Graphic-Rich Items within High-Stakes Tests: Indonesia National Exam (UN), PISA, and TIMSS

Destina Wahyu Winarti
University of Canberra
<Destina.Winarti@canberra.edu.au>

Sitti Maesuri Patahuddin
University of Canberra
<Sitti.Patahuddin@canberra.edu.au>

In this paper, we undertake a content analysis of mathematics assessment tasks to understand how often graphical representations are embedded within high-stakes national and international tests. A total of 274 items were analysed, consisting of 160 Grade 9 UN items, 88 Grade 8 TIMSS items, and 26 PISA items. Analysis showed that all items in the PISA test were embedded with graphics, with far fewer graphical items in the TIMMS and national UN tests (47% and 33% respectively). We also found that graphical items in UN tests are distinct from PISA and TIMSS, suggesting a misalignment between what is represented in UN tests and international instruments.

Graphics-based representations are a powerful tool to communicate information (Cleveland & McGill, 1985), and a number of studies provide strong evidence for the importance of graphic representation in mathematics assessment (Kulm, Dager Wilson, & Kitchen, 2005; Lowrie & Diezmann, 2009; Lowrie, Diezmann, & Logan, 2012). The emphasis on the importance of graphical representation within mathematics and teaching has encouraged many parties to place concern on principles and standards for graphical representation within school mathematics, including assessments. Such parties include the U.S. National Council of Teachers of Mathematics (NCTM), Department of Education U.K., and Australian Association of Mathematics Teachers (AAMT). Since curriculum development is influenced by assessment practices, we investigated the characteristics of graphical items within three high-stakes tests: Indonesia National Exam (UN), Trends in International Mathematics and Science Study (TIMSS), and Programme for International Student Assessment (PISA).

Although high-stakes testing is viewed as problematic by some (Abrams, Pedulla, & Madaus, 2003), observing assessment trends has its merits. These large-scale tests are used to make important decisions that affect students, teachers, administrators, communities, schools, and districts. Graphics-based representations in international tests are of particular interest, as Lowrie and Diezmann (2009) noted that graphics-rich tasks have become increasingly used in national tests over the past decade. It is important to understand the characteristics of graphical items within high-stakes tests, as test results are used for ranking and categorising schools, teachers, and children. Results are also reported to the public as part of the accountability movement (Au, 2007).

Graphical and Non-Graphical Mathematics Items

Graphics are defined as representations used to store, understand, and communicate essential information in a visual form (Bertin, 1983). Graphics include number lines, scales, maps, charts, and Venn diagrams (Logan, Lowrie, & Diezmann, 2009). Diezmann and Lowrie (2008) divided the roles of graphics in mathematics into two categories: context and information. In the present study, we extend this classification beyond the dichotomous categories of information and contextual graphics to include items that contain both of these attributes. The new category “combination graphic” is used to signify items that have both a contextual and information graphic embedded within the item. In

(2017). In A. Downton, S. Livy, & J. Hall (Eds.), *40 years on: We are still learning! Proceedings of the 40th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 569-576). Melbourne: MERGA.

this study, we categorise graphical items into three types: contextual, information, and combination graphics. Non-graphics items are described as items that only contain texts and/or symbols, and are categorised into two types: word problems and symbolic. Explanations and examples of each category are presented in Table 1.

Table 1
Definitions of Types of Graphics and Non-Graphics Test Items

Type		Definition	Examples in Appendix A
Graphics	Information	Conveys mathematical information that is required to solve the task (Diezmann & Lowrie, 2008)	Item 1 and 3
	Contextual	Represents objects, people or locations for illustrative purposes only, with no mathematical information related to the task (Diezmann & Lowrie, 2008)	Item 2
	Combination	Represents information that contain both the attributes of information and contextual graphics	Item 4
Non-Graphics	Word problem	Includes a text passage without a picture or graphic associated with it	Item 5
	Symbolic	Consists of numbers or symbols with short instructions, such as ‘do the following’ and ‘solve the following’	Item 6

The Nature of National and International Mathematics Assessment: UN, TIMSS, and PISA

A test is identified as a high-stakes test when its results are utilised to make critical decisions that affect test participants such as students, teachers, and administrators. It is often considered part of a policy design, and results are reported to the public. For instance, tests are administered to decide grade promotion or categorise school performances (Au, 2007; McNeil, 2000). In this study, we explored three high-stakes mathematics tests: the Indonesian national assessment (UN) and international tests TIMSS and PISA.

