (outside school contexts, usually presented in verbal form), than problems inside school contexts presented in written form, and symbolic computation exercises; Carraher et al. (1987) found that real-world knowledge facilitated the task's accessibility, hence it could lead students to a greater performance.

Baranes, Perry, and Stigler (1989) intended to replicate Carraher et al. (1987)'s findings with Year 3 American students (who do not work as street sellers and almost certainly had more consistent schooling) but found no contextual effects in either performance or strategy use for success with the American sample (n=18); that is to say, the students did not generally activate their real-world knowledge and representation of it in the solution of the tasks. Participating students sometimes activated their real-world knowledge when the numbers used in the word problems induced students to use "a culturally constituted system of quantification, such as money" (Baranes et al., 1989, p. 316).

Context and Engagement

In the examination of the positive effect of context on students' performance, better performance by students can be associated with more involvement with the task context. Stillman (2002) investigated exhaustively how forty-three Year 11 and 12 students used the task context during modelling tasks in selecting the mathematics they needed initially, in keeping on track throughout the particular task they were attempting (and generally in their task solving), and as a final check at the end of a task. Her study found that mathematical modelling tasks resulted in more understanding and thus greater involvement with the context than standard application tasks; and that better performance was correlated with medium to high involvement with the context. Nevertheless, students who performed well in mathematics were found to be an exception: they did well with low involvement.

Context Familiarity

A well-known postulate is that a familiar context "provides a less abstract and more directly experienced grounding for the new domain, thus enhancing the use of particular strategies that allow for more efficient processing" (Huang, 2004, p. 279). From a theoretical stance it follows that familiar contexts might enhance students' performance. This has been found in studies such as that of Chipman, Marshall, and Scott (1991).

An increasing body of empirical research evidence determines, however, that familiarity with a context can be associated with either negative or neutral impact on students' performance (Huang, 2004; van den Heuvel-Panhuizen, 2005). Huang (2004) investigated to what extent the performance and perception of task difficulty of 48 Taiwan Year 4 students were influenced by familiar or unfamiliar contexts in four everyday shopping mathematical tasks. The data did not support the hypothesis that familiar contexts improve performance; instead students were statistically significantly better on tasks embedded in unfamiliar contexts. Moreover, students spent a statistically significant longer time in solving tasks with familiar contexts.

Van den Heuvel-Panhuizen (2005) also acknowledged that familiar contexts are not always helpful to students and may also generate difficulties in students' problem solving. Familiar contexts can hinder some students finding an answer, while others may focus on contextual aspects and fail to engage with the necessary mathematics required to solve the task. Boaler (1994) gave an example of this situation. She analysed performance of 50 female students on two sets of questions intended to assess the same mathematical content (equivalence of fractions) but set in different contexts (i.e. soccer season, planting plants, cutting pieces of wood, and a fashion workshop). Results show that females underachieved in task contexts with which they were probably more familiar (e.g. fashion rather than soccer). They often took excessive account of contextual information in the tasks. Boaler (1994) speculated that the relative underachievement on the fashion task was because the attractive and familiar context distracted the students from the mathematical structure.

The above review has raised several issues concerning contextual influences on students' performance, which this study aims to examine. The results in the literature on students' performance with reference to the effects of context on mathematical tasks are mixed; it can have a positive, a negative or a neutral impact for all the factors reviewed above (i.e. familiarity, activation of real-world knowledge, etc.). It is also possible that context will operate differently with different types of tasks (word problems, extended modelling tasks, etc.). Hence, the effects of context on PISA mathematics items are undetermined, due to the lack of studies addressing this issue.

The Study

This study is a pilot study to investigate the effects of (1) familiarity and (2) the level of context use on students' performance on PISA items. The item context was varied whilst holding constant other features of the item that are known to affect performance (performance is interpreted as average percent correct). There are three research questions:

1. Is students' performance better on items with more familiar contexts?
2. Is students' performance on items that require second-order context use more affected by change of context than on items that require first-order context use?
3. How does context affect students' solutions of mathematics items and their performance?

