Images of Numeracy

Investigating the effects of visual representations of problem situations in contextual mathematical problem solving

Kees Hoogland

Images of Numeracy

Investigating the effects of visual representations of problem situations in contextual mathematical problem solving

Kees Hoogland

This doctoral thesis was financially supported by SLO and facilitated by Eindhoven School of Education (Eindhoven University of Technology). A major part of the research was carried out in the context of the programme Onderwijsbewijs II of the Ministry of Education, Culture and Science.

A catalogue record is available from the Eindhoven University of Technology Library.
ISBN: 978-90-386-4156-0
Print: T-Point Print
Cover design: Simone Analbers (SLO)
Cover photos: Bigstock™
© 2016 Kees Hoogland

Images of Numeracy

Investigating the effects of visual representations of problem situations in contextual mathematical problem solving

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op dinsdag 8 november 2016 om 16:00 uur

door

Cornelis Peter Hoogland

geboren te Vlaardingen

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. P.J. den Brok
promotor:	prof.dr. B.E.U. Pepin
copromotor:	dr. A. Bakker (Universiteit Utrecht)
leden:	prof.dr. P.H.M. Drijvers (Universiteit Utrecht)
	prof.dr. K.P.E. Gravemeijer
	prof.dr. J. de Koning (Erasmus Universiteit Rotterdam)
	prof.dr. J.S.H. van Leeuwaarden
	prof.dr. L. Verschaffel (Katholieke Universiteit Leuven)

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Contents

Chapter 1. Problem statement and conceptual framework

1.1 Problem statement and research question: the rationale

In mathematics education the focus has been shifting from a strong emphasis on mathematical procedures towards more attention to problem solving and creating a problem-solving attitude in students. This holds particularly true for the domains of mathematics education that deal with numeracy and mathematical literacy. In these particular branches of mathematics education the central theme is how people can learn to use mathematics to cope with quantitative problems from daily life (Kilpatrick, 1996; Niss, 1996; OECD, 1999; Toner, 2011).

Contextual mathematical problems play an important role in the realisation of mathematics education with a focus on problem solving. A contextual mathematical problem consists of a representation of a real-life problem situation and a question. The dominant form of representing the problem situation is describing the problem situation in so-called “word problems” (Lave, 1992; Verschaffel, Greer, & De Corte, 2000; Verschaffel, Greer, Van Dooren, & Mukhopadhyay, 2009). This genre of problems is used worldwide in mathematics education that pursues mathematical problem-solving goals. However, in the literature many difficulties are reported about the use of word problems in the mathematics classroom (Dewolf, Van Dooren, Ev Cimen, & Verschaffel, 2014; Gellert & Jablonka, 2009; Inoue, 2009; Reusser & Stebler, 1997; Roth, 2009; Verschaffel, De Corte, & Borghart, 1997; Verschaffel, De Corte, & Lasure, 1994). Students and teachers alike (Depaepe, De Corte, & Verschaffel, 2010) persistently see word problems as procedural exercises in disguise, and hence students develop an “answer-getting” mindset, that is, they pick numbers from the text and perform an operation upon them, without making sense of the problem and without taking realistic constraints into consideration. As sense-making is seen as crucial in problem-solving processes (Schoenfeld, 1992), the answer-getting mindset is seen as counterproductive to good problem solving. As a result of an answer-getting mindset many inadequate or nonsensical solutions can be observed in students’ work. Schoenfeld (1991) calls this “suspension of sense-making”, and it is seen as a serious challenge in mathematics education, along with the language difficulties many students encounter with the verbal, sometimes verbose, description of the problem situations.

Over the years, mathematics educators have suggested various measures to counteract or prevent the suspension of sense-making and the calculational approach that has been witnessed when students solve word problems. Most suggestions are in the direction of making the problem-solving setting feel more authentic for students. Examples include changing the setting in which the problem is posed, giving instructions to take reality into account, and adding pictorial elements to the problems (Cooper & Harries, 2002, 2003; DeFranco & Curcio, 1997; Dewolf et al., 2014; Dewolf, Van Dooren, Hermens, & Verschaffel, 2015; Säljö, Riesbeck, & Wyndhamn, 2009; Wyndhamn & Säljö, 1997). Most of these measures have yielded modest results in improving the performance of students. These studies, however, have provided indications that the ways in which problem situations are represented in contextual mathematical problems have an effect on students’ sense-making, behaviour, motivation, and results.

In the research project on which this thesis is based, we regarded the dominance of verbal descriptions of problem situations as typical, and looked for alternatives by representing the problem situations with a depictive representation: for example, using photographs representing the problem situation as close as possible to the represented actual problem in real life. We wanted to know whether a specific change in the representation of the problem situation, from descriptive to mainly depictive, was likely to improve students’ performance in making sense of the problem. Hence we singled out students’ performance as a measure for the change in students’ behaviour as a result of changing the representation of the problem situation from descriptive to depictive. We investigated whether this change would improve the chance that students will achieve an adequate solution to the posed problem. The central research question for this dissertation is the following:

> In presenting contextual mathematical problems, what is the effect on student performance of changing a descriptive representation of the problem situation to a mainly depictive one?

We predicted a better performance of students on contextual mathematical problems with a mainly depictive problem situation, as an indicator of students making more sense of a depictive representation of a problem situation than of a descriptive one. A further rationale for this research question was the connection of the research question with a broader issue in education: how to represent problems from real life in educational settings or for educational purposes, and the effect of this on students' performance? Using a descriptive representation of the problem situation has a long history not only in mathematics education but also in other subject matters. Using depictive representations of problem situations could pave the way for other representations, including multimedia, in a more realistic and authentic way, which is likely to make it easier for students to make sense of the situation.

We regarded this approach as challenging and, from our review of the literature, under-researched in mathematics education research. We designed a controlled randomised trial to measure the effect of changing the representation of the problem situation on students' performance. In 2010 this methodological approach to the research question was awarded a grant from the programme "Onderwijsbewijs II", instigated by the Dutch Ministry of Education, Culture and Sciences to encourage evidence-based research in education.

In the following section, we elaborate on this rationale from different theoretical perspectives and with references to the relevant literature.

1.2 Conceptual framework

1.2.1 Goals of mathematics education

In this section we argue that the focus of mathematics education, especially in secondary and vocational education, has shifted and that more attention is now given to mathematical reasoning and problem solving, and to making sense of quantitative situations. Mathematical reasoning is an essential part of many human cognitive activities used in education, in the sciences, in work environments, and in daily life. Although many mathematical operations and techniques can be executed by technological tools, mathematical reasoning and mathematical skills are useful for people on a daily basis in their professional and personal lives. High level mathematical knowledge and skills are used in technical, social and economic sciences in a wide variety of areas, such as operations research, medical imaging, economic modelling, navigation, agriculture, manufacturing, security, and communication, to name but a few. There seems to be a broad international consensus in the field of mathematics education to pursue goals that reflect the necessity to prepare students for future problem solving in practical day-to-day situations (Kilpatrick, 1996; Niss, 1996; OECD, 1999; Toner, 2011).

In the Netherlands, an analysis has recently been commissioned by the Platform Wiskunde Nederland (PWN) [Dutch Organisation for Mathematics], and conducted by Deloitte on the role and value of mathematics in Dutch society. In the report *Mathematical sciences and their value for the Dutch economy* (Deloitte, 2014) the authors claim that in approximately 900,000 full-time equivalent jobs, highly educated employees use some sort of mathematical sciences in their work – including, for example, employees in economics-oriented jobs, bankers, and medical professionals – contributing to up to a quarter of the national income. At the personal level, mathematics is used by individuals to participate in the "mathematised" society to solve a wide array of quantitative problems in everyday life, such as they occur in shopping, rent and mortgage issues, medication, and planning, for example. In the case of using mathematical knowledge and skills in everyday life situations, we speak more commonly about mathematical literacy. In the last few decades numbers have come to be generated more easily by the ubiquitous availability of spreadsheets and the accompanying visual representations of (big) data gathered. Related to that, one can observe that those (visualised) numbers also play an important role in policy decision making in both commerce and politics. More than ever this "mathematisation of society" (Gellert & Jablonka, 2007) calls upon mathematics education to provide students with the appropriate knowledge and skills to participate adequately and autonomously in this "mathematised world" and to make sense of the ubiquitous quantitative problems they encounter. Consequently, in many countries the discussion on appropriate and relevant goals for mathematics education is on the political agenda.