The UN is administered every year in Indonesia. It targets students in Grade 6 (11- to 13-year-olds), Grade 9 (14- to 16-year-olds), and Grade 12 (17- to 19-year-olds). It is intended to ensure that Indonesian education providers assess their students’ achievement against national education standards. It has been used to decide whether students can progress to the next level of schooling (Kemendikbud, 2007). Thus, UN significantly impacts classroom instruction.

TIMSS is administered every four years and targets students in Grade 4 (9- to 10-year-olds) and Grade 8 (13- to 14-year-olds). It is administered by International Association for the Evaluation of Educational Achievement (IEA), and is well known for providing international comparative assessments of educational achievement.

PISA is administered every three years, and assesses Grade 10 (15- to 16-year-olds) to determine the extent to which they have acquired key knowledge and skills that are fundamental for full participation in society. PISA is designed to assess whether students can reproduce what they have learned. It also examines how well students can extrapolate from what they have learned and apply that knowledge in unfamiliar settings, both inside and outside of school. In other words, it assesses what students can do with what they know (Organisation for Economic Co-Operation and Development [OECD], 2012).

Both PISA and TIMSS are widely accepted as performance benchmarks by participating countries (Mullis, Martin, Foy, & Arora, 2012; OECD, 2012). The key difference is that TIMSS aims to assess the coverage of mathematics curriculum of participating countries, while PISA aims to assess mathematical literacy that is considered critical for a student's life (Stacey, 2014).

Items in these three high-stakes tests have different formats. UN includes only multiple choice questions, while TIMSS includes multiple choice and short response questions. PISA items are group-based with a passage of text setting out a real-life situation (OECD, 2012). Students are required to provide their own answers. Each PISA group item can have three to four questions, and each question refers to the passage. Although TIMSS and PISA characterise item test differently, items can be categorised into four common strands: numbers, algebra, geometry, and statistics (Gronmo, Lindquist, Arora, & Mullis, 2015; Kemdikbud, 2016; OECD, 2012).

Method

This study is a part of an ongoing PhD project investigating the correlation between Indonesian students' spatial ability and mathematics performances. Mathematics performance is often measured through high-stakes tests, and the type of questions in the tests determines how much spatial reasoning is needed in solving them. For example, graphical items often require more spatial reasoning than non-graphical items (Lowrie & Diezmann, 2007). Students with better spatial ability can decode graphics relatively easy compared to those with lower spatial ability (e.g., Hegarty & Mayer, 2002; Vekiri, 2002). This led us to address the following questions:

1. What proportion of graphic and non-graphic items appears in high-stakes tests (UN, TIMSS, and PISA)?
2. What is the nature of graphical items in four different mathematical strands in these high-stakes tests?

Instruments and Procedure

To select sample tests for analysis, we mainly considered two aspects: the students' age (i.e., 14 to 16 years of age) and the time frame of the tests. UN is administered every year, TIMSS every four years (Mullis et al., 2012), and PISA every three years (OECD, 2012). Therefore ,we analysed test items from UN 2011–2014 inclusive, TIMSS 2011, and PISA 2012. As a result, the data set included 160 items from UN, 88 released items from TIMSS 2001, and 26 released main survey items from PISA 2012.

UN items were downloaded from various websites created by teachers in Indonesia. TIMSS items were retrieved from the National Center for Education Statistics website (IEA, 2013). PISA items were downloaded from the Organisation for Economic Co-operation and Development website (OECD, 2013).

Data Analysis

Content analysis was employed to classify test items. Two researchers independently coded all the items with the following procedure: First, each item test was coded as either a graphical (G) or a non-graphical (NG) item. Second, all graphical items were further assigned with one of the three codes: information graphic (IG), contextual graphic (CG), or combination graphic (MIX). Third, each of the non-graphical items were coded as word problems (WP) or symbolic (SM). An example of each type of graphic item is provided in Appendix A. Fourth, each item was classified into four strands: number, algebra, geometry, and statistics (IEA, 2013). Coded items were recorded in an Excel spreadsheet for descriptive data analysis. Coding reliability was high, with 95% agreement between the two researchers. The remaining 5% of the codes were agreed upon after discussion.

Results

Results from this study are presented in two parts, according to the research question they address. Part 1 presents the proportion of graphics and non-graphical items within three high-stakes tests. Part 2 reports the nature of graphical items in mathematical strands.

