The Mathematics Enthusiast

Volume 11
Number 3 *Number 3*

Article 8

12-2014

Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors

Ariyadi Wijaya

Marja van den Heuvel-Panhuizen

Michiel Doorman

Alexander Robitzsch

Recommended Citation

Wijaya, Ariyadi; van den Heuvel-Panhuizen, Marja; Doorman, Michiel; and Robitzsch, Alexander (2014) "Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors," *The Mathematics Enthusiast*: Vol. 11 : No. 3 , Article 8.
DOI: https://doi.org/10.54870/1551-3440.1317
Available at: https://scholarworks.umt.edu/tme/vol11/iss3/8

Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors

Ariyadi Wijaya[a, b1]; Marja van den Heuvel-Panhuizen[a]; Michiel Doorman[a]; Alexander Robitzsch[c]

[a] *Freudenthal Institute for Science and Mathematics Education, Utrecht University, The Netherlands*

[b] *Mathematics Education Department, Universitas Negeri Yogyakarta, Indonesia*

[c] *Federal Institute for Education Research, Innovation and Development of the Austrian School System, Austria*

Abstract: The intention of this study was to clarify students' difficulties in solving context-based mathematics tasks as used in the Programme for International Student Assessment (PISA). The study was carried out with 362 Indonesian ninth- and tenth-grade students. In the study we used 34 released PISA mathematics tasks including three task types: reproduction, connection, and reflection. Students' difficulties were identified by using Newman's error categories, which were connected to the modeling process described by Blum and Leiss and to the PISA stages of mathematization, including (1) comprehending a task, (2) transforming the task into a mathematical problem, (3) processing mathematical procedures, and (4) interpreting or encoding the solution in terms of the real situation. Our data analysis revealed that students made most mistakes in the first two stages of the solution process. Out of the total amount of errors 38% of them has to do with understanding the meaning of the context-based tasks. These comprehension errors particularly include the selection of relevant information. In transforming a context-based task into a mathematical problem 42% of the errors were made. Less errors were made in mathematical processing and encoding the answers. These types of errors formed respectively 17% and 3% of the total amount of errors. Our study also revealed a significant relation between the error types and the task types. In reproduction tasks, mostly comprehension errors (37%) and transformation errors (34%) were made. Also in connection tasks

[1] a.wijaya@uu.nl

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 11, no. 3**, pp. 555-584

students made mostly comprehension errors (41%) and transformation errors (43%). However, in reflection tasks mostly transformation errors (66%) were made. Furthermore, we also found a relation between error types and student performance levels. Low performing students made a higher number of comprehension and transformation errors than high performing students. This finding indicates that low performing students might already get stuck in the early stages of the modeling process and are unable to arrive in the stage of carrying out mathematical procedures when solving a context-based task.

Keywords: Context-based mathematics tasks; Mathematization; Modeling process; Newman error categories; Students' difficulties.

1. Introduction

Employer dissatisfaction with school graduates' inability to apply mathematics stimulated a movement favoring both the use of mathematics in everyday situations (Boaler, 1993a) and a practice-orientated mathematics education (Graumann, 2011). The main objective of this movement is to develop students' ability to apply mathematics in everyday life (Graumann, 2011) which is seen as a core goal of mathematics education (Biembengut, 2007; Greer, Verschaffel, & Mukhopadhyay, 2007). In addition, this innovation was also motivated by theoretical developments in educational psychology such as situated cognition theory (Henning, 2004; Nunez, Edwards, & Matos, 1999) and socio-cultural theory (Henning, 2004). Finally, this context-connected approach to mathematics education emerged from studies of mathematics in out-of-school settings such as supermarkets (Lave, 1988) and street markets (Nunes, Schliemann, & Carraher, 1993).

In line with this emphasis on application in mathematics education, the utilitarian purpose of mathematics in everyday life has also become a concern of the Programme for International Student Assessment (PISA) which is organized by the Organisation for Economic Co-operation and Development (OECD). PISA is a large-scale assessment which aims to determine whether students can apply mathematics in a variety of situations. For this purpose, PISA uses real world problems which require quantitative reasoning, spatial reasoning or problem solving (OECD, 2003b). An analysis of PISA results showed that the competencies measured in PISA surveys are better predictors for 15 year-old students' later success than the qualifications reflected in school marks (Schleicher, 2007). Therefore, the PISA survey has become an influential factor in reforming educational practices (Liang, 2010) and making decisions about educational policy (Grek, 2009; Yore, Anderson, & Hung Chiu, 2010).

Despite the importance of contexts for learning mathematics, several studies (Cummins, Kintsch, Reusser, & Weimer, 1988; Palm, 2008; Verschaffel, Greer, & De Corte, 2000) indicate that contexts can also be problematic for students when they are used in mathematics tasks. Students often miscomprehend the meaning of context-based tasks (Cummins et al., 1988) and give solutions that are not relevant to the situation described in the tasks (Palm, 2008). Considering these findings, the intention of this study was to clarify students' obstacles or difficulties when solving context-based tasks.

As a focus we chose the difficulties students in Indonesia have with context-based problems. The PISA 2009 study (OECD, 2010) showed that only one third of the Indonesian students could answer the types of mathematics tasks involving familiar contexts including all information necessary to solve the tasks and having clearly defined questions. Furthermore, less than one percent of the Indonesian students could work with tasks situated in complex situations which require mathematical modeling, and well-developed thinking and reasoning skills. These poor results ask for further research because the characteristics of PISA tasks are relevant to the mathematics learning goals mandated in the Indonesian Curriculum 2004. For example, one of the goals is that students are able to solve problems that require students to understand the problem, and design and complete a mathematical model of it, and interpret the solution (Pusat Kurikulum, 2003).

Although the Indonesian Curriculum 2004 takes the application aspect of mathematical concepts in daily life into account (Pusat Kurikulum, 2003), the PISA results clearly showed that this curriculum did not yet raise Indonesian students' achievement in solving context-based mathematics tasks. This finding was the main reason to set up the 'Context-based Mathematics Tasks Indonesia' project, or in short the CoMTI project. The general goal of this project is to contribute to the improvement of the Indonesian students' performance on context-based mathematics tasks. The present study is the first part of this project and aimed at clarifying Indonesian students' difficulties or obstacles when solving context-based mathematics tasks. Having insight in where students get stuck will provide us with a key to improve their achievement. Moreover, this insight can contribute to the theoretical knowledge about the teaching and learning of mathematics in context.

2. Theoretical background

2.1. Learning mathematics in context

Contexts are recognized as important levers for mathematics learning because they offer various opportunities for students to learn mathematics. The use of contexts reduces students' perception of mathematics as a remote body of knowledge (Boaler, 1993b), and by means of contexts students can develop a better insight about the usefulness of mathematics for solving daily-life problems (De Lange, 1987). Another benefit of contexts is that they provide students with strategies to solve mathematical problems (Van den Heuvel-Panhuizen, 1996). When solving a context-based problem, students might connect the situation of the problem to their experiences. As a result, students might use not only formal mathematical procedures, but also informal strategies, such as using repeated subtraction instead of a formal digit-based long division. In the teaching and learning process, students' daily experiences and informal strategies can also be used as a starting point to introduce mathematics concepts. For example, covering a floor with squared tiles can be used as the starting point to discuss the formula for the area of a rectangle. In this way, contexts support the development of students' mathematical understanding (De Lange, 1987; Gravemeijer & Doorman, 1999; Van den Heuvel-Panhuizen, 1996).

In mathematics education, the use of contexts can imply different types of contexts. According to Van den Heuvel-Panhuizen (2005) contexts may refer to real-world settings, fantasy situations or even to the formal world of mathematics. This is a wide interpretation of context in which contexts are not restricted to real-world settings. What important is that contexts create situations for students that are experienced as real and related to their common-sense understanding. In addition, a crucial characteristic of a context for learning mathematics is that there are possibilities for mathematization. A context should provide information that can be organized mathematically and should offer opportunities for students to work within the context by using their pre-existing knowledge and experiences (Van den Heuvel-Panhuizen, 2005).