In the Netherlands the discussion on modernisation of the current curricula and on the development of new and relevant goals for education is ongoing (Platform Onderwijs 2032, 2016). In this discussion a perspective on different goals for education was provided by Biesta (2010, 2013) among others. From his perspective, goals for education can be categorised in three major strands: qualification, socialisation, and subjectivation/personal development. In the case of mathematics education this can be translated to a broad array of goals: qualifications for the use of mathematics in future studies and jobs; socialisation in becoming a responsible and reflective citizen who can understand the quantitative aspects of society and cope with them (OECD, 2012); and personal development for which mathematics education could give students a critical attitude towards the use of mathematics and numbers in their personal life and confidence to engage autonomously in quantitative problems in so many aspects of daily life that could be solved with mathematics knowledge and skills. Common to all three strands of goals is the practical and relevant use of mathematical knowledge and skills by individuals. Consequently, mathematics education has to deal with the intricate relation between the formation of mathematical knowledge and skills and the actual and real situations where these knowledge and skills are applied.

The tendency to focus on the practical use of mathematical knowledge is reflected in various concepts used in the literature on goals for mathematics education. The concept of numeracy (Gal, Groenestijn, Manly, Schmitt, & Tout, 2003; Tout & Gal, 2015), which is also used in the title of this dissertation, was originally an abbreviation of numerical literacy, indicating that it is a part of the basic practical knowledge and skills people should have to cope with modern society, comparable with prose literacy, document literacy, financial literacy, and science literacy (OECD, 2012). Next to the concept of numeracy, one speaks also of mathematical literacy (Jablonka, 2003, 2015; Jablonka & Niss, 2014), quantitative literacy (Madison & Steen, 2008; Steen, 2001), matheracy (D'Ambrosio, 1997), functional mathematics (Forman & Steen, 1999), and a little more remotely, ethnomathematics (D'Ambrosio, 1997; Gerdes, 1994), each with its own specific interpretations. What they all have in common is that they look at individuals trying to understand and make sense of situations in the world around them, and thereby using some form of mathematical reasoning, knowledge, and skills.

All of the most recent approaches for defining numeracy fit into what Maguire and O'Donoghue (2002) would call an *integrative* definition of numeracy. In this definition numeracy is viewed as a complex, multifaceted, and sophisticated construct incorporating each individual’s mathematics, communication, cultural, social, emotional, and personal aspects in context. These more integrative approaches to numeracy have become influential over the last few decades, as illustrated by projects that define standards for numeracy instructional content such as the PISA, ALL, and PIAAC (Gal et al., 2003; OECD, 2013a, 2013c, 2014; PIAAC Numeracy Expert Group, 2009; Tsatsaroni & Evans, 2014). Numeracy is defined by Gal et al. (2003) in the ALL Survey as follows:

> the knowledge and skills required to manage and respond effectively to the mathematical demands of diverse situations; in addition, numerate behaviour is observed when people manage a situation or solve a problem in a real context; it involves responding to information about mathematical ideas that may be represented in a range of ways; it requires the activation of a range of enabling knowledge, factors, and processes. (p. 142)

The content definitions in all these international projects on numeracy or mathematical literacy emphasise the need for people to adjust to the increasing technological demands of the knowledge economy. That is seen as the core of all definitions of numeracy, mathematical literacy, or quantitative literacy. This was also the definition of numeracy we used in designing the image-rich contextual mathematical problems for the instrument in our studies (see Chapter 4).

1.2.2 Word problems and sense-making

In our studies we compared students’ performance on two types of problems: word problems that use descriptive representations of the problem situations and image-rich numeracy problems that use mainly depictive representations of the problem situation. This section elaborates on word problems for this purpose. Pictorial or tally representations of reality are almost as old as humankind (Zaslavsky, 1999) and can be found in prehistoric cave paintings in many parts of the world. More recent, but still centuries old, is the use of verbal descriptions of real-life situations as a way to bring real-life problems into the classroom. Representing problem situations with words to be solved with mathematical tools dates back to Fibonacci in his *Liber Abaci* (Leonardo Pisano (Fibonacci), 1202; Sigler, 2002), one of the oldest known examples of the use of word problems.

In mathematics education of the twentieth century, the concept "word problems" has become the label for this genre of contextual mathematical problems. Word problems are an important part of the present investigation into the effect of representations of problem situations on students' performance. According to Verschaffel, Depaepe, and Van Dooren (2014) word problems can be defined as: "verbal descriptions of problem situations wherein one or more questions are raised the answer to which can be obtained by the application of mathematical operations to numerical data available in the problem statement" (p.641).

In word problems both the description of the problem situation and the actual problem statement are presented in words. Many studies deal with the difficulties that occur when descriptive representations of real-life situations are used in mathematics education (Verschaffel et al., 2000; Verschaffel et al., 2009). Word problems are referred to in the Netherlands as *talige contextopgaven* or as *wiskundige contextopgaven* [contextual mathematical problems], or sometimes more disparagingly as *verhaaltjesommen* [mathematical story problems] or *redactiesommen* [composed mathematical problems]. The genre of word problems has become a goal in itself, with the decoding of the description of the problem situation as intended learning outcome, and in that way drifting away from the more fundamental goal of solving the actual problem that is represented by the description.

The literature also reveals the difficulties that are encountered in the use of word problems. Over the past 20 years the use of word problems in assessing mathematical potential for solving real-life problems has been intensively researched. Many research findings mention serious drawbacks of using straightforward word problems to assess accurately students' mathematical potential in problem-solving (Verschaffel et al., 2000; Verschaffel et al., 2009). Numerous studies on student behaviour when solving word problems report on students experiencing “blockages” in the problem-solving process (Galbraith & Stillman, 2006). The most reported behaviour is that students base their analysis and calculations on a rather arbitrary association between certain salient quantitative elements of the problem situation and certain mathematical operations (Thompson, Philipp, Thompson, & Boyd, 1994; Verschaffel et al., 2000). Studies of word problems further show that students tend not to consider the possible constraints imposed by reality (Caldwell, 1995; Cooper & Harries, 2003; Dewolf et al., 2015; Lave, 1992; Reusser & Stebler, 1997; Schoenfeld, 1992; Verschaffel et al., 1994; Verschaffel et al., 2009; Wyndhamn & Säljö, 1997), but rather value more the bare outcome of the calculation than the realism of the outcome. There is strong evidence that students are taught to approach word problems as “school maths” problems and not as problems of authentic life situations. This behaviour of students is reported to be reinforced by teachers’ approaches that tend to foreground the mathematical structure of the problem rather than the contextual aspects (Depaepe et al., 2010).

The use of superficial strategies and dissociation from reality in problem solving with word problems has become known as “suspension of sense-making” (Reusser & Stebler, 1997; Schoenfeld, 1991; Verschaffel et al., 2000). This phenomenon is quite persistent in classroom situations and has generated serious concern as to whether students can show their full potential in solving quantitative real-life problems when confronted with word problems in the classroom or in an assessment setting. Suspension of sense-making in solving word problems is closely connected to a rigid mindset focusing on the result of the possible operations involved, rather than making sense of the situation for solving the problem at hand. This behaviour is also known as a “calculational orientation” (Thompson et al., 1994) which for many decades has been, and to a great extent still is (Madison & Steen, 2008), dominant in mathematics classrooms. Thompson et al. (1994) distinguish between a calculational orientation, where the focus is primarily on procedures and operations with numbers, and a conceptual orientation approach, where the focus is primarily on explaining, reasoning, and interpreting the quantitative situation. Using quantitative problems to train students only in executing arithmetical operations inevitably leads to a calculational orientation.

1.2.3 Problem solving and sense-making

To understand better the mechanisms involved when students solve mathematical contextual problems – word problems or otherwise – many studies in mathematics education have recently focused on modelling and problem solving (Blum, 2002, 2015; Burkhardt, 2006; Haines, Galbraith, Blum, & Khan, 2007; Schoenfeld, 1992, 2007, 2014; Törner, Schoenfeld, & Reiss, 2007). These studies focus not only on the representation of the problem, but in a much broader sense on the processes and mechanisms that take place when students solve problems of a quantitative nature, including sense-making.