Proportion of Graphic and Non-Graphic Items by Instrument

Analysis of the data revealed a wide diversity in the proportion of graphic and non-graphic items across the three tests. Table 2 shows that graphical items accounted for approximately 33% in UN, and 47% in TIMSS, while PISA items are completely (100%) graphics-based.

The next stage of analysis revealed that UN and TIMSS graphical items were mostly information graphics, with a small proportion of combination graphics and no contextual graphics. Most PISA items were coded as combination, but there was also a reasonable proportion of information and contextual graphics. Data analysis also revealed that non-graphic items were mostly in the form of word problems. Results are displayed in Table 2.

Table 2
Item Representation across the Three High-Stakes Tests

Type of test	Graphics (%)				Non-Graphics (%)		
	IG	CG	MIX	Total	WP	SM	Total
UN (160)	32	0	1	33	53	14	67
TIMSS (88)	38	0	9	47	41	13	54
PISA (26)	27	35	38	100	0	0	0

Note: Percentage is rounded to the nearest whole number.

Graphic Items by Strand

Another level of analysis involved identifying the proportion of graphic items in each of the four mathematical strands: numbers, algebra, geometry, and statistics. As shown in Figure 1, graphical items appeared in all of the four strands.

In each instrument, geometry items were the most prevalent among the high-stakes tests, with the UN having the highest proportion of geometry items. This indicates that the Indonesian graphic items are most commonly aligned to geometry content. The proportion of graphic items within the number strand varies across the tests, with UN having the least and PISA having the most. PISA more than doubled the proportion of graphics in number items, compared to TIMSS. In the algebra strand, UN had the lower proportion compared to TIMSS and PISA, which have the same proportion. In the statistics strand, TIMMS has the highest proportion of graphics elements, followed by PISA and UN. Generally, PISA has a more balanced proportion across strands, compared to UN and TIMSS.

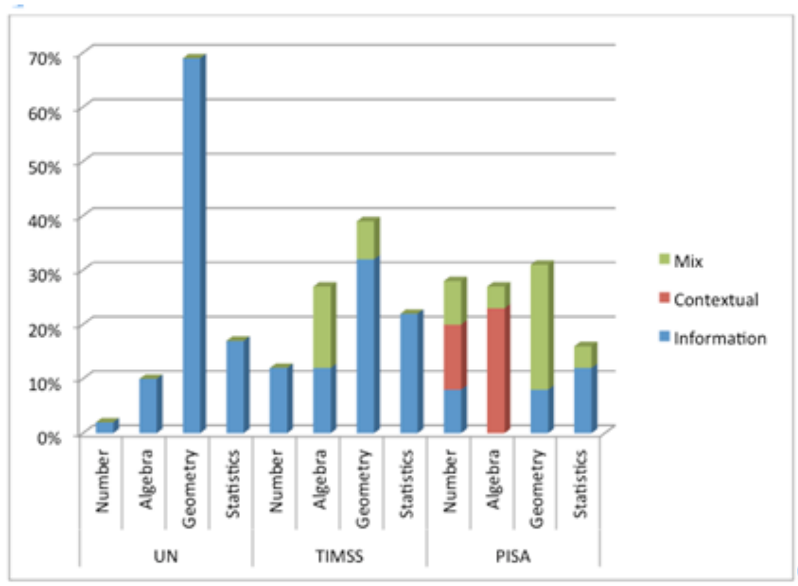


Figure 1. Proportion of graphical items in three high-stakes tests by strands and types.

Discussion and Conclusion

Our analysis of items in high-stakes tests identifies three discussion points. First, this study highlights the importance of graphics-based representations in mathematics assessments today. Content analysis showed that graphics-based items were used frequently in high-stakes tests. Overall, approximately 65% of all mathematical items in the three high-stakes tests contained a graphic. Since high-stakes tests often inform education policies (Au, 2007; Madaus, 1988), this can impact students' lives. In the case of UN tests, results can determine whether a student progresses to the next grade or graduates, so it is critical for students to learn graphical mathematics tasks (Logan et al., 2014). A rigorous design for graphical tasks in high-stakes tests should be applied, to ensure that students are not unfairly disadvantaged (Lowrie, Diezmann, & Logan, 2012).

Second, the findings suggest that Indonesian students are exposed to more word problems than graphics, despite Indonesian society using a vast array of information graphics outside of school, such as graphs, diagrams, tables, and maps. The Indonesian National Exams (UN) use the fewest graphical items, compared to TIMSS and PISA. UN items are predominantly word problems (55%) across each of the strands (number, algebra, geometry, and statistics). It is somewhat the opposite with the other two high-stakes tests, which have a lower proportion of word problems. In TIMSS, for instance, word problem items took up on average 45% of the total items, while PISA does not have any questions categorised as word problems.