The PISA study also uses a broad interpretation of context, defining it as a specific setting within a 'situation' which includes all detailed elements used to formulate the problem (OECD, 2003b, p. 32). In this definition, 'situation' refers to the part of the students' world in which the tasks are placed. This includes personal, educational/occupational, public, and scientific situation types. As well as Van den Heuvel-Panhuizen (2005), the PISA researchers also see that a formal mathematics setting can be seen as a context. Such context is called an 'intra-mathematical context' (OECD, 2003b, p. 33) and refers only to mathematical objects, symbols, or structures without any reference to the real world. However, PISA only uses a limited number of such contexts in its surveys and places most value on real-world contexts, which are called 'extra-mathematical contexts' (OECD, 2003b, p. 33). To solve tasks which use extra-mathematical contexts, students need to translate the contexts into a mathematical form through the process of mathematization (OECD, 2003b).

The extra-mathematical contexts defined by PISA are similar to Roth's (1996) definition of contexts, which also focuses on the modeling perspective. Roth (1996, p. 491) defined context as "a real-world phenomenon that can be modeled by mathematical form." In comparison to Van den Heuvel-Panhuizen and the PISA researchers, Roth takes a narrower perspective on contexts, because he restricts contexts only to real-world phenomena. However, despite this restriction, Roth's focus on the mathematical modeling of the context is close to the idea of mathematization as used in PISA.

Based on the aforementioned definitions of context, in our study we restricted contexts to situations which provide opportunities for mathematization and are connected to daily life. This restriction is in line with the aim of PISA to assess students' abilities to apply mathematics in everyday life. In conclusion, we defined context-based mathematics tasks as tasks situated in real-world settings which provide elements or information that need to be organized and modeled mathematically.

2.2. Solving context-based mathematics tasks

Solving context-based mathematics tasks requires an interplay between the real world and mathematics (Schwarzkopf, 2007), which is often described as a modeling process (Maass, 2010) or mathematization (OECD, 2003b). The process of modeling begins with a real-world problem, ends with a real-world solution (Maass, 2010) and is considered to be carried out in seven steps (Blum & Leiss, as cited in Maass, 2010). As the first step, a solver needs to establish a 'situation model' to understand the real-world problem. The situation model is then developed into a 'real model' through the process of simplifying and structuring. In the next step, the solver needs to construct a 'mathematical model' by mathematizing the real model. After the mathematical model is established, the solver can carry out mathematical procedures to get a mathematical solution of the problem. Then, the mathematical solution has to be interpreted and validated in terms of the real-world problem. As the final step, the real-world solution has to be presented in terms of the real-world situation of the problem.

In PISA, the process required to solve a real-world problem is called ‘mathematization’ (OECD, 2003b). This process involves: understanding the problem situated in reality; organizing the real-world problem according to mathematical concepts and identifying the relevant mathematics; transforming the real-world problem into a mathematical problem which represents the situation; solving the mathematical problem; and interpreting the mathematical solution in terms of the real situation (OECD, 2003b). In general, the stages of PISA’s mathematization are similar to those of the modeling process. To successfully perform mathematization, a student needs to possess mathematical competences which are related to the cognitive demands of context-based tasks (OECD, 2003b). Concerning the cognitive demands of a context-based task, PISA defines three types of tasks:

a. Reproduction tasks

   These tasks require recalling mathematical objects and properties, performing routine procedures, applying standard algorithms, and applying technical skills.

b. Connection tasks

   These tasks require the integration and connection from different mathematical curriculum strands, or the linking of different representations of a problem. The tasks are non-routine and ask for transformation between the context and the mathematical world.

c. Reflection tasks

   These tasks include complex problem situations in which it is not obvious in advance which mathematical procedures have to be carried out.

Regarding students’ performance on context-based tasks, PISA (OECD, 2009a) found that cognitive demands are crucial aspects of context-based tasks because they are – among other task characteristics, such as the length of text, the item format, the mathematical content, and the contexts – the most important factors influencing item difficulty.

2.3. Analyzing students’ errors in solving context-based mathematics tasks

To analyze students’ difficulties when solving mathematical word problems, Newman (1977, 1983) developed a model which is known as Newman Error Analysis (see also Clarkson, 1991; Clements, 1980). Newman proposed five categories of errors based on the process of solving mathematical word problems, namely errors of reading, comprehension, transformation, process skills, and encoding. To figure out whether Newman’s error categories are also suitable for analyzing students' errors in solving context-based tasks which provide information that needs to be organized and modeled mathematically, we compared Newman's error categories with the stages of Blum and Leiss' modeling process (as cited in Maass, 2010) and the PISA's mathematization stages (OECD, 2003b).

Table 1. Newman's error categories and stages in solving context-based mathematics tasks

Newman's error categories[a]	Stages in solving context-based mathematics tasks	
	Stages in Blum and Leiss' Modeling[b]	Stages in PISA's Mathematization[c]
Reading: Error in simple recognition of words and symbols	--	--
Comprehension: Error in understanding the meaning of a problem	Understanding problem by establishing situational model	Understanding problem situated in reality
--	Establishing real model by simplifying situational model	--
--	--	Organizing real-world problems according to mathematical concepts and identifying relevant mathematics
Transformation: Error in transforming a word problem into an appropriate mathematical problem	Constructing mathematical model by mathematizing real model	Transforming real-world problem into mathematical problem which represents the problem situation
Process skills: Error in performing mathematical procedures	Working mathematically to get mathematical solution	Solving mathematical problems
Encoding: Error in representing the mathematical solution into acceptable written form	Interpreting mathematical solution in relation to original problem situation Validating interpreted mathematical solution by checking whether this is appropriate and reasonable for its purpose	Interpreting mathematical solution in terms of real situation
--	Communicating the real-world solution	--

[a] (Newman, 1977, 1983; Clarkson, 1991; Clements, 1980); [b] (as cited in Maass, 2010); [c] (OECD, 2003b)

Table 1 shows that of Newman's five error categories, only the first category that refers to the technical aspect of reading cannot be matched to a modeling or mathematization stage of the solution process. The category comprehension errors, which focuses on students' inability to understand a problem, corresponds to the first stage of the modeling process ("understanding problem by establishing situational model") and to the first phase of the mathematization process ("understanding problem situated in reality"). The transformation errors refer to errors in constructing a mathematical problem or mathematical model of a real-world problem, which is also a stage in the modeling process and in mathematization. Newman's category of errors in mathematical procedures relates to the modeling stage of working mathematically and the mathematization stage of solving mathematical problems. Lastly, Newman's encoding errors correspond to the final stage of modeling process and mathematization at which the mathematical solution is interpreted in terms of the real-world problem situation. Considering these similarities, Newman's error categories can be used to analyze students' errors in solving context-based mathematics tasks.

2.4. Research questions

The CoMTI project aims at improving Indonesian students' performance in solving context-based mathematics tasks. To find indications of how to improve this performance, the first CoMTI study looked for explanations for the low scores in the PISA surveys by investigating, on the basis of Newman's error categories, the difficulties students have when solving context-based mathematics tasks such as used in the PISA surveys.

Generally expressed our first research question was:

1. *What errors do Indonesian students make when solving context-based mathematics tasks?*

A further goal of this study was to investigate the students' errors in connection with the cognitive demands of the tasks and the student performance level. Therefore, our second research question was:

2. *What is the relation between the types of errors, the types of context-based tasks (in the sense of cognitive demands), and the student performance level?*

3. Method

3.1. Mathematics tasks

A so-called 'CoMTI test' was administered to collect data about students' errors when solving context-based mathematics tasks. The test was based on the released 'mathematics units'[1] from PISA (OECD, 2009b) and included only those units which were situated in extra-mathematical context.

Furthermore, to get a broad view of the kinds of difficulties Indonesian students encounter, units were selected in which Indonesian students in the PISA 2003 survey (OECD, 2009b) had either a notably low or high percentage of correct answer. In total we arrived at 19 mathematics units consisting of 34 questions. Hereafter, we will call these questions 'tasks', because they are not just additional questions to a main problem but complete problems on their own, which can be solved independently of each other. Based on the PISA qualification of the tasks we included 15 reproduction, 15 connection and 4 reflection tasks. The tasks were equally distributed over four different booklets according to their difficulty level, as reflected in the percentage correct answers found in the PISA 2003 survey (OECD, 2009b). Six of the tasks were used as anchor tasks and were included in all booklets. Every student took one booklet consisting of 12 to 14 tasks.