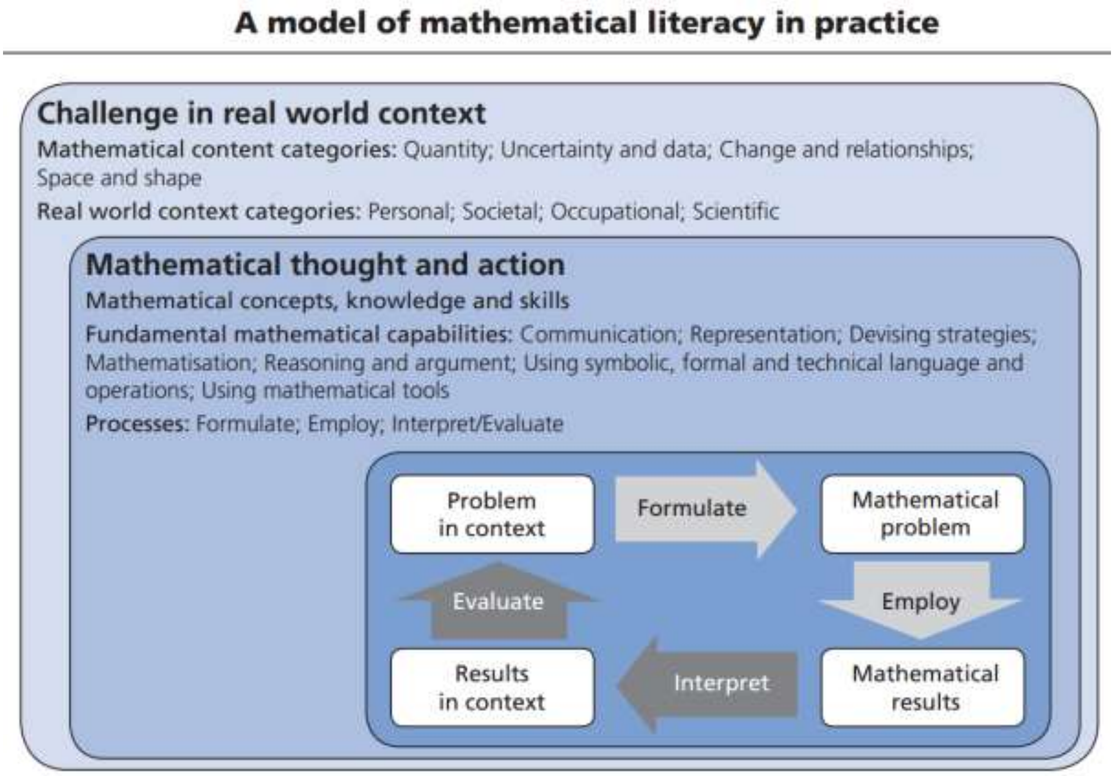


Fig. 1.1 A model of mathematical literacy in practice. From OECD (2013a) (p. 26).

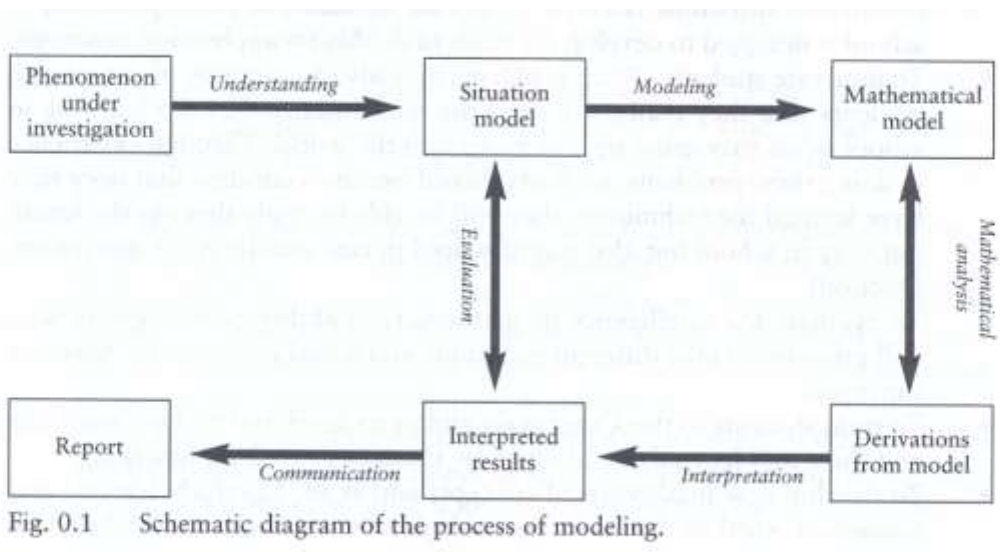


Fig. 1.2 Schematic diagram of the process of modelling. From Verschaffel et al. (2000) (p. xii), reproduced with permission of the author.

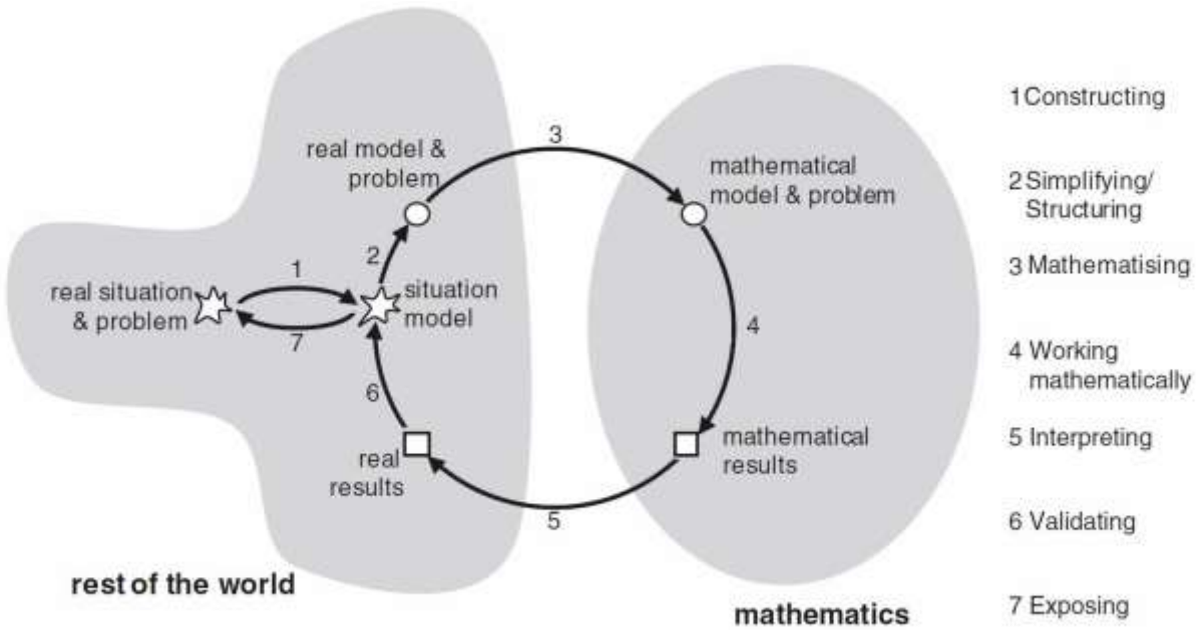


Fig. 1.3 Seven-step modelling schema. From Blum (2015) (p. 76), reproduced with permission of the author.

In the literature on problem solving and modelling, several diagrams (Figures 1.1, 1.2, and 1.3) are used to visualise the process of solving contextual mathematical problems (Blum & Leiß, 2007; OECD, 2013a; Verschaffel et al., 2000).

The three diagrams above show a variety of cognitive activities that are involved in solving quantitative problems from everyday life, but the basic structure in the diagrams is comparable. In the first phase, the starting point is a real problem (problem in real-world context, problem under investigation, real situation and problem) and an assertion that this problem must be modelled (problem in context, situation model). In the second phase, a clear and well-defined phase of mathematising is stipulated. In the successive diagrams these initial phases are called: formulating the mathematical problem, modelling and problem solving, understanding and modelling the problem, mathematising, and planning (Stacey, 2015). These initial phases are particularly important for our investigation, because changing the representation of the problem situation most likely influences these first phases of mathematising and modelling. Moreover we can argue that sense-making is an important part of these initial phases.

In the third phase the actual calculating takes place (employing, mathematical analysis, working mathematically). Sense-making is less important in this phase, and in both representations of the tasks used in our studies the same operations and the same calculation must be conducted to solve the problem.

In the fourth phase, at the end of the cycle, the solver has to interpret the outcome of the mathematical activity and again make sense of it in the perspective of the original problem (interpret and evaluate, interpretation and communication, interpreting, validating, exposing). We speak of a “problem-solving mindset” when students fully engage in all activities displayed in these diagrams and not only focus on the calculational phase. The image-rich numeracy problems we designed aimed at keeping the students in a problem-solving mindset and engaged in the whole cycle of problem solving.

1.2.4 Other perspectives on the use of depictive elements

In the previous section, the potential positive effect of using mainly depictive representations of problem situations on sense-making, engaging in a problem-solving mindset, and performance were presented. However, theories on the effect on students’ performance of adding depictive elements have not shown to be conclusive, and even sometimes contradictory in their predictions. This reinforced the need to answer our research question empirically, but could also indicate some mitigating effects on the prediction that students' performance would improve by changing the representation of the problem situation to mainly depictive. These theories are elaborated in this section.