Selection, Characteristics and Creation of Test Items

Four PISA Mathematics items released after the 2000, 2003 and 2006 surveys were selected (OECD, 2006). The items had approximately similar difficulty, and two were judged by the researchers to require first-order use of context, and two required second-order use of context (taking into account that the subjects are 15 year old students). Items with low percentage correct were chosen because it was thought that the contexts used may have made them more difficult, and also that any improved performance on items with increased familiarity of the context could show up.

Table 1 gives the names and PISA codes of the selected items, along with the average percentage of students in all OECD countries receiving full credit (OECD, 2009) and level of context use, as judged by the researchers. Table 1 also names the four 'PISA-like' items that were specially constructed for this study. Each PISA item had a PISA-like sibling item where the context was changed to be more familiar to students, but with all other relevant aspects of the problem held constant. To exemplify this, Pi-A was about scheduling internet chat between students in Germany and Australia accounting for time zone differences. The context of its PISA-like sibling PLi1-A was to schedule TV broadcasts for India of events at the Commonwealth Games that were being held in Australia at the time of data collection. The timeliness of the Commonwealth Games, and frequent discussions about when events are shown on TV, led the researchers to judge that this would be a more familiar context for the sample population than scheduling internet chat. Held constant were the mathematical core of the problems, the level of context use, the required amount of receptive and expressive communication, response type, etc. In this pilot study, increased familiarity of the siblings was judged by the researchers. The PISA items and siblings are fully described in Almuna Salgado (2010).

Creating sibling items was difficult, and reveals deep aspects about how context is embedded in the problem-solving process. Finding an appropriate new context requires imagination and understanding of students' interests. Especially for the second-order use of context, although the basic underlying mathematical model may be the same, the new assumptions and information required were often very different and changed the problems. This foreshadows a need for further work to refine the notion of mathematical core. For example, Pi-D *Robberies* gave a graph of the number of robberies in a town, and asked students to evaluate a reporter's statement that there had been a big increase. Truncated axes made the graph potentially misleading. The mathematical core related to considering both absolute and relative change. However, whether a change is truly large or small and whether absolute or relative changes are most important varies markedly with context: consider increased train fares, house prices, change in CO_2 concentration in the air or average temperature of the earth. Context seems to be a dynamic feature of a task even when the cognitive demand, competencies required and the model remain constant. In summary, very careful consideration was needed to construct the siblings.

Table 1
PISA Items Selected for the Study and their Sibling Items

Item code	PISA item name	PISA item code[A]	Percentage receiving full credit[A]	Item's use of context	Code of sibling item	Context of sibling item
Pi-A	Internet Chat Relay	M402Q02	29%	First	PLi1-A	2010 Commonwealth Games on TV
Pi-B	Rock Concert	M552Q01	28%[B]	Second	PLi1-B	Car park
Pi-C[C]	Support for the President	M702Q01	36%	First	PLi1-C	Support for a new shopping centre
Pi-D	Robberies	M179Q01	30%	Second	PLi1-D	Train ticket prices

A Source OECD (2009). **B** Field trial data (source OECD, 2013). **C** Pi-C was simplified for this study by omitting one of the factors (survey time) given in M702Q01.

Participants and the Rotated Test Design

The target population in this study was students approximately 15 years old, because PISA is administered at this age. The performance of thirty students from two Year 10 classes at one volunteer school in Melbourne, Australia, on four PISA items was compared using paper-and-pencil tests. Participants were tested in a class period. Because of the limited time for testing and potential inter-item effects, a rotated design of three booklets was used, so that students did not solve both an item and its sibling. This means that direct comparison of performance at the level of the student was not possible. Each booklet contained PISA items and PISA-like items with first and second-order use of context (see Table 2). Booklets were distributed at random to 10 students each. Booklets also contained some additional items not discussed here. To answer Research Question 3, a sample of students was interviewed about their solution processes using stimulated recall. Space limitations preclude a proper report here (see Almuna Salgado (2010)), although insights have been included in the discussion section below.