The CoMTI test was administered in the period from 16 May to 2 July, 2011, which is in agreement with the testing period of PISA (which is generally between March 1 and August 31) (OECD, 2005). In the CoMTI test the students were asked to show how they solved each task, while in the PISA survey this was only asked for the open constructed-response tasks. Consequently, the time allocated for solving the tasks in the CoMTI test was longer (50 minutes for 12 to 14 tasks) than in the PISA surveys (35 minutes for 16 tasks) (OECD, 2005).

3.2. Participants

The total sample in this CoMTI study included 362 students recruited from eleven schools located in rural and urban areas in the province of Yogyakarta, Indonesia. Although this selection might have as a consequence that the students in our sample were at a higher academic level than the national average[2], we chose this province for carrying out our study for reasons of convenience (the first author originates from this province).

To have the sample in our study close to the age range of 15 years and 3 months to 16 years and 2 months, which is taken in the PISA surveys as the operationalization of fifteen-year-olds (OECD, 2005), and which also applies to the Indonesian sample in the PISA surveys, we did our study with Grade 9 and Grade 10 students who generally are of this age. However, it turned out that in our sample the students were in the age range from 14 years and 2 months to 18 years and 6 months (see Table 2), which means that our sample had younger and older students than in the PISA sample.

Table 2. Composition of the sample

Grade	Boys	Girls	Total	Min. age	Max. age	Mean age (SD)
Grade 9	85	148	233	14 y; 2 m	16 y; 7 m	15 y; 3 m (5 m)[a]
Grade 10	59	70	129	14 y; 10 m	18 y; 6 m	16 y; 4 m (7 m)
Total	144	218	362			

[a] y= year; m = month

Before analyzing students' errors, we also checked whether the ability level of the students in our sample was comparable to the level of the Indonesian students who participated in the PISA surveys. For this purpose, we compared the percentages of correct answers of 17 tasks included in the 2003 PISA survey (OECD, 2009b), with the scores we found in our sample.

To obtain the percentages of correct answers in our sample, we scored the students' responses according to the PISA marking scheme, which uses three categories: full credit, partial credit, and no credit (OECD, 2009b). The interrater reliability of this scoring was checked by conducting a second scoring by an external judge for approximately 15% of students' responses to the open constructed-response tasks. The multiple-choice and closed constructed-response tasks were not included in the check of the interrater reliability, because the scoring for these tasks was straightforward. We obtained a Cohen's Kappa of .76, which indicates that the scoring was reliable (Landis & Koch, 1977).

The calculation of the Pearson correlation coefficient between the percentages correct answers of the 17 tasks in the PISA 2003 survey and in the CoMTI sample revealed a significant correlation, $r\ (15) = .83$, $p < .05$. This result indicates that the tasks which were difficult for Indonesian students in the PISA 2003 survey were also difficult for the students participating in the CoMTI study (see Figure 1). However, the students in the CoMTI study performed better than the Indonesian students in the PISA 2003 survey. The mean percentage of correct answers in our study was 61%, which is a remarkably higher result than the 29% correct answers of the Indonesian students in the PISA survey. We assume that this result was due to the higher academic performance of students in the province of Yogyakarta compared to the performance of Indonesian students in general (see Note 2).

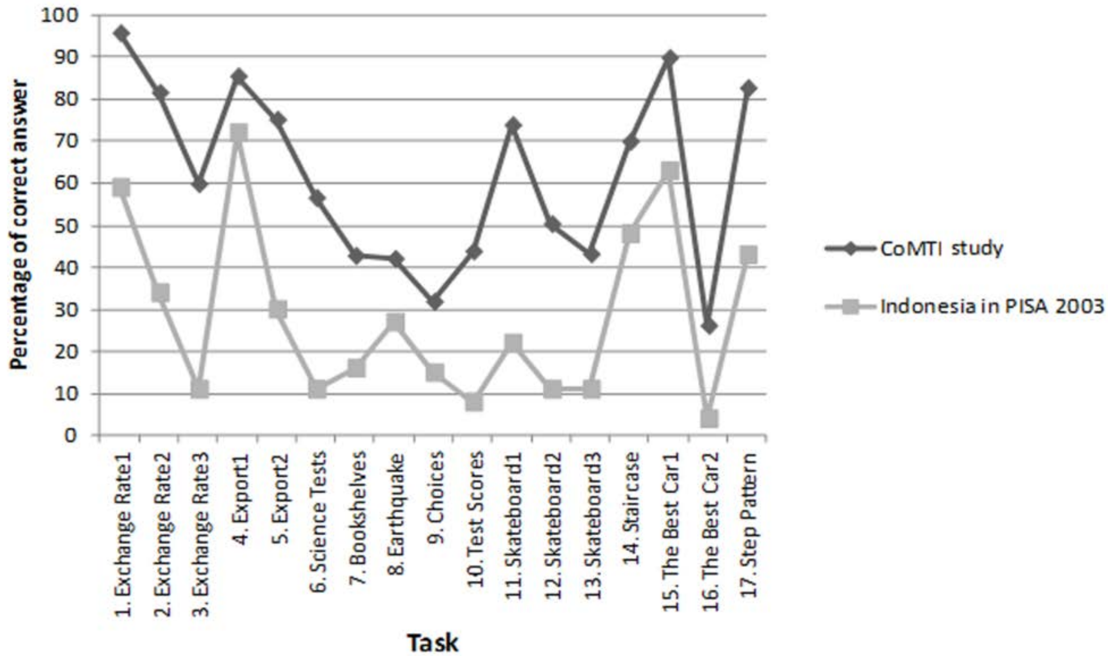


Figure 1. Percentage correct answers in the CoMTI sample and the Indonesian PISA 2003 sample

3.3. Procedure of coding the errors

To investigate the errors, only the students' incorrect responses, i.e., the responses with no credit or partial credit, were coded. Missing responses which were also categorized as no credit, were not coded and were excluded from the analysis because students' errors cannot be identified from a blank response.

The scheme used to code the errors (see Table 3) was based on Newman's error categories and in agreement with Blum and Leiss' modeling process and PISA's mathematization stages. However, we included in this coding scheme only four of Newman's error categories, namely 'comprehension', 'transformation', 'mathematical processing', and 'encoding' errors. Instead of Newman's error category of 'process skills', we used the term 'mathematical processing', because in this way it is more clear that errors in process skills concern errors in processing mathematical procedures. The technical error type of 'reading' was left out because this type of error does not refer to understanding the meaning of a task. Moreover, the code 'unknown' was added in the coding scheme because in about 8% of the incorrect responses, the written responses did not provide enough information for coding the errors. These responses with the code 'unknown' were not included in the analysis.

Table 3. Coding Scheme for error types when solving context-based mathematics tasks

Error type	Sub-type	Explanation
Comprehension	Misunderstanding the instruction	Student incorrectly interpreted what they were asked to do.
	Misunderstanding a keyword	Student misunderstood a keyword, which was usually a mathematical term.
	Error in selecting information	Student was unable to distinguish between relevant and irrelevant information (e.g. using all information provided in a task or neglecting relevant information) or was unable to gather required information which was not provided in the task.
Transformation	Procedural tendency	Student tended to use directly a mathematical procedure (such as formula, algorithm) without analyzing whether or not it was needed.
	Taking too much account of the context	Student's answer only referred to the context/real world situation without taking the perspective of the mathematics.
	Wrong mathematical operation/concept	Student used mathematical procedure/concepts which are not relevant to the tasks.
	Treating a graph as a picture	Student treated a graph as a literal picture of a situation. Student interpreted and focused on the shape of the graph, instead of on the properties of the graph.
Mathematical Processing	Algebraic error	Error in solving algebraic expression or function.
	Arithmetical error	Error in calculation.
	Error in mathematical interpretation of graph:	
	- Point-interval confusion	Student mistakenly focused on a single point rather than on an interval.
	- Slope-height confusion	Student did not use the slope of the graph but only focused on the vertical distances.
	Measurement error	Student could not convert between standard units (from m/minute to km/h) or from non-standard units to standard units (from step/minute to m/minute).
	Improper use of scale	Student could not select and use the scale of a map properly.
	Unfinished answer	Student used a correct formula or procedure, but they did not finish it.
Encoding		Student was unable to correctly interpret and validate the mathematical solution in terms of the real world problem. This error was reflected by an impossible or not realistic answer.
Unknown		Type of error could not be identified due to limited information from student's work.