The Dual Coding Theory (Paivio, 1986), for instance, focused on individual characteristics and categorised people as visual learners (visualisers) or verbal learners (verbalisers). Massa and Mayer (2006), however, found no empirical evidence that in order to gain better results verbal learners should be given verbal instruction and visual learners should be given visual instruction. Instead they found that adding pictorial aids to an online lesson that was heavily text-based tended to help both visualisers and verbalisers. Earlier studies on learning languages (Plass, Chun, Mayer, & Leutner, 1998) have also shown that a combination of text and pictures yielded better results in terms of learning outcomes than text alone. Various studies from the perspective of multimedia learning suggest that under the right conditions, that is, where text and pictures are mutual relevant elements of the problem and allow the students to interpret and integrate the information with minimum cognitive processing, the combination of text and pictures can promote better comprehension (Fletcher & Tobias, 2005; Mayer, 2005, 2009). In mathematics education, Dewolf, Van Dooren, Ev Cimen, and Verschaffel (Dewolf et al., 2014; Dewolf et al., 2015) studied the effect of adding decorative, representational, and informational pictures to mathematical problems, but found no significant effects. One important factor Dewolf et al. (2015) offered as a possible explanation for this is "that the illustrations used did not depict in a sufficient salient way the realistic modelling complexity involved in the items" (p. 167). This provides valuable clues on the importance of choosing visual representations which are closely connected to the "real" problem, and an integral part of the problem situation, to be of effect.

In recent years, research on performance in mathematical problem solving has been carried out from a perspective of Cognitive Load Theory (Sweller, 2005, 2010). These studies argued that redundancy of information, for example, by adding illustrations, can put extra cognitive load on students' working memories, and split-attention effects can occur when students have to jump between text and illustration elements (Berends & van Lieshout, 2009; Rasmussen & Bisanz, 2005; Scheiter, Gerjets, & Catrambone, 2006). From this perspective, an effect may occur opposite to the predicted positive effects of image-rich numeracy problems on students' performance. This is elaborated on this in Chapter 7, where future research questions are discussed.

Cognitive psychology also offers theories and insights on the effect of depictive and descriptive representations on creativity and problem solving (Schnotz, 2002; Schnotz, Baadte, Müller, & Rasch, 2010; Schnotz & Bannert, 2003). Schnotz and Bannert (2003) concluded that task-appropriate graphics may support learning and task-inappropriate graphics may interfere with mental model construction. Schnotz et al. (2010) stated that, to solve a quantitative problem, a task-oriented construction of a mental mathematical representation is necessary, provided that it is task-appropriate. Their line of reasoning is that depictive representations can help students to make a relevant mathematical mental model of the situation, and that depictive representations have a high inferential power because the information can "be read off more directly from the representation" (p. 21). This perspective added to the plausibility of our prediction, which we tested in our empirical studies.

1.3 The Dutch context

In the Dutch situation the issue of how to represent problem situations from real life in mathematics education has been a subject of ongoing political debates (SLO, 2014). Although vast experience with contextual mathematics education has been accrued in primary education since the 1970s, in upper secondary education since the 1980s, and in vocational education since the 1990s, the debate on how to represent problem situations from real life in mathematics education in a fruitful and effective way is still very much alive. The use of contexts, and more specifically the use of word problems in examinations, is a recurring subject in political debates and in (social) media in the Netherlands.

For a better understanding of this debate in this section I elaborate further on the political and societal context of the past 15 years. In the Netherlands at the start of the twenty-first century one could notice a rising concern about an alleged decline in mathematics performance. In the discussions, a kind of "backward utopia" was discernible. It manifested in a proclaimed belief that there once was a time when all children mastered all mathematical multiplication facts and were all able to execute all the algorithms in addition, subtraction, multiplication, and division of whole numbers, decimal numbers and fractions quickly and flawlessly. There was no factual underpinning for this belief, but it provided a frame for the discussions on the state of mathematics education which was strong enough to provoke policy changes.

The political debate resulted in a demand for an education-wide framework regarding literacy and numeracy. In 2010 the "Referentiekader Taal en Rekenen" [Dutch Literacy and Numeracy Framework (LaNF)] was passed by law. This LaNF described the goals for Dutch language (as mother tongue) and mathematics for students aged 4-18 years. Furthermore, the Ministry of Education decreed by law a set of high stakes final examinations in numeracy (next to a mathematics curriculum) at the end of secondary education. Amidst vivid discussions in public media on contrasting desires for the content, the resulting LaNF contained a broad perspective on numeracy, comparable with other international frameworks like PISA (OECD, 2013c), PIAAC (PIAAC Numeracy Expert Group, 2009; Tsatsaroni & Evans, 2014), or for example Norway's framework for basic skills (Vox - Norwegian Agency for Lifelong Learning, 2013). The breadth of the LaNF showed in the four content domains – numbers, proportions, measurement & geometry, and relations (tables, diagrams, graphs, formula) –, and the emphasis on functional mathematics, problem solving, and mathematical literacy. The specific content of the Dutch LaNF is described in more detail in Chapter 2 and in other resources (Hoogland & Stelwagen, 2012; Ministerie van OCW, 2009). Although in 2010 a law was passed imposing an obligatory examination at the end of each educational track in secondary education, its implementation has only progressed to approximately 10% of the intended student population. The debate on further implementation is still ongoing.

1.4 Overview of the dissertation

This dissertation reports on four empirical studies: an exploratory study; the design and validation of an instrument to measure students' performance on numeracy tasks; a large-scale randomised controlled trial; and a replication study. These studies provided results that constituted the building blocks for answering the research question. Based on interview data, the exploratory study indicated what kind of mathematical representations were making sense for students. This knowledge was used to design image-rich, relevant, and task-appropriate numeracy problems as alternatives to word problems. The designed tasks were subsequently used in the design and validation of an instrument that could measure the difference in students' performance when changing the representation of the problem situation of the numeracy tasks from descriptive to mainly predictive. The instrument was used in a large-scale trial with students aged 11-20 years, and was replicated with adults as a target group.

The dissertation is structured as follows. Chapter 1 provides the rationale for the research question, the theoretical frameworks that informed the research, and the context of the studies from national and international perspectives. Chapter 2 reports on an exploratory study on the numerate world of vocational students. Through a series of semi-structured interviews with students, data were collected on the "numeracy incidents" that were to be identified in the transcription of the interviews. These numeracy incidents were categorised according to two major numeracy frameworks: the international PISA framework on numeracy (OECD, 2013c) and the Dutch Literacy and Numeracy Framework (Ministerie van OCW, 2009). The findings showed that there is a gap between the numerate world as expressed by the interviewed students and the representation of the numerate reality to students in the most common numeracy assessments of students.

Chapters 3, 4, and 5 report on the large-scale trial carried out with over 30,000 participants from primary, general secondary, and vocational education. Chapter 3 presents the design and validation of the instrument that was used in the large-scale trial and in the replication study. Chapter 4 provides the main findings of the large-scale trial with the general conclusions on the effect on students' performances. The effect size and the interdependency of the other measured and not-measured variables are discussed on the basis of a probit analysis.

Chapter 5 reports on the analysis on item level of the trial conducted. This analysis sheds light on the relation of the item characteristics and the measured difference in students' performance, giving an indication on which domains of the LaNF the measured effect was largest.

Chapter 6 reports on a replication study of the developed instrument with 420 adult participants in the Groot Nationaal Rekenonderzoek [Great National Numeracy Survey], a research initiative by the public broadcasting organisations VPRO and NTR, supported by the Netherlands Organisation for Scientific Research (NWO). The findings of the replication study were compared with the findings of the large-scale trial.

In Chapter 7 the main findings are summarised, conclusions provided, and the limitations of these studies discussed. This chapter also provides suggestions for future research. During the large-scale trial more data were gathered than were analysed and reported on. There are many possibilities for further analysis of the collected data to test hypotheses and gain in-depth knowledge of the processes that are involved in the assessments of numeracy and the representation of results in a school context. Some tentative research designs are described that can give impetus to further research.

Chapters 2, 3, 4, 5, and 6 of this dissertation have been disseminated to the international scientific community in the form of peer-reviewed journal articles and conference papers. Some of these articles have been accepted, others are still under revision or under review. This way of sharing our findings with the scientific community makes it possible to read the chapters independently. Therefore, there is an unavoidable overlap in some sections of the chapters, especially in the sections on the theoretical background. For the coherence and consistency of this dissertation the references are combined in one appendix and the spelling is made uniform as British English.