Results

Table 2 presents the item names and booklets, sample size, distribution of students' scores (assessed using the published PISA scheme), item means and standard deviations. The more familiar context of the sibling items, compared to the PISA items, resulted in the expected improved performance only for Pi-C and PLi1-C (Presidential poll compared to shopping centre). Performance was not affected for the A sibling pair, and the more familiar items for the B and D sibling pairs had lower performance. Overall more familiar contexts did not improve performance. Instead, performance was worse with familiar contexts, but not to a statistically significant extent ($F(1,78) = 1.179, p = 0.281$).

Table 2
Distribution of Students' Scores (Items scored out of 2, sample size 10 students per item)

Item	Booklet	Full credit (2)	Partial credit (1)	No credit (0)	Mean	Item s.d.
Pi-A	A	50%	-	50%	1.00	1.05
Pi-B	C	50%	-	50%	1.00	1.05
Pi-C	C	60%	0%	40%	1.20	1.03
Pi-D	B	20%	40%	40%	0.80	0.78
Average performance on Pi					1.00	0.97
PLi1-A	C	50%	-	50%	1.00	1.05
PLi1-B	A	20%	-	80%	0.40	0.84
PLi1-C	B	50%	40%	10%	1.40	0.69
PLi1-D	C	0%	30%	70%	0.30	0.48
Average performance on PLi1					0.77	0.76

Note: Partial credit not available on Pi-A, Pi-B and siblings based on the PISA marking scheme.

Research Question 2 hypothesised that changing context would change performance more on items with second-order than first-order use of context. Since this hypothesis predicted the magnitude of change, rather than whether one context of a pair made an item easier or harder than another, the hypothesis was examined by analysing the absolute values of the differences in performance. The data in Table 2 support this hypothesis. Because there is no student level data on change, only the average changes for the items in the sibling pairs can be compared. The absolute changes in mean scores for the two first-order sibling pairs are 0.00 and 0.20 respectively, and for the second-order sibling pairs are 0.60 and 0.50 respectively. With only four sibling pairs, no statistical tests are appropriate, but the data agree with the hypothesis.

Discussion and Conclusions

The study aimed to examine differences in the performance of participants in two Year 10 classes when they solved PISA items with the same mathematical core whilst varying the familiarity of contexts. Embedding items in a context more familiar to students did not assist students' performance. This might be explained by the fact that the greater familiarity of items in this study was not empirically determined, but was only established from the researchers' opinions. It may also be explained if the new and more familiar sibling items were not technically as well constructed as the multiply-trialled PISA items, but this is unlikely because the PISA items were such a close model for their siblings.

Qualitative evidence from the interviews revealed that, in more familiar contexts, students tended to bring personal information into arguments rather than using a mathematical argument; a familiar context was sometimes interpreted and judged as personal rather than from a mathematical point of view. Students' comments suggested that certain familiar contexts (e.g. train ticket prices) encouraged them to think and argue in personal terms rather than mathematically. Students who considered that the absolute proposed ticket price increase was a very large amount of money (which is part of their subjective figurative context) may not have considered at all whether the proposed increase was a small percentage, quite in line with normal inflation—indeed, they may not know about the likely size of normal inflation. In addition, with this familiar context (i.e. train ticket prices) some students were reality bound; they considered a familiar context as real using Melbourne metropolitan ticket prices to judge their answers which produced a large scenario for a narrow task. For example, some students took into consideration the different kinds of tickets and prices in Melbourne; the subjective figurative context was seen to guide the selection of arguments to communicate students' answers. In this vein, one difficulty with familiar contexts is that they tend to elicit responses in students that may be based on integration of personal knowledge and values with mathematics in order to build an intended solution. Familiar contexts also may be borderline cases where the relatively stronger understanding of a task plays a role when students communicate an answer; they may assume that it is not necessary to give a very detailed answer because everyone already knows the arguments. For example, some students did not explain that their answers were based on proportional reasoning (so they did not get full credit), perhaps assuming that would be well understood.

On the other hand, the results in Table 2 provided some support for the hypothesis that the items involving the second-order use of context were more affected by change of context than items involving the first-order use of context. The experience of creating the sibling pairs for the items with second-order use of context also pointed in this direction. The research hypothesis above seemed therefore to be supported, but the data set (using two pairs of items for each level of context use) was too small to make a strong claim.