To make the coding more fine-grained we specified the four error types into a number of sub-types (see Table 3), which were established on the basis of a first exploration of the data and a further literature review. For example, the study of Leinhardt, Zaslavsky, and Stein (1990) was used to establish sub-types related to the use of graphs, which resulted in three sub-types: ‘treating a graph as a picture’, ‘point-interval confusion’, and ‘slope-height confusion’. The last two sub-types belonged clearly to the error type of mathematical processing. The sub-type ‘treating a graph as a picture’ was classified under the error type of transformation because it indicates that the students do not think about the mathematical properties of a graph. Because students can make more than one error when solving a task, a multiple coding was applied in which a response could be coded with more than one code.

The coding was carried out by the first author and afterwards the reliability of the coding was checked through an additional coding by an external coder. This extra coding was done on the basis of 22% of students’ incorrect responses which were randomly selected from all mathematics units. In agreement with the multiple coding procedure, we calculated the interrater reliability for each error type and the code unknown, which resulted in Cohen’s Kappa of .72 for comprehension errors, .73 for transformation errors, .79 for errors in mathematical processing, .89 for encoding errors, and .80 for unknown errors, which indicate that the coding was reliable (Landis & Koch, 1977).

3.4. Statistical analyses

To investigate the relationship between error types and task types, a chi-square test of independence was conducted on the basis of the students’ responses. Because these responses are nested within students, a chi-square with a Rao-Scott adjustment for clustered data in the R survey package was used (Lumley, 2004, 2012).

For studying the relationship of student performance and error type, we applied a Rasch analysis to obtain scale scores of the students’ performance. The reason for choosing this analysis is that it can take into account an incomplete test design (different students got different test booklets with a different set of tasks). A partial credit model was specified in ConQuest (Wu, Adams, Wilson, & Haldane, 2007). The scale scores were estimated within this item response model by weighted likelihood estimates (Warm, 1989) and were categorized into four almost equally distributed performance levels where Level 1 indicates the lowest performance and Level 4 the highest performance. To test whether the frequency of a specific error type differed between performance levels, we applied an analysis of variance based on linear mixed models (Bates, Maechler, & Bolker, 2011). This analysis was based on all responses where an error could be coded and treated the nesting of task responses within students by specifying a random effect for students.

4. Results

4.1. Overview of the observed types of errors

In total, we had 4707 possible responses (number of tasks done by all students in total) which included 2472 correct responses (53%), 1532 incorrect responses (33%), i.e., no credit or partial credit, and 703 missing responses (15%). The error analysis was carried out for the 1532 incorrect responses. The analysis of these responses, based on the multiple coding, revealed that 1718 errors were made of which 38% were comprehension errors and 42% were transformation errors. Mathematical processing errors were less frequently made (17%) and encoding errors only occurred a few times (3%) (see Table 4).

Table 4. Frequencies of error types

Type of error	N	%
Comprehension (C)	653	38
Transformation (T)	723	42
Mathematical processing (M)	291	17
Encoding (E)	51	3
Total of observed errors	1718[a]	100

[a] Because of multiple coding, the total of observed errors exceeds the number of incorrect responses (i.e. n = 1532). In total we had 13 coding categories (including combinations of error types); the six most frequently coded categories were C, CM, CT, M, ME, and T.

4.1.1. Observed comprehension errors

When making comprehension errors, students had problems with understanding the meaning of a task. This was because they misunderstood the instruction or a particular keyword, or they had difficulties in using the correct information. Errors in selecting information included half of the 653 comprehension errors and indicate that students had difficulty in distinguishing between relevant and irrelevant information provided in the task or in gathering required information which was not provided in the task (see Table 5).

Table 5. Frequencies of sub-types of comprehension errors

Sub-type of comprehension error	n	%
Misunderstanding the instruction	227	35
Misunderstanding a keyword	100	15
Error in selecting information	326	50
Total of observed errors	653	100

Figure 2 shows an example of student work which contains an error in selecting information. The student had to solve the *Staircase* task, which was about finding the height of each step of a staircase consisting of 14 steps.

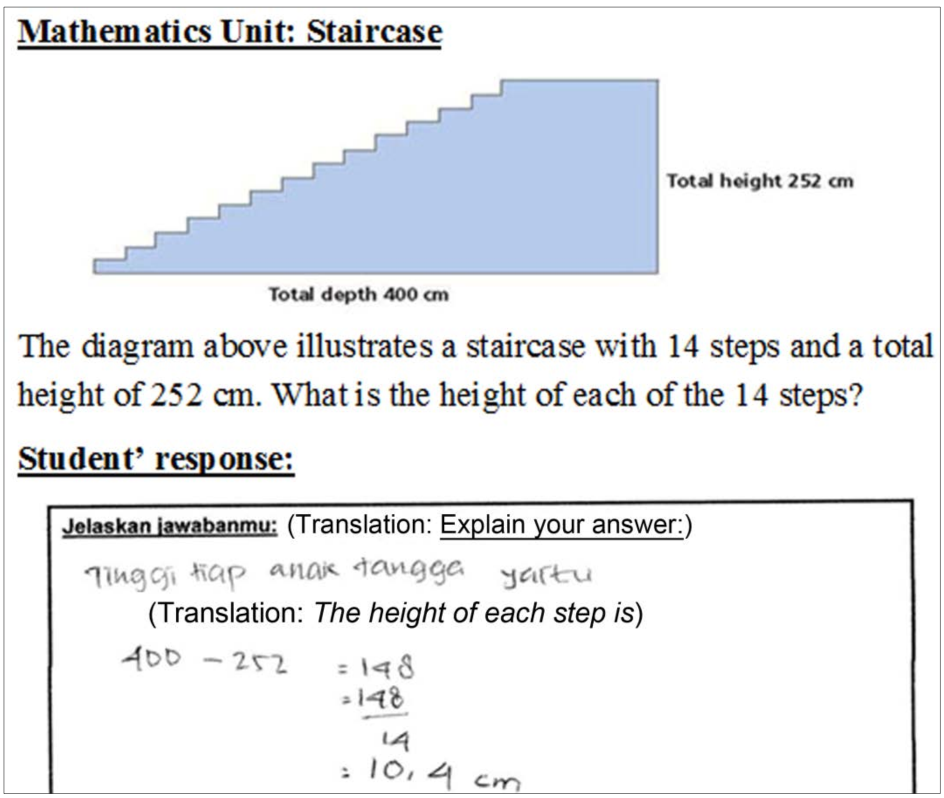


Figure 2. Example of *comprehension error*

The student seemed to have deduced correctly that the height covered by the staircase had to be divided by the number of steps. However, he did not divide the total height 252 (cm) by 14. Instead, he took the 400 and subtracted 252 from it and then he divided the result of it, which is 148, by 14. So, the student made a calculation with the given total depth of the staircase, though this was irrelevant for solving the task.