Chapter 2. The numeracy of vocational students: Exploring the nature of the mathematics used in daily life and work

Abstract

In 2010 a Dutch Literacy and Numeracy Framework (LaNF) was imposed by law, with a compulsory examination at the end of vocational tracks. To explore the relevance of the content of this framework, we conducted semi-structured interviews with vocational students which focused on daily and work related activities. In the analysis we identified “numeracy incidents” and categorised them in the domain categories of both the PISA 2015 and the Dutch LaNF. This gave an indication that the frameworks could be related to the numeracy practices of vocational students. In the work related activities, however, we found that the students showed better understanding of mathematical concepts with produced artefacts at hand and that they used supporting gestures to express their numeracy competences. This gave rise to serious concern whether the digital test used in the Netherlands was an appropriate way to assess the numeracy of vocational students.

This chapter is based on two peer-reviewed conference papers.

Hoogland, K. (2005). Mind and gesture: The numeracy of a vocational student. In M. Horne & B. Marr (Eds.), *Connecting voices in adult mathematics and numeracy: Practitioners, researchers and learners. Proceedings of the adults learning mathematics 12th annual international conference (ALM)* (pp. 150-158). Melbourne, Australia: ACU.

Hoogland, K., & Pepin, B. (accepted). The numeracy of vocational students: Exploring the nature of the mathematics used in daily life and work *Proceedings of the 13th International Congress on Mathematical Education.* Hamburg, Germany.

2.1 Introduction

In the Netherlands, despite high rankings in international comparative studies, there is a persistent concern with regard to the literacy and numeracy levels of young adults entering advanced vocational education and the workplace. The political reaction to these complaints and concerns was the introduction in 2010 of a Literacy and Numeracy Framework (LaNF) with a compulsory numeracy examination at the end of all vocational educational tracks (Ministerie van OCW, 2009). The developed numeracy framework is described in more detail by Hoogland and Stelwagen (2012). The compulsory numeracy examination consists of 45 mathematical problems, of which 15 are strictly procedural and 30 are contextual problems (Cito, 2015). Many teachers and mathematics educators have questioned the relevance of this way of assessing vocational students, and they see a gap between numeracy used by students in their everyday life and (future) work, and numeracy as operationalised in the final examination.

To gain a better understanding of the numeracy practices of vocational students the research question of this study was: “How can the numerate world of vocational students be seen and heard in interviews with the students about everyday life and (future) work related activities?”. The rationale for the study was that a better understanding and research-informed insights into the way vocational students use numeracy in their daily lives could contribute to a reconsideration of appropriate ways to assess such skills.

2.2 Theoretical background

In the research question the concept of numeracy is used. Numeracy, though, is a concept without well-defined content nor clear boundaries (Geiger, Goos, & Forgasz, 2015; Ruthven, 2016; Tout & Gal, 2015). Most definitions of numeracy have in common that they acknowledge a relation with mathematical knowledge and skills, and that there is a relevancy for the use in everyday life, in and outside the home. This conceptual vagueness and contested definitions are also a product of divergent individual and societal concerns and values (Jablonka, 2003).

In this study we used a definition of numeracy coined in the Netherlands: Gecijferdheid is de combinatie van kennis, vaardigheden en persoonlijke kwaliteiten die een individu nodig heeft om adequaat en autonoom om te gaan met de kwantitatieve kant van de wereld om ons heen (Hoogland, 2006). This translates to: Numeracy is a combination of knowledge, skills and personal qualities which an individual needs to deal adequately and autonomously with the quantitative side of the world around us. The reference in this definition to the combination of knowledge, skills, and personal qualities like self-confidence, perseverance, curiosity, a lack of fear of the quantitative, assertiveness, etc., is consistent with the definition of numeracy used by Gal et al. (2003), which states that numerate behaviour is also enabled by dispositional elements as prior beliefs, attitudes and habits. In the organising framework of Maguire and O'Donoghue (2002) this definition fits in the integrative phase of the development of the conceptualisation of numeracy.

A further knowledge base for this study is the body of knowledge that is built around research on mathematics for work (Bakker, 2014; Coben, 2003; Straesser, 2015; Wake, 2014). From this research a picture is arising of a strong necessity to refine and retune the current mathematics education to a more usable and functional content to better accommodate the future workforce, for instance, by focusing on techno-mathematical literacies (Hoyles, Noss, Kent, & Bakker, 2010), modelling (Wake, 2015), and problem solving (Schoenfeld, 2007; Törner et al., 2007).

2.3 Research design and methods

The methodology of investigating the nature of numeracy by interviewing individuals is what Straesser (2015) would classify as a case study in the field of numeracy. The numerate world of a vocational student is the world as the student constructs and experiences it. In this study we tried to get a glimpse of that world through the lens of the definition of numeracy in this study. The study was exploratory and phenomenological and used a stimulated recall approach, also known from studies by Zevenbergen and Zevenbergen (2000; 2009).

2.3.1 Numeracy incidents

To identify aspects of numeracy in the interviews we defined a "numeracy incident" as the combination of a situation and the numerate (re)action of the individual. This made it possible to identify, categorise, and analyse these incidents and draw conclusions on their nature.

2.3.2 Participants

The participants in the interviews were 19 vocational students aged 15-19 years in the lower streams of pre-vocational and vocational education. The students agreed to be interviewed and video-recorded. They agreed also that data from the interviews could be used anonymously in this study, and only in this research, and that after publication of the anonymised results the video-recordings would be destroyed.

2.3.3 Data collection

Over a period of two years, in two cycles, 19 students were interviewed on issues related to their daily life and (future) work. In the first cycle the interviews were semi-structured and the focus was on examples of numeracy in daily life, more specifically regarding lifestyle issues. The lifestyle issues we identified for these contemporary young adults were: using a mobile phone; earnings in jobs and apprenticeships; holidays and travelling; spending and prices. We narrowed it down to a list of topics that might be discussed with the respondents:

- hobbies
- clothes
- part-time jobs
- holidays/travel
- future jobs
- mobile phone
- savings and allowances

In the second cycle we changed three things. Firstly, the topics of the interviews changed to artefacts the students produced in (future) work related assignments. Secondly, the interview technique shifted to more open questions, aiming at “letting the students talk about the making of the artefacts”. Thirdly, the analysis was broadened to include positive or negative attitude in relation to numeracy and use of gestures that supported the showing of numeracy concepts by the interviewees. The interviews were conducted by six different pairs of interviewer plus observer. They included two mathematics teachers and four mathematics educators. All of them were well informed on the concepts of numeracy and the aim of the study.

2.3.4 Categorisation schemes

In the first cycle of interviews we identified and categorised the numeracy incidents according to the content domains in the PISA 2012 Framework (OECD, 2013c) and the Dutch LaNF respectively (see Table 2.1).

In the Dutch framework there is a large emphasis on proportions, fractions and percentages, and topic areas such as probability and data (representation) are backgrounded and incorporated in the domain of relations.

Table 2.1
Overview of content categories in two frameworks

PISA 2012 Framework	**Dutch Literacy and Numeracy**
Quantity	Numbers
	Proportions
Shape & Space	Measurement & Geometry
Change & Relationship	Relations
Uncertainty & Data	

Note. Presented by similarity (horizontal).

The categories from PISA 2012 and LaNF 2010 are not very refined, and they are aimed more at classifying curriculum goals, lesson materials, and assessments items than they are constructed to classify human expressions or actions. So, in the second cycle of our study we expanded the categorisation scheme to include three categories that were not content oriented, but were oriented to the attitude (positive or negative) shown towards the numeracy incidents reported and the use of supporting gestures in the numeracy incidents.

2.4 Results

In the first cycle of interviews the questions were on a number of lifestyle issues typical for these contemporary young adults: hobbies; clothes; jobs; holiday and travelling; mobile phones; and money. In about 102 minutes of interviews some 135 numeracy incidents were identified, most of them around the topics: holiday and travelling (27), jobs (24), and mobile phones (23). Relating to the PISA framework around 70% of the numeracy incidents were in the domain of quantity, another 15% in the domain of shape & space, and another 15% in the domain of change & relationship. Relating to the LaNF framework, around 50% of the numeracy incidents were in the domain of numbers, another 10% in the domain of proportions, and another 40% in the domain of measurement & geometry. The complete transcription of the numeracy incidents identified will be translated and made available under open access. We show some examples below as a kind of vignettes.

> A diesel car is economical, but an American car is not. I do not know how much petrol an American car uses. We went to xx and when we arrived we already had to fill up again. I do not know how many litres can be filled up. I guess the distance was 100 kilometres. A litre of petrol cost 1 euro.

> I want to go into the hotel business. And then if you know how many guests are coming, it is convenient when you know how to do the arithmetic. I don't know. Depends on having one star or having five stars – 50 euro a night? Maybe 75? In a five-star hotel you might pay 500 euro and then you have only breakfast with it.

> A small (beauty) treatment costs something like 21 euro. A small treatment is about 30 minutes, quarter of an hour or three quarters of an hour. A long treatment costs around 50 or 90 euro.