Third, this study found that graphics-based items mostly appeared in geometry strands for all three high-stakes tests. This is not surprising, as school geometry is highly linked to interpreting shapes and other graphics. However, number and algebra questions in UN tests hardly contained any graphical items, compared to TIMSS and PISA. This indicates that number and algebra items in international high-stakes test often include context, while UN items tend to measure fluency (See examples in Appendix A: Item 3, Item 5, and Item 6). This pattern has been recognised by other researchers (e.g., Edo, Ilma, & Hartono, 2014), and Indonesian researchers recently advocated for the use of context with assessment (e.g., Kohar, Zulkardi, & Darmawijoyo, 2014).

This investigation highlights differences in the types of graphics used in UN tests compared to other international tests. Since UN is the national exam in Indonesia, it represents the type of content to which Indonesian students are exposed at school. The item structure of the UN test is quite different from that of international comparison tests, and Indonesian students may find it difficult to decode representations that frame mathematics thinking within contexts. As other researchers (e.g., Greenlees , 2015; Logan & Greenlees, 2008) maintained, contextual information embedded within graphics can dramatically influence sense-making. Though further research is required in this area, our findings provide some understanding of why Indonesian students perform poorly on TIMSS and PISA assessments.

References

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice*, *42*(1), 18-29.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267.

Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps.* Madison, WI: The University of Wisconsin Press.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science, 229*(4716), 828-833.

Diezmann, C., Lowrie, T., Sugars, L., & Logan, T. (2009). The visual side to numeracy: Students' sensemaking with graphics. *Australian Primary Mathematics Classroom, 14*(1), 16-20.

Diezmann, C. M., & Lowrie, T. (2008). The role of information graphics in mathematical proficiency. In M. Goos, R. Brown, & K. Makar (Eds.), *Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 647-650). Brisbane: MERGA.

Edo, S. I., Ilma, R., & Hartono, Y. (2014). Investigating secondary school students' difficulties in modeling problems PISA-model level 5 and 6. *Journal on Mathematics Education, 4*(1), 41-58.

Greenlees, J. (2015). The past, present and future of Australian national assessment. In C. Vistro-Yu (Ed.), *In pursuit of quality mathematics education for all: Proceedings of the 7th ICMI-East Asia Regional Conference on Mathematics Education* (pp. 75-86). Quezon City, Phillipines: Philippine Council of Mathematics Teacher Educators.

Gronmo, L. S., Lindquist, M., Arora, A., & Mullis, I. V. S. (2015). TIMSS 2015 assessment frameworks. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 assessment frameworks* (pp. 11-24). Boston, MA: TIMSS & PIRLS International Study Centre.

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction*, *20*(1), 47-77.

International Association for the Evaluation of Educational Achievement. (2013). *TIMSS 2011 Grade 8 released mathematics items*. Retrieved from https://nces.ed.gov/timss/pdf/TIMSS2011_G8_Math.pdf

Kemdikbud. (2016). *Kisi-Kisi UN SMP/MTS 2017.* Retrieved from http://bsnp-indonesia.org/wp-content/uploads/2016/12/KISI-KISI-UN-SMP-MTs-2017.pdf

Kemdikbud. (2007). *Peraturan menteri pendidikan nasional Republik Indonesia nomor 34 tahun 2007.* Retrieved from http://pendis.kemenag.go.id/file/dokumen/permen342007.pdf

Kohar, A. W., Zulkardi, Z., & Darmawijoyo, D. (2014). *Developing PISA-like mathematics tasks to promote students' mathematical literacy.* Retrieved from http://eprints.unsri.ac.id/5186/1/Ahmad_Wachidul_Kohar.pdf

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

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

Logan, T., Lowrie, T., & Diezmann, C. M. (2014). Co-thought gestures: Supporting students to successfully navigate map tasks. *Educational Studies in Mathematics, 87*, 87-102. doi:10.1007/s10649-014-9546-2

Lowrie, T. (2012). Visual and spatial reasoning: The changing form of mathematics representation and communication. In *Reasoning, communication and connections in mathematics: Yearbook 2012, Association of Mathematics Educators* (pp. 149-168). Singapore: World Scientific.