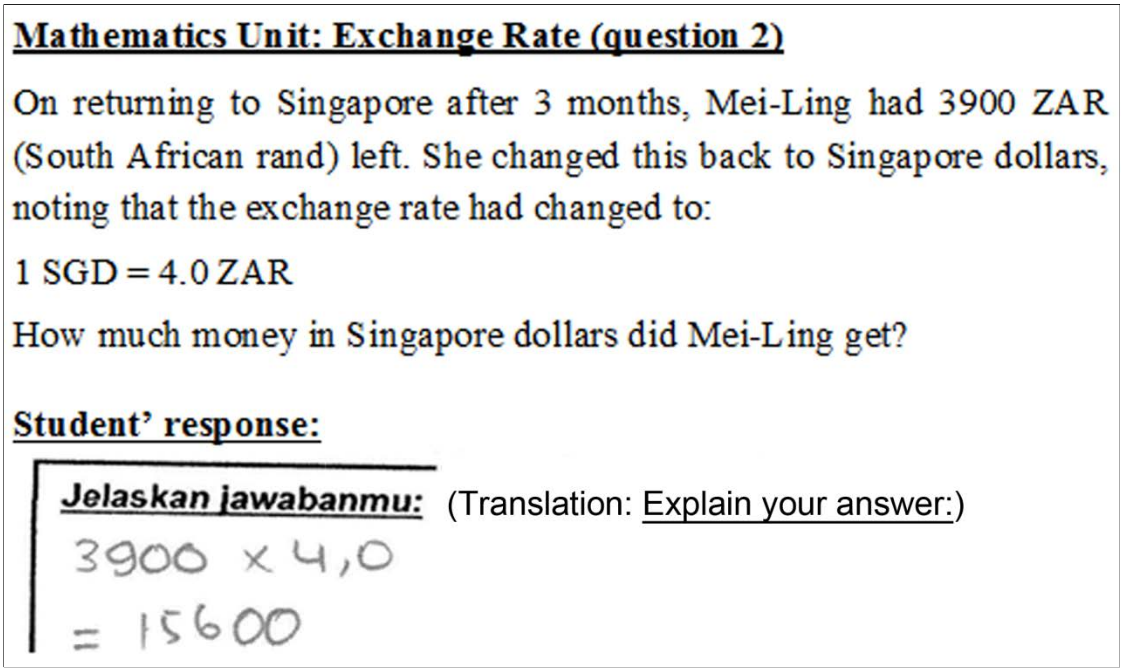
4.1.2. Observed transformation errors

Within the transformation errors, the most dominant sub-type was using a wrong mathematical operation or concept. Of the 723 transformation errors, two thirds belonged to this sub-type (see Table 6).

Table 6. Frequencies of sub-types of transformation errors

Sub-type of transformation error	n	%
Procedural tendency	90	12
Taking too much account of the context	56	8
Wrong mathematical operation/concept	489	68
Treating a graph as a picture	88	12
Total of observed errors	723	100

Figure 3 shows the response of a student who made a transformation error. The task was about the concept of direct proportion situated in the context of money exchange. The student was asked to change 3900 ZAR to Singapore dollars with an exchange rate of 1 SGD = 4.0 ZAR and chose the wrong mathematical procedure for solving this task. Instead of dividing 3900 by 4.0, the student multiplied 3900 by 4.0.



Mathematics Unit: Exchange Rate (question 2)

On returning to Singapore after 3 months, Mei-Ling had 3900 ZAR (South African rand) left. She changed this back to Singapore dollars, noting that the exchange rate had changed to:

1 SGD = 4.0 ZAR

How much money in Singapore dollars did Mei-Ling get?

Student' response:

Jelaskan jawabanmu: (Translation: Explain your answer:)

3900 × 4,0

= 15600

Figure 3. Example of *transformation error*

4.1.3. Observed mathematical processing errors

Mathematical processing errors correspond to students' failure in carrying out mathematical procedures (for an overview of these errors, see Table 3). This type of errors is mostly dependent on the mathematical topic addressed in a task. For example, errors in interpreting a graph do not occur when there is no graph in the task. Consequently, we calculated the frequencies of the sub-types of mathematical processing errors only for the related tasks, i.e. tasks in which such errors may occur (see Table 7).

Table 7. Frequencies of sub-types of mathematical processing errors

Sub-type of mathematical processing error	Related tasks	All errors in related tasks	Mathematical processing errors in related tasks	
	n	n	n	%
Algebraic error	8	243	33	14
Arithmetical error	20	956	94	10
Error in interpreting graph	6	155	43	28
Measurement error	1	74	15	20
Error related to improper use of scale	1	177	49	28
Unfinished answer	26	1125	79	7

An example of a mathematical processing error is shown in Figure 4. The task is about finding a man's pace length (*P*) by using the formula $\frac{n}{P}=140$ in which *n*, the number of steps per minute, is given. The student correctly substituted the given information into the formula and came to $\frac{70}{P}=140$. However, he took 140 and subtracted 70 instead of dividing 70 by 140. This response indicates that the student had difficulty to work with an equation in which the unknown value was the divisor and the dividend is smaller than the quotient.

4.1.4. Observed encoding errors

Encoding errors were not divided in sub-types. They comprise all errors that have to do with the students' inability to interpret a mathematical answer as a solution that fits to the real-world context of a task. As said only 3% of the total errors belonged to this category. The response of the student in Figure 4, discussed in the previous section, also contains an encoding error. His answer of 70 is, within the context of this task, an answer that does not make sense. A human's pace length of 70 meter is a rather unrealistic answer.

Mathematics Unit: Walking

For men, the formula, $\frac{n}{P} = 140$, gives an approximate relationship between n and P where,
n = number of steps per minute, and
P = pacelength in meters

Question 1:
If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

Students' responses:

Jelaskan jawabanmu: (Translation: Explain your answer:)

$\frac{n}{P} = 140$

$\frac{70}{P} = 140$

$P = 140 - 70$

$= 70$

Figure 4. Example of *mathematical processing error* and *encoding error*

4.2. The relation between the types of errors and the types of tasks

In agreement with the PISA findings (OECD, 2009a), we found that the reproduction tasks were the easiest for the students, whereas the tasks with a higher cognitive demand, the connection and the reflection tasks, had a higher percentage of completely wrong answers (which means no credit) (see Figure 5).

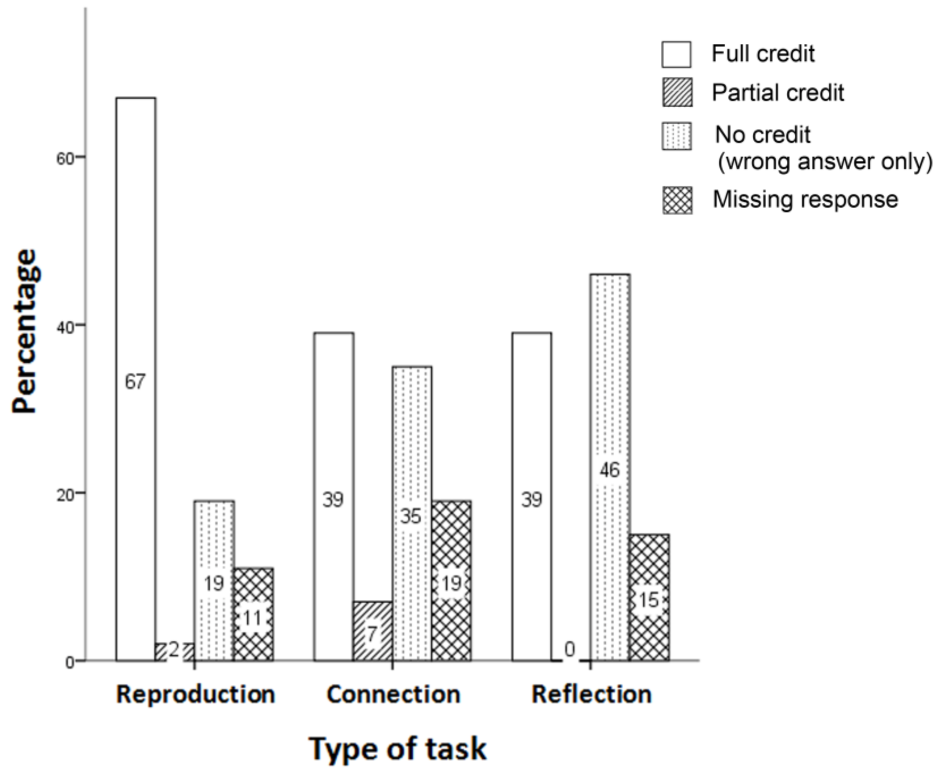


Figure 5. Percentage of full credit, partial credit, no credit, and missing response per task type

To investigate the relation between the error types and the tasks types we performed a chi-square test of independence based on the six most frequently coded error categories (see the note in Table 4). The test showed that there was a significant relation (χ^2 (10, N = 1393) = 91.385, p < .001). Furthermore, we found a moderate association between the error types and the types of tasks (Phi coefficient = .256; Cramer's V = .181).