> I get 50 euro clothing money a month. A shirt doesn't have to be 180 euro, in that case I don't want it. That expensive is not necessary. In the sales I bought some shirts, that costs nothing. Now I am saving again. The sales are stopping. A shirt first was 74 euro and I bought it

for 34 euro. That is something like 50% off. They normally cost over 100 guilders. I would need a lot of money if I had to buy my clothes myself. I don't know. If I want to buy brand clothing, I think around 300 euro a month.

A: I have so many jobs.
Q: Do you have an overview of your earnings?
A: No, I just put it in a box for the holidays. No, I do not know how much I earn. I guess around 150 euro?? 5 euro per hour just 15 + 15 + ….. I don't know. I get 3 euro per hour. A fixed evening Tuesday, 3 hours babysitting and sometimes on the weekend.
Q: If you want to save for something. How much is it about per week?
A: Per day 5 euro, for three hours, no, 7 euro 50. That is on one evening. Maybe 10 euro per week.

In the second cycle of interviews the vocational students were interviewed on an artefact they had produced. The identification and categorisation of the numeracy incidents took place in the same way as in the first cycle. Our first result was that in the interviews on future work situations with an artefact at hand it was possible to pinpoint numeracy incident, under the assumption that a numeracy incident is a numerical situation and the (re)action of the individual to that situation. Our second result was that in this case also the domain categories of PISA 2012 and LaNF 2010 are feasible and useful to categorise the numeracy incidents. Our third result was that, compared with the interviews on lifestyle issues held in the first cycle, the category of space & shape (PISA 2012 Framework) occurred in a higher frequency, likewise with the domain of proportions (LaNF), as was expected in a design and production setting.

But much more striking was that students talked for longer, more openly, with more positive attitude, and with more numerate sophistication about the artefacts at hand to support their answers. It was also notable that the talking about mathematical aspects of the artefact at hand was in many cases accompanied by gestures that supported the mathematical concepts, for instance, by using fingers to point in directions, using thumb and index finger to show angles, using a horizontal flat hand indicating levels to indicate rising numbers. We give a few vignette-like examples.

> It is not that difficult. You just measure here (gesture) and draw a line like ... (gesture). It has to have this length (pointing at diagram). And so on.
>
> First you have to put your bevel gauge to 45 degrees, like this (simulating with his right hand). And then you put it like that (gesture), and then you can saw (gesture).

The expansion of the categorisation scheme with categories on disposition and gestures made the analysis provide a more complete picture of the numeracy practices of the vocational students.

2.5 Discussion and conclusion

In the last few decades there has been increasing interest within the world of mathematics education on the use of mathematics in real situations, especially on the solving of quantitative problems from daily life (Blum, Galbraith, Henn, & Niss, 2007; Bonotto, 2007). Developing students' abilities to solve quantitative problems from daily life is seen increasingly as a goal of mathematics education. Much research indicates that the common classroom practice to teach or assess these abilities is by using word problems, and many reports show that this classroom practice is not effective due to the use of stereotypical questions, due to (language) issues regarding word problems (Verschaffel et al., 2000; Verschaffel et al., 2009), and due to a classroom culture that is more focused on answer-getting than on problem solving (Daro, 2013). In our research project the focus is on a more authentic representation of real-life situations in the classroom. We have a special interest in vocational students, because they are most likely to be the students who will use mathematics as a practical tool in their daily life and occupation.

We concluded that both the PISA framework and the Dutch LaNF were adequate to categorise the content of the numeracy practices we identified in the interviews with the students. However, the sophistication and efficacy of the demonstrated numeracy practices were much higher when the students were interviewed with the artefact at hand. This observation was even invigorated by paying attention to the gestures the students made to support their mathematical concepts. Also notable in the second cycle was that the students hardly used any explicit mathematical concepts. Instead gestures and words from their own (practical) vocabulary were used.

To conclude, to assess numeracy practices we recommend making practical settings and artefacts at hand part of the assessments. Such recommendations are in line other mathematics educators who advocate the use of real-life situations in mathematics classrooms to teach and assess students' ability to deal with quantitative problems from everyday life (Bonotto, 2007, 2009; Frankenstein, 2009; Zevenbergen & Zevenbergen, 2009).

The findings of this study added to the concern on the relevancy of assessing the numeracy competences of vocational students solely with a digital answer-oriented series of procedural and word problems. The current Dutch legislation for now makes it impossible to use formative or practice oriented assessment methods for vocational students that might do more justice to the nature of the numeracy competences of these practical oriented young adults.

Chapter 3. Representing contextual mathematical problems in descriptive or depictive form: Design and validation of an instrument

Abstract

The aim of this study is to contribute to the body of knowledge on the use of contextual mathematical problems. Word problems are a predominant genre in mathematics classrooms in assessing students' abilities to solve problems from everyday life. Research on word problems, however, reveals a range of difficulties in their use in mathematics education. In our research we took an alternative approach: we designed image-rich numeracy problems as alternatives to word problems. A set of word problems was modified by systematically replacing the descriptive representation of the problem situation by a mainly depictive representation and an instrument was designed to measure the effect of this modification on students' performance. The instrument can measure the effect of this alternative approach in a randomised controlled trial. In order to use the instrument at scale, we made this instrument usable also as a diagnostic test for an upcoming nationwide examination on numeracy. In this article we explain and discuss the design of the instrument and the validation of its intended uses.

This chapter is published as:

Hoogland, K., Pepin, B., Bakker, A., de Koning, J., & Gravemeijer, K. (2016). Representing contextual mathematical problems in descriptive or depictive form: Design of an instrument and validation of its uses. *Studies in Educational Evaluation, 50,* 22-32. doi:10.1016/j.stueduc.2016.06.005

3.1 Introduction

In mathematics education there is an increasing focus on the usability of acquired mathematical knowledge and skills (Kilpatrick, 1996; OECD, 1999; Toner, 2011), and hence there is a growing need for materials and tools to teach and assess the use of mathematical knowledge and skills in real-life situations. For decades it has been common practice to use word problems to teach and assess students' abilities to solve quantitative problems in practical day-to-day situations (Verschaffel et al., 2000). However, the current practice of using word problems to assess students' abilities to solve quantitative problems from everyday life also gives rise to serious concerns. The question is whether word problems are adequate for this purpose (Verschaffel et al., 2000; Verschaffel et al., 2009). According to Verschaffel et al. (2014) word problems can be defined as: "verbal descriptions of problem situations wherein one or more questions are raised the answer to which can be obtained by the application of mathematical operations to numerical data available in the problem statement" (p. 641). In this study we used "word problems" for problems in which both the description of the problem situation as well as the actual problem statement are presented in words.

The reported difficulties with word problems are so persistent that in this study we investigated an alternative to word problems as a means to evaluate students' abilities to solve quantitative problems in practical day-to-day situations. In this alternative the descriptive representation of the problem situation, as is common in word problems, is replaced as much as possible by a depictive representation, which means using visual elements, mostly photographs, that are as close as possible to the real-life problem situation. To contrast them with word problems we call these problems image-rich numeracy problems. The choice for a mainly depictive representation of the problem situation is informed by research on difficulties with word problems (Verschaffel et al., 2009), considerations on the sometimes problematic relation between language, context, and sense-making in solving word problems (Sepeng, 2013), considerations on authenticity in mathematical problem solving (Palm, 2009; Verschaffel et al., 2000), and research on problem solving in cognitive psychology (Schnotz, 2002, 2005; Schnotz et al., 2010). These research perspectives combined strongly suggest that using real-life images, such as photographs, to represent the problem situation is more effective in keeping students in a problem-solving mindset instead of falling back to an answer-getting mindset (Daro, 2013). Photographs are more easily associated with real-life situations, and arguably can feel more authentic for students, and therefore increase the likelihood that students will continue using considerations of reality in the problem-solving activities. Furthermore, it is likely that difficulties of language and text comprehension are reduced by mainly depictive representations of the problem situation. The effects of changing the representation of the problem situation on students' performance are still under-researched and we decided a validated instrument was needed to measure these effects. In this article we described the design of such an instrument and the validation of its intended uses. The instrument was subsequently used in the Dutch context (see Chapters 4, 5, and 6), and the English version of the instrument is now available under open access (Hoogland & De Koning, 2013).

3.2 Theoretical perspectives

3.2.1 Mathematics as usable knowledge

Over the past 50 years situations from real life have increasingly been used in school mathematics classrooms. There are several developments in mathematics education that have pushed this trend forward. Firstly, there is a plea by mathematics educators for a stronger emphasis in school mathematics on the ways in which mathematics is used in daily life (Kilpatrick, 1996; OECD, 1999). Kilpatrick (1996) observes that "the curriculum had shifted (…) away from an emphasis on abstract structures towards efforts to include more realistic applications, with an emphasis on the ways in which mathematics is used in daily and professional life" (p. 7). Secondly, there is an increasing use of examples from reality as an integral part of an instruction theory for mathematics. For instance, in Realistic Mathematics Education (RME) (Freudenthal, 1973; Gravemeijer, 1994, 1999, 2004; Van den Heuvel-Panhuizen, 2000; Van den Heuvel-Panhuizen & Drijvers, 2014) contexts, models and representations play an important role in the educational process. The central idea in RME is that students should be supported in reinventing mathematics with the support of the teacher and the curriculum materials (e.g., textbooks). The starting points of such reinvention processes should be experientially real for the students. Problems situated in everyday life contexts often fulfil this requirement. Hence, in RME, situations from real life are not just used to prepare students for solving applied problems. The main function of real-life situations in RME is to offer a conceptual basis for reinventing the mathematics the students are to learn. In relation to this, Freudenthal's (1983) didactical phenomenology suggests looking for phenomena that – as he puts it – "beg to be organised" by the mathematical procedures, concepts or tools one wants the students to (re)invent. In addition to this, Treffers (1987) recommends a broad phenomenological exploration in order to incorporate various inroads into the mathematical procedures, concepts or tools under consideration. Thirdly, in mathematics education research there is an increasing focus on problem-solving and modelling (Blum et al., 2007; Burkhardt, 2006; Kaiser, Blomhøj, & Sriraman, 2006; Lesh & Zawojewski, 2007; Schoenfeld, 1992; Sriraman, Kaiser, & Blomhøj, 2006). Schoenfeld (2014) signals a reframing of what it means to understand mathematics:

> At the core of that reframing is the notion of mathematics as a sense making activity—that learning mathematics entails developing deep understandings of certain culturally and historically transmitted ideas, and employing those ideas in ways that reflect the perception of objects and relations, their mathematization, and the meaningful use of mathematical symbols in the service of solving problems. (p. 498)

The aforementioned developments concerning applicability, instruction, problem solving, and modelling can be seen as branches of a larger tree that represents the delicate relationship between reality and mathematics (education) and subsequently the way that reality is represented in classroom practice. As a consequence of these developments, in classroom practice we see a great variety of examples and problems from real life, which are used as tasks. The predominant form used for these tasks is the word problem, which uses above all a descriptive representation of reality (Madison & Steen, 2008; Sriraman et al., 2006). This use of word problems, however, is not without difficulties.

In the next section we will show the difficulties that are encountered when using word problems in classroom situations; which ideas for counteracting those difficulties are conveyed in the literature; and a potential alternative to be designed.

3.2.2 Difficulties with sense-making in solving word problems

In the literature on problem solving and modelling several schemas are used to visualise the process of solving word problems (Blum et al., 2007; Burkhardt, 2006; OECD, 2013a; Verschaffel et al., 2000). In Figure 3.1 the OECD (2013a) schema is shown. These schemas have in common that they show a variety of cognitive activities that are possible and necessary to solve quantitative problems from everyday life. It is assumed that somewhere in the process the solver must formulate a (mental) mathematical problem from the problem situation, which gives him the opportunity to employ the relevant mathematical reasoning and calculations towards a mathematical solution to the problem. And typically with this kind of problems, the solver has to interpret the outcome of the employment of mathematical actions to make sense of it in the perspective of the original problem. We can speak of a problem-solving mindset if a student fully engages in the activities formulated in these schemas. However, each step also stands for possible barriers in the problem-solving process.

Many studies on students' behaviour in solving word problems report that the steps of understanding the situation are often superficially executed by the students (Verschaffel et al., 2000). Other difficulties with word problems arise when students disregard possible constraints imposed by reality when they experience unfamiliarity with the situation at hand (Cooper & Harries, 2003; Lave, 1992; Reusser & Stebler, 1997; Verschaffel et al., 2009), or lack the proper meta-cognitive skills for solving the problem (Caldwell, 1995). The most reported difficulty is that students base their analysis and calculations on a loose association of certain salient quantitative elements of the problem situation with certain mathematical operations (Verschaffel et al., 2000).

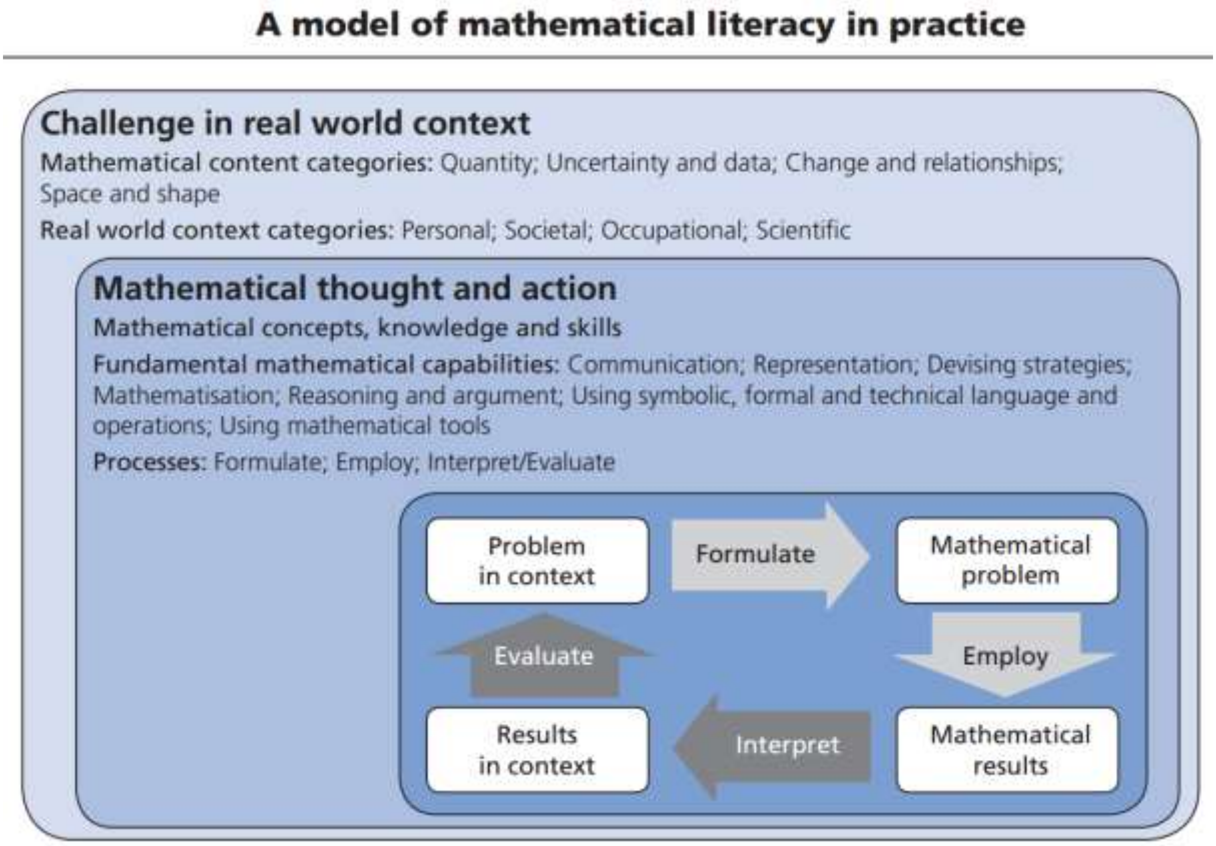


Fig. 3.1 A model of mathematical literacy in practice. From OECD (2013a) (p. 26).

Students seem to focus merely on the right-hand side vertical step of the diagram in Figure 3.1 and on the outcome of the calculation, and they seem to value these outcomes more highly than the "realism" of the outcome. This student behaviour is an indication that students see a word problem as a "school maths" problem disguising some arithmetic algorithms, and not as a representation of a real-life problem that has to be understood. This behaviour stems from a strong calculational orientation in mathematics classrooms (Thompson et al., 1994), whereby the focus is on procedures and operations and not on explaining and reasoning. In the calculational orientation, quantitative problems seem first and foremost to be used to train students in procedural fluency. From this perspective, the representation of the problem situation is not a particularly important aspect, and the word problem chosen is typically straightforward, scarcely hiding the aim of merely being a means of training for a certain type of calculation. The phenomenon that students often fail to note the meaninglessness of the stated problems and believe that every word problem has to be solved by a single numerical answer was termed by Schoenfeld (1991) "suspension of sense-making". This phenomenon is further investigated in the monograph *Making sense of word problems* (Verschaffel et al., 2000).

The persistence of the perceived difficulties cannot be ascribed exclusively to student attitude. The literature contends that this student behaviour is likely to be reinforced by teachers' approaches that tend to emphasise the mathematical structure of the problem rather than the contextual aspects (Depaepe et al., 2010). Gravemeijer (1997) and Yackel and Cobb (1995) have pointed out that the typical student and teacher behaviour of focusing on outcomes is widespread and persistent, and based on established implicit or explicit sociomathematical norms in classroom culture. Furthermore, they argue that one norm stands out – in school mathematics solving a word problem is about finding and performing the right calculation.

The mentioned difficulties give rise to serious concerns whether students can show their full potential of solving quantitative real-life problems when confronted with word problems in an assessment or evaluation situation. Hence, it is proposed that an alternative to word problems should be considered and that such alternative should be tested in practice and its effects should be measured in a systematic way.

3.2.3 Counteracting difficulties with word problems

Over the past decades many attempts have been made to formulate quantitative problems, instruct students, and create situations in such a way that the calculational approach and suspension of sense-making around word problems would be counteracted or avoided. Verschaffel et al. (1994), whose work has been reproduced in many countries, researched the effect of using paired standard and "problematic" items, whereby for the problematic items the students were almost forced to take into account the realistic constraints of the situation. The conclusion was that students have such a strong tendency to exclude realistic considerations that "problematising" the word problem was not enough to trigger realistic considerations to any great extent. Reusser and Stebler (1997) researched the effect of alerting students explicitly to the need to weigh realistic considerations, and of suggesting meta-cognitive activities, such as making a sketch or studying the picture. They found a small positive effect for these measures, but their main conclusion was that the classroom culture of not taking into account realistic considerations and hastily looking for the right operation and the right answer persists. Wyndhamn and Säljö (1997) studied the effect of increasing the authenticity of the experimental setting by actually bringing in concrete materials that played a role in the given problems. They found that students scored significantly better in the more authentic settings, which they saw as evidence that the presentation of the problem was the main cause for the students' unrealistic answers. Similar research was done by DeFranco and Curcio (1997) who compared the results of a written problem – "328 senior citizens are going on a trip. A bus can seat 40 people. How many buses are needed so that all the senior citizens can go on a trip?" – with a real-life simulation of this problem where the students actually had to contact the bus operator to order buses. They found a reduction in the reluctance of students to take realistic considerations into account. Cooper and Harries (2002, 2003) singled out a problem from an English national test about the number of times an elevator had to go up to transport a certain number of people. They made an alternative problem where students were encouraged to consider seriously the practical constraints and the assumptions in the calculations. They found that around a quarter of the students proved able and willing to offer reasons for their choices made. Dewolf, Van Dooren, and Verschaffel (2011) examined the effect of presenting the same fair-sharing problem in different settings – mathematics class versus religion class – and found that the context of problem solving has an important influence on the interpretation and solution to the problem. A substantial number of students in the religion class combined and weighted a larger variety of criteria for the problem statement, while students in the mathematics class limited themselves mainly to one criterion for fair sharing.

In research on alternatives to word problems, other researchers and practitioners have asked whether the problem of suspension of sense-making can be counteracted or avoided by changing the problems, or adding incentives to students to make realistic considerations (Bonotto, 2007, 2009; Frankenstein, 2009; Lave, 1992; Zevenbergen & Zevenbergen, 2009). They advocate the creation of actual real-life situations in mathematics lessons, which could feel more authentic to students, to keep students more in a problem-solving mindset. Bonotto (2007) argues for encouraging students to analyse mathematical facts, embedded in appropriate "cultural artefacts", such as supermarket receipts, bottle and can labels, railway schedules, or a weekly TV guide. Although many of the arguments for using real-life situations to teach students relevant problem-solving skills are convincing, there is not a widespread dissemination of such practices. In our alternative we decided to change the representation of the problem situation from descriptive to mainly depictive to get it closer to the real-life situation.

3.2.4 Perspectives on the use of depictive elements

Recent studies by Dewolf and colleagues (Dewolf et al., 2014; Dewolf et al., 2015) and earlier studies of Elia and colleagues (Elia, Gagatsis, & Demetriou, 2007; Elia & Philippou, 2004) addressed the effect of adding decorative, representational, and informational pictures to mathematical word problems, and they found small or no significant effects. These studies encouraged us to look into the designing process for images closely connected to the real-life situation that was to be represented. So, in our approach, the problem situation was represented with images from real life, in the form of photographs, headlines from newspapers, and handwritten notes (see Figure 3.2). We hypothesised that the use of images from real life would increase the association with real-life situations and support the student to solve “a truly engaging dilemma” (Verschaffel et al., 2000), p. 44) and therefore decrease the suspension of sense-making and the strong calculational focus.

Recent studies have also investigated the relation between the use of pictorial elements by students during the problem-solving process and their performance on word problems (Boonen, 2015; Hegarty, 2004; Krawec, 2014; Van Garderen, 2006; Van Garderen & Montague, 2003). These studies show that students perform better when they make use of relevant visual or schematic elements in their problem-solving process. However, this only held when the pictorial elements of the students were relevant for and consistent with the mathematical model needed to solve the problem. When students used less relevant pictorial drawings the effect on their performance was negative.

Furthermore, a supporting argument for the change in the direction of mainly depictive representations can be found in research on the linguistic issues students encounter in solving word problems. For example, research studies investigating the role that language plays in the actual solving of word problems (Sepeng & Webb, 2012), for instance through discussions, show that the linguistic development of students has an effect on their problem-solving skills and on the interplay between contexts and sense-making (Sepeng, 2013). This relation between linguistic development and mathematical problem-solving performances is corroborated by the 2012 PISA studies which show that there is a latent correlation of 0.85 between reading and mathematics performances (OECD, 2014).

These studies provided us with arguments that designing a specific, relevant, and close to real-life form of visual representation of the problem situation might help students to make sense of the problem, adopt a relevant mathematical model, and interpret the outcomes of the mathematical operations. This led to a set of choices we made in designing the tasks for the instrument.

3.2.5 Our choices for an alternative approach

In our research we sought an alternative to word problems that could be used in assessment and evaluation instruments, and at the same time portray real-life situations, that feel authentic to students. Authenticity in education is a widely debated subject dating back to the writings of Dewey (1916). Palm (2009) argues in his Theory of Authentic Situations: “that a strong argument can be made that the fidelity of the simulations (…) clearly has an impact on the extent to which students, when dealing with school tasks, may engage in the mathematical activities attributed to the real situations that are simulated” (p. 9). From Palm’s framework (2009) the aspects of realism and presentation of the problem were used in our alternatives to make the represented problem situation feel more authentic to students. The other aspects, like event and circumstances, were not changed in our particular comparison. We contend that those first two aspects contributed to a higher representativeness of the tasks which arguably could have a positive effect on the quality of problem-solving by the students.

From a cognitive psychological perspective Schnotz et al. (2010) distinguished between descriptive and depictive representations, to each of which they ascribed specific representational and inferential powers. In a much similar approach as was used in the problem-solving schematic of Figure 3.1, Schnotz (2002) stated that to solve a quantitative problem, a task-oriented construction of a mental mathematical model has been necessary. His line of reasoning was that depictive representations could better support students to make a relevant mental mathematical model of the situation. He further stipulated that depictive representations had a high inferential power and were in many cases close to the form of mental mathematical models: sketches; diagrams; and/or drawings in that they captured the essence of the problem (Schnotz et al., 2010). Based on this reasoning we inferred that using images from real-life situations could support students in their problem-solving task.

3.3 Design of the instrument and validity issues

Informed by the aforementioned research on word problems, cognitive psychological perspectives on problem solving, and linguistic considerations, we came to the design of an instrument to measure the effect of changing the representation of the problem situation on students' performance. Aiming for large-scale use, we designed the instrument in such a way that it could simultaneously serve as a diagnostic test for a nationwide examination on numeracy which was to be imposed in the Netherlands. In the design of the instrument we carefully considered issues of validity, according to the AERA/APA/NCME recommendations (Joint Committee on the Standards for Educational and Psychological Testing of the American Educational Research Association the American Psychological Association and the National Council on Measurement in Education, 2014).

For measuring the effect of changing the representation of the problem situation on students' performance, we designed the instrument as a randomised controlled trial, making it possible to analyse the data with sophisticated limited dependent variable models, in order to establish the plausibility of the effect measured and to interpret the effect of possible interdependent variables. For the diagnostic use of the instrument we focused on the validity of the content of the test. In the method section we detailed the evidence we gathered to support the validity of the uses and interpretations of the intended instrument.

3.3.1 Context of the use of the instrument

For international readers we provide information on the context in which the instrument was trialled. In the Netherlands the relevance of the developed instrument is