Lowrie, T., & Diezmann, C. M. (2007). Solving graphics problems: Student performance in junior grades. *The Journal of Educational Research, 100*(6), 369-378.

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.

McNeil, L. (2002). *Contradictions of school reform: Educational costs of standardized testing*. New York, NY: Routledge.

Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Herengracht, The Netherlands: International Association for the Evaluation of Educational Achievement.

Organisation for Economic Co-Operation and Development. (2012). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Retrieved from https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Organisation for Economic Co-Operation and Development. (2013). *PISA 2012 released mathematics items*. Retrieved from http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf

Stacey, K. (2014). The PISA view of mathematical literacy in Indonesia. *Journal on Mathematics Education*, *2*(2), 95-126.

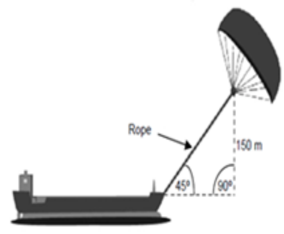
Vekiri, I. (2002). What is the value of graphical displays in learning? *Educational Psychology Review*, *14*(3), 261-312.

Appendix A: Examples of Graphical Type of Items

Question 3: SAILING SHIPS PM923Q03

Approximately what is the length of the rope for the kite sail, in order to pull the ship at an angle of 45° and be at a vertical height of 150 m, as shown in the diagram opposite?

A 173 m
B 212 m
C 285 m
D 300 m



Note: Drawing not to scale.
© by skysails

SAILING SHIPS SCORING 3

QUESTION INTENT:

Description: Use Pythagorean Theorem within a real geometric context
Mathematical content area: Space and shape
Context: Scientific
Process: Employ

Item 1. PISA: Information graphics (Geometry strand).

CLIMBING MOUNT FUJI

Mount Fuji is a famous dormant volcano in Japan.

Question 1: CLIMBING MOUNT FUJI PI

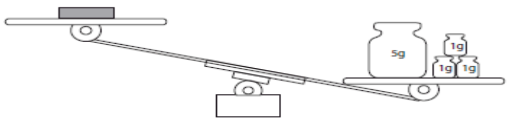
Mount Fuji is only open to the public for climbing from 1 July to 27 August each year. About 200 000 people climb Mount Fuji during this time.

On average, about how many people climb Mount Fuji each day?

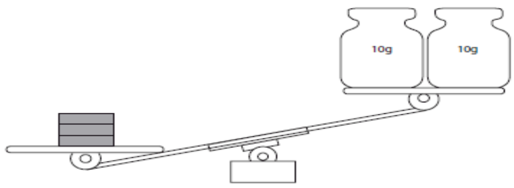
A 340
B 710
C 3400
D 7100
E 7400

Item 2. PISA: Contextual graphics (Number strand).

Jo has three metal blocks. The weight of each block is the same.
When she weighed one block against 8 grams, this is what happened.



When she weighed all three blocks against 20 grams, this is what happened.



Which of the following could be the weight of one metal block?

A. 5 g
B. 6 g
C. 7 g
D. 8 g

Item 3. TIMSS: Information graphics (Algebra strand).

WHICH CAR?

Chris has just received her car driving licence and wants to buy her first car.

This table below shows the details of four cars she finds at a local car dealer.

Model:	**Alpha**	**Bolte**	**Castel**	**Dezal**
Year	2003	2000	2001	1999
Advertised price (zeds)	4800	4450	4250	3990
Distance travelled (kilometres)	105 000	115 000	128 000	109 000
Engine capacity (litres)	1.79	1.796	1.82	1.783

Question 1: WHICH CAR?

Chris wants a car that meets **all** of these conditions:

- The distance travelled is **not** higher than 120 000 kilometres.
- It was made in the year 2000 or a later year.
- The advertised price is **not** higher than 4500 zeds.

Which car meets Chris's conditions?

A Alpha
B Bolte
C Castel
D Dezal

Item 4. TIMSS: Combination graphics (Statistics strand).

A workman cut off $\frac{1}{5}$ of a pipe. The piece he cut off was 3 meters long.

How many meters long was the original pipe?

A. 8 m
B. 12 m
C. 15 m
D. 18 m

Item 5. TIMSS: Non-graphical word problems (Number strand).

Hasil dari $\sqrt{3} \times \sqrt{8}$ adalah ....

A. $2\sqrt{6}$
B. $3\sqrt{6}$
C. $4\sqrt{3}$
D. $4\sqrt{6}$

Item 6. UN: Non-graphical symbolic (Number strand).