When examining the proportion of error types within every task type, we found that in the reproduction tasks, mostly comprehension errors (37%) and transformation errors (34%) were made (see Table 8). Also in the connection tasks students made mostly comprehension errors (41%) and transformation errors (43%). However, in the reflection tasks mostly transformation errors (66%) were made. Furthermore, the analysis revealed that out of the three types of tasks the connection tasks had the highest average number of errors per task.

Table 8. Error types within task type

Type of task	Tasks n	Incorrect responses n[a]	Comprehension error n	Comprehension error %	Transformation error n	Transformation error %	Mathem. Processing error n	Mathem. Processing error %	Encoding error n	Encoding error %	Total errors n	Total errors %[a]	Total errors average number of errors per task
Reproduction	15	518	211	37	194	34	136	24	23	4	564	99	38
Connection	15	852	404	41	424	43	140	14	28	3	996	101	66
Reflection	4	162	38	24	105	66	15	9	0	0	158	99	40
Total	34	1532	653		723		291		51		1718[b]		

[a] Because of rounding off, the total percentages are not equal to 100%

[b] Because of multiple coding, the total errors exceeds the total number of incorrect responses

4.3. The relation between the types of errors, the types of tasks, and the students' performance level

When testing whether students on different performance levels differed with respect to the error types they made, we found, for all task types together, that the low performing students (Level 1 and Level 2) made more transformation errors than the high performing students (Level 3 and Level 4) (see Figure 6)[3]. For the mathematical processing errors the pattern was opposite. Here we found more errors in the high performing students than in the low performing students. With respect to the comprehension errors there was not such a difference. The low and high performing students made about the same amount of comprehension errors.

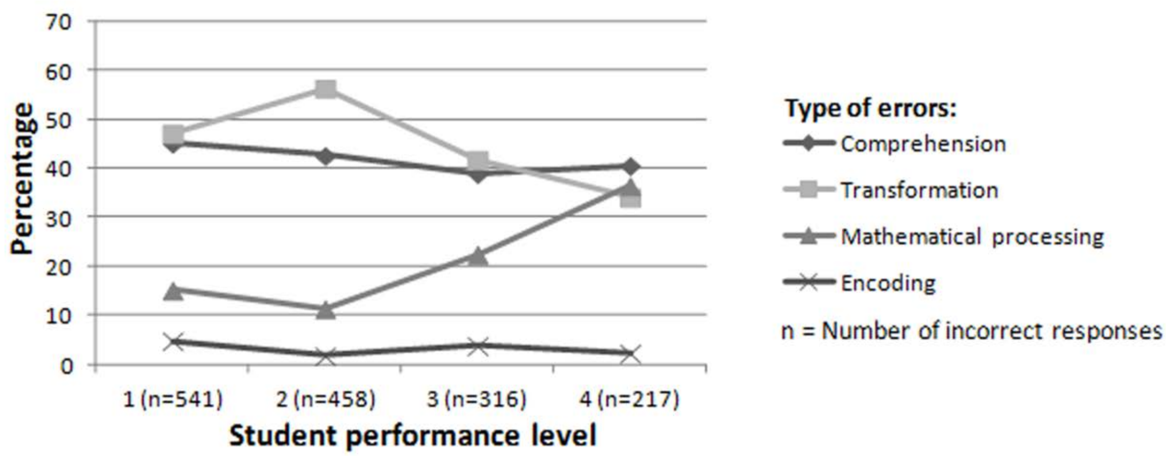


Figure 6. Types of error in *all tasks* for different performance levels

A nearly similar pattern of error type and performance level was also observed when we zoomed in on the connection tasks (see Figure 7).

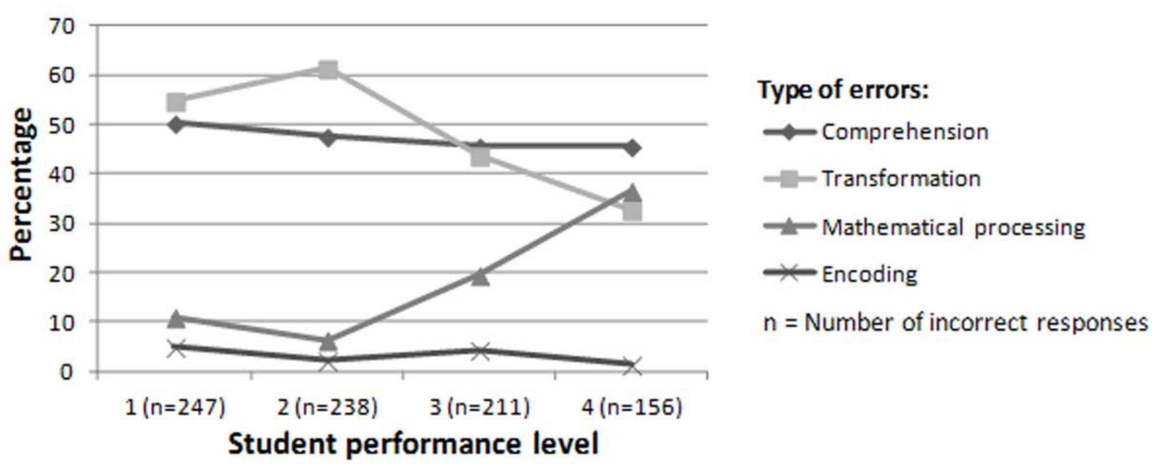


Figure 7. Types of errors in *connection tasks* for different performance levels

For the reproduction tasks (see Figure 8) the pattern was also quite comparable. The only difference was that the high performing students made less comprehension errors than the low performing students.

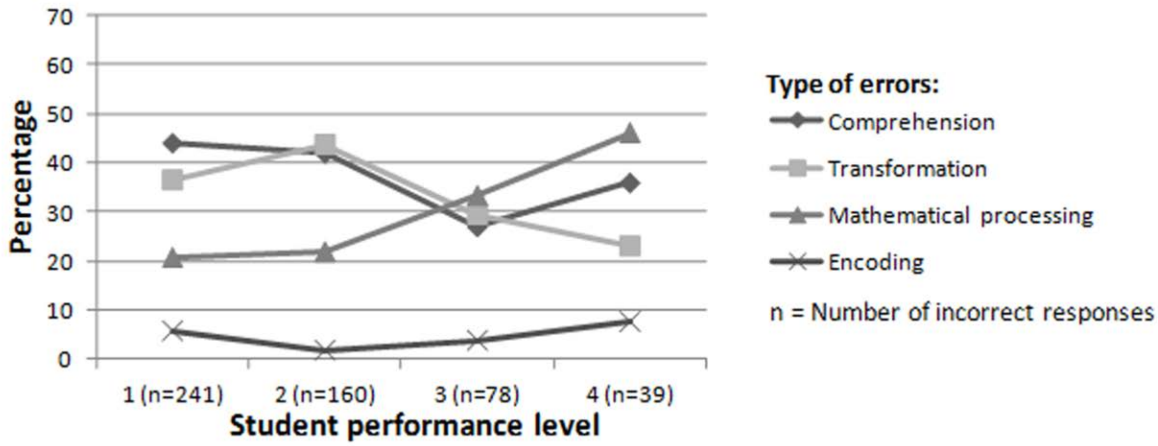


Figure 8. Types of errors in *reproduction tasks* for different performance levels

For the reflection tasks we found that the high performing students made more mathematical processing errors than the low performing students (see Figure 9). For the other error types, we did not found remarkable differences across student performance levels.

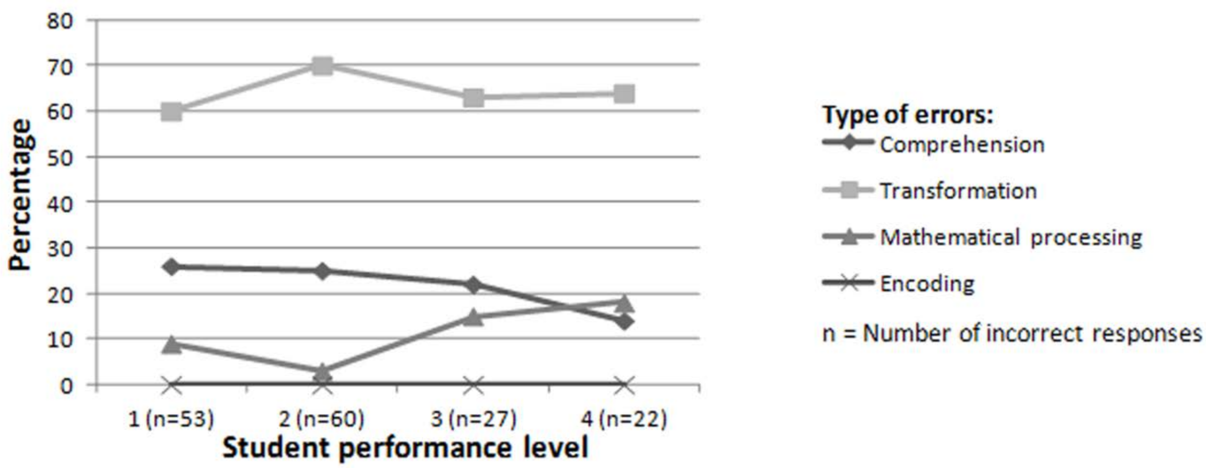
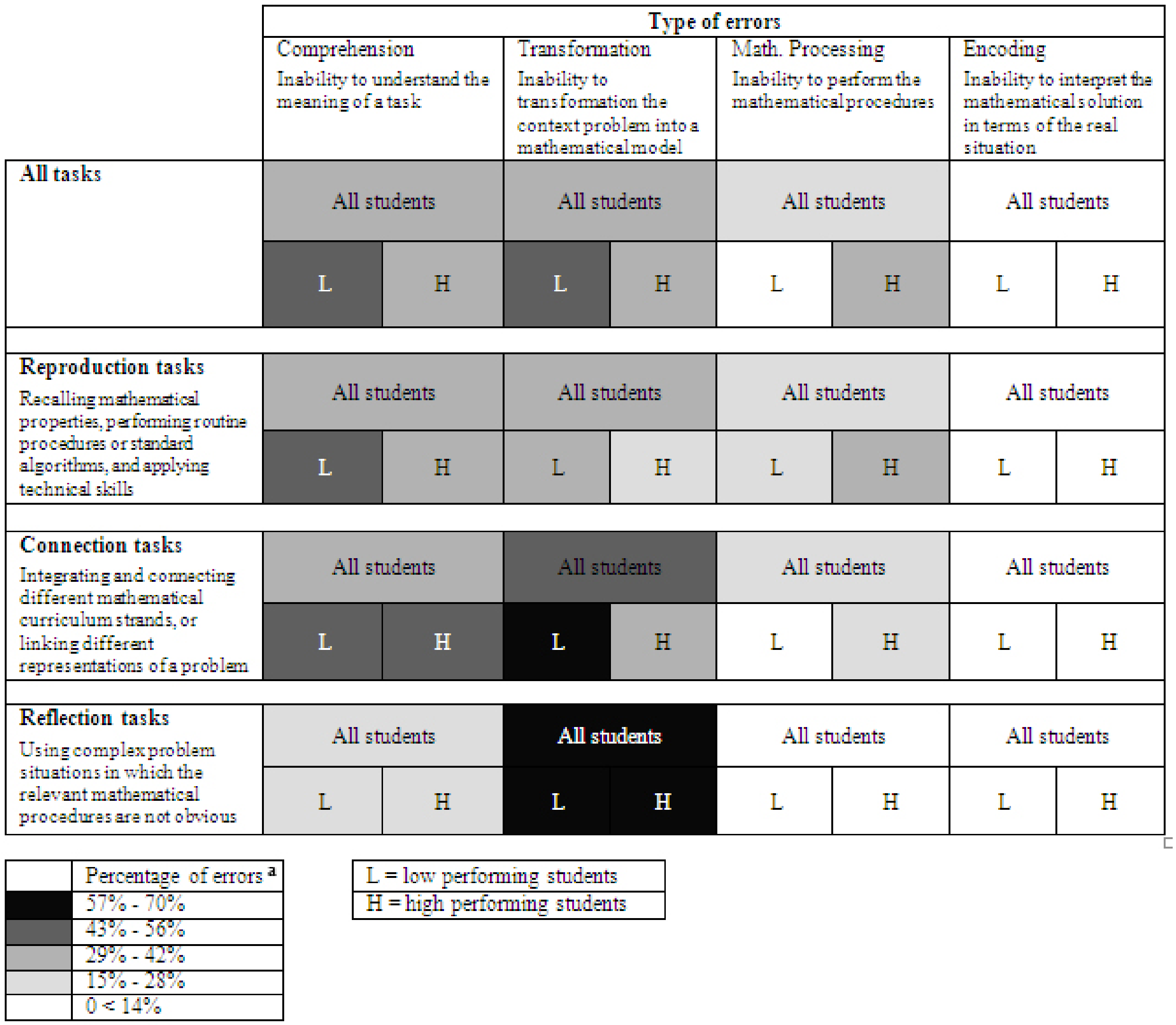


Figure 9. Types of errors in *reflection tasks* for different performance levels

5. Conclusions and discussion

The present was study aimed at getting a better understanding of students' errors when solving context-based mathematics tasks. Figure 10 summarizes our findings regarding the types of errors Indonesian nine- and ten-graders made when solving these tasks. Out of the four types of errors that were derived from Newman (1977, 1983), we found that comprehension and transformation errors were most dominant and that students made less errors in mathematical processing and in the interpretation of the mathematical solution in terms of the original real-world situation. This implies that the students in our sample mostly experienced difficulties in the early stages of solving context-based mathematics tasks as described by Blum and Leiss (cited in Maass, 2010) and PISA (OECD, 2003b), i.e. comprehending a real-world problem and transforming it into a mathematical problem.



[a] The percentages of errors range from 0% to 70% because the maximum percentage of students' errors is 66%. This range is divided into five equal levels. In the figure, these levels are represented in cells with different degrees of shading.

Figure 10. Summary of research findings

Furthermore, within the category of comprehension errors our analysis revealed that most students were unable to select relevant information. Students tended to use all numbers given in the text without considering their relevance to solving the task. This finding provides a new perspective on students' errors in understanding real-world tasks because previous studies (Bernardo, 1999; Cummins et al., 1988) mainly concerned students' errors in relation to the language used in the presentation of the task. For example, they found that students had difficulties to understand the mathematical connotation of particular words.

The above findings suggest that focusing on the early stages of modeling process or mathematization might be an important key to improve students' performance on context-based tasks. In particular for comprehending a real-world problem, much attention should be given to tasks with lacking or superfluous information in which students have to use their daily-life knowledge or have to select the information that is relevant to solve a particular task.

A further focus of our study was the relation between the type of tasks and the types of students' errors. In agreement with the PISA findings (OECD, 2009a) we found that the cognitive demands of the tasks are an important factor influencing the difficulty level of context-based tasks. Because we did not only look at the correctness of the answers but also to the errors made by the students, we could reveal that most errors were not made in the reflection tasks, i.e. the tasks with the highest cognitive demand, but in the connection tasks. It seems as if these tasks are most vulnerable for mistakes. Furthermore, our analysis revealed that in the reflections tasks students made less comprehension errors than in the connection tasks and the reproduction tasks. One possible reason might be that most reflection tasks used in our study did not provide either more or less information than needed to solve the task. Consequently, students in our study did hardly have to deal with selecting relevant information.

Regarding the relation between the types of errors and the student performance level, we found that generally the low performing students made more comprehension errors and transformation errors than the high performing students. For the mathematical processing errors the opposite was found. The high performing students made more mathematical processing errors than the low performing students. A possible explanation for this is that low performing students, in contrast to high performing students, might get stuck in the first two stages of solving context-based mathematics tasks and therefore are not arriving at the stage of carrying out mathematical procedures. These findings confirm Newman's (1977) argument that the error types might have a hierarchical structure: failures on a particular step of solving a task prevents a student from progressing to the next step.

In sum, our study gave a better insight into the errors students make when solving context-based tasks and provided us with indications for improving their achievement. Our results signify that paying more attention to comprehending a task, in particular selecting relevant data, and to transforming a task, which means identifying an adequate mathematical procedure, both might improve low performing students' ability to solve context-based tasks. For the high performing students, our results show that they may benefit from paying more attention to performing mathematical procedures.

However, when making use of the findings of our study this should be done with prudence, because our study clearly has some limitations that need to be taken into account. What we found in this Indonesian sample does not necessarily apply to students in other countries with different educational practices. In addition, the classification of task types – reproduction, connection and reflection – as determined in the PISA study might not always be experienced by the students in a similar way. For example, whether a reproduction task is a reproduction task for the students also depends on their prior knowledge and experiences. For instance, as described by Kolovou, Van den Heuvel-Panhuizen, and Bakker (2009), students who have not learned algebra cannot use a routine algebraic procedure to split a number into several successive numbers (such as splitting 51 into 16, 17, and 18). Instead, they might use an informal reasoning strategy with trial-and-error to solve it. In this case, for these students the task is a connection task, whereas for students who have learned algebra it might be a reproduction task.

Notwithstanding the aforementioned limitations, the results of our study provide a basis for further research into the possible causes of students' difficulties in solving context-based mathematics tasks. For finding causes of the difficulties that students encounter, in addition to analyzing students' errors, it is also essential to examine what *opportunities-to-learn* students are offered in solving these kinds of tasks. Investigating these learning opportunities will be our next step in the CoMTI project.

Notes

1. In PISA (see OECD, 2003b, p. 52) a 'mathematics unit' consists of one or more questions which can be answered independently. These questions are based on the same context which is generally presented by a written description and a graphic or another representation.

2. For example, of the 33 provinces in Indonesia the province of Yogyakarta occupied the 6th place in the national examination in the academic year of 2007/2008 (Mardapi & Kartowagiran, 2009).
3. The diagram in Figure 6 (and similarly in Figure 7, Figure 8, and Figure 9) can be read as follows: the students at Level 1 gave in total 541 incorrect responses of which 45% contained comprehension errors, 47% transformation errors, 15% mathematical processing errors, and 5% encoding errors. Because of multiple coding, the total percentage exceeds 100%.

Acknowledgement

This research was supported by the Indonesian Ministry of National Education under the Better Education through Reformed Management and Universal Teacher Upgrading (BERMUTU) Project.

References

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42.

Bernardo, A. B. I. (1999). Overcoming obstacles to understanding and solving word problems in mathematics. *Educational Psychology, 19*(2), 149-163.

Biembengut, M. S. (2007). Modelling and applications in primary education. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 451-456). New York: Springer.

Boaler, J. (1993a). Encouraging transfer of 'school' mathematics to the 'real world' through the integration of process and content, context and culture. *Educational Studies in Mathematics, 25*(4), 341-373.

Boaler, J. (1993b). The role of contexts in the mathematics classroom: Do they make mathematics more real? *For the Learning of Mathematics, 13*(2), 12–17.

Clarkson, P. C. (1991). Language comprehension errors: A further investigation. *Mathematics Education Research Journal, 3*(2), 24-33.

Clements, M. A. (1980). Analyzing children's errors on written mathematical task. *Educational Studies in Mathematics, 11*(1), 1-21.

De Lange, J. (1987). *Mathematics, insight and meaning*. Utrecht: OW & OC, Rijksuniversiteit Utrecht.

Graumann, G. (2011). Mathematics for problem in the everyday world. In J. Maasz & J. O'Donoghue (Eds.), *Real-world problems for secondary school mathematics students: case studies* (pp. 113-122). Rotterdam: Sense Publishers.

Gravemeijer, K.P.E. & Doorman, L.M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics, 39*(1-3), pp. 111-129.

Greer, B., Verschaffel, L., & Mukhopadhyay, S. (2007). Modelling for life: Mathematics and children's experience. In W. Blum, P. L. Galbraith, H.-W. Henn & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 89-98). New York: Springer.

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy, 24*(1), 23-37.

Henning, P. H. (2004). Everyday cognition and situated learning. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology : A project of the Association for Educational Communications and Technology (2nd)* (pp. 143-168). New Jersey: Lawrence Erlbaum Associates, Inc.

Kolovou, A., Van den Heuvel-Panhuizen, M., & Bakker, A. (2009). Non-routine problem solving tasks in primary school mathematics textbooks – A Needle in a Haystack. *Mediterranean Journal for Research in Mathematics Education, 8*(2), 31-68.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60*(1), 1-64.

Liang, X. (2010). Assessment use, self-efficacy and mathematics achievement: Comparative analysis of PISA 2003. Data of Finland, Canada and the USA. *Evaluation & Research in Education, 23*(3), 213-229.

Lumley, T. (2004). Analysis of complex survey samples. Journal *of Statistical Software, 9*(1), 1-19.

Lumley, T. (2012). *Survey: Analysis of complex survey samples. R Package version 3.28-1*.

Maass, K. (2010). Classification scheme for modelling tasks. *Journal für Mathematik-Didaktik, 31*(2), 285-311.

Mardapi, D., & Kartowagiran, B. (2009). *Dampak Ujian Nasional*. Yogyakarta: Universitas Negeri Yogyakarta.

Newman, M. A. (1977). An analysis of sixth-grade pupils' errors on written mathematical tasks. *Victorian Institute for Educational Research Bulletin*, *39*, 31-43.

Newman, M. A. (1983). *Strategies for diagnosis and remediation*. Sydney: Harcourt Brace Jovanovich.

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics*. New York: Cambridge University Press.

Nunez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics, 39*(1-3), 45-65.

OECD. (2003a). *Literacy skills for the world of tomorrow. Further results from PISA 2000*. Paris: OECD.

OECD. (2003b). *The PISA 2003 assessment framework – Mathematics, reading, science, and problem solving knowledge and skills*. Paris: OECD.

OECD. (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.

OECD. (2005). *PISA 2003. Technical report*. Paris: OECD.

OECD. (2007). *PISA 2006: Science competencies for tomorrow's world*. Paris: OECD.

OECD. (2009a). *Learning mathematics for life. A view perspective from PISA*. Paris: OECD.

OECD. (2009b). *Take the test. Sample questions from OECD's PISA assessments*. Paris: OECD.

OECD. (2010). *PISA 2009 results: What students know and can do. Student performance in reading, mathematics, and science (Vol. I)*. Paris: OECD.

Pusat Kurikulum. (2003). *Kurikulum 2004. Standar Kompetensi mata pelajaran matematika Sekolah Menengah Pertama dan Madrasah Tsanawiyah*. Jakarta: Departemen Pendidikan Nasional.

Roth, W.-M. (1996). Where is the context in contextual word problems? Mathematical practices and products in grade 8 students' answers to story problems. *Cognition and Instruction, 14*(4), 487-527.

Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Educational Change, 8*(4), 349-357.

Schwarzkopf, R. (2007). Elementary modelling in mathematics lessons: The interplay between "real-world" knowledge and "mathematical structures". In W. Blum, P. L. Galbraith, H.-W.

Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 209-216). New York: Springer.

Van den Heuvel-Panhuizen, M. (1996). *Assessment and Realistic Mathematics Education*. Utrecht: CD-β Press, Center for Science and Mathematics Education.

Van den Heuvel-Panhuizen, M. (2005). The role of context in assessment problems in mathematics. *For the Learning of Mathematics*, *25*(2), 2-9, and 23.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427-450.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). ACER conquest version 2.0. Mulgrave: ACER Press.

Yore, L. D., Anderson, J. O., & Hung Chiu, M. (2010). Moving PISA results into the policy arena: Perspectives on knowledge transfer for future considerations and preparations. *International Journal of Science and Mathematics Education, 8*(3), 593-603.

Yusuf, S. (2004). *Analisis tes PISA. Literasi membaca, matematika, dan sains*. Jakarta: Pusat Penelitian Pendidikan. Departemen Pendidikan Nasional.