To conclude, the findings in this study are in line with the research literature, which suggests that task context can impact on students' performance in variable ways. This leads us to infer that context needs more careful study with deeper analysis and experimental control of the way in which context is involved. Plans for carrying out such further study include (a) getting data from students on familiarity of the contexts, (b) extending the operationalised definition of mathematical core beyond the notion of mathematical model to include the mathematical competencies (OECD, 2009), mathematical procedure, computation and strategies and solving challenges of every mathematical task, (c) using more sensitive marking schemes than the PISA scheme, which is necessarily simple in order to be used by many teams for large sample sizes in many languages, and (d) refining the definition and operationalisation of the levels of context use. Better information about how context affects student performance might help teachers to instruct students how to work more effectively with problems in context.

References

Almuna Salgado, F. (2010). *Investigating the effect of item-context on students' performance on mathematics items.* Masters Coursework thesis, Melbourne Graduate School of Education, The University of Melbourne. Retrieved from http://repository.unimelb.edu.au/10187/10041

Baranes, R., Perry, M., & Stigler, J. W. (1989). Activation of real-world knowledge in the solution of word problems. *Cognition and Instruction, 6*(4), 287-318.

Blum, W. (2002). ICMI STUDY 14: Applications and modelling in mathematics education. *Educational Studies in Mathematics, 51*(1/2), 149-171.

Boaler, J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to 'realistic' mathematics contexts. *British Educational Research Journal, (5),* 551-564.

Busse, A., & Kaiser, G. (2003). Context in application and modelling - an empirical approach. In Q. X. Ye, W. Blum, S. Houston & Q. Y. Jiang (Eds.), *Mathematical modelling in education and culture: ICTMA 10* (pp. 3-15). Chichester, UK: Albion Publishing.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology, 3*(1), 21-29.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education, 18*(2), 83-97. doi: 10.2307/749244

Chipman, S., Marshall, S. & Scott, P. (1991). Content effects on word problem performance: A possible source of test bias? *American Educational Research Journal*, *28*(4), 897 - 915.

Clarke, D., & Helme, S. (1998). Context as construction. In O. Björkqvist (Ed.), *Mathematics teaching from a constructivist point of view* (pp. 129-147). Vasa, Finland: Faculty of Education, Abo Akademi University.

de Lange, J. (1979). Contextuele problemen [Contextual problems]. *Euclides, 55*, 50-60.

de Lange, J. (2007). Large-scale assessment and mathematics education. In F. K. Lester (Ed.) *Second handbook of research on mathematics teaching and learning*. (pp. 1112 – 1142) Charlotte, N.C.: Information Age Publishing.

Huang, H. (2004). The impact of context on children's performance in solving everyday mathematical problems with real-world settings. *Journal of Research in Childhood Education, 18*(4), 278-292.

Organisation for Economic Co-Operation and Development. (2006). *PISA released items-Mathematics*. Retrieved March 12, 2014, from http://www.oecd.org/pisa/38709418.pdf

Organisation for Economic Co-Operation and Development (2009). *Learning mathematics for life: A perspective from PISA*. Paris: Organisation for Economic Co-operation and Development. Retrieved March 12, 2014, from http://browse.oecdbookshop.org/oecd/pdfs/free/9809111E.PDF

Organisation for Economic Co-Operation and Development. (2013). *PISA 2012. Assessment and analytical framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. doi: 10.1787/9789264190511-en

Stillman, G. (1998). The emperor's new clothes? Teaching and assessment of mathematical applications at the senior level. In P. Galbraith, W. Blum, G. Booker & D. Huntley (Eds.), *Mathematical modelling: Teaching and assessment in a technology-rich world* (pp. 243-253). Chichester: Horwood Publishing.

Stillman, G. (2000). Impact of prior knowledge of task context on approaches to applications tasks. *The Journal of Mathematical Behavior, 19*(3), 333-361.

Stillman, G. (2002). *Assessing higher order mathematical thinking through applications.* (Unpublished Doctoral dissertation), The University of Queensland, Brisbane.

van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics, (2)*, 2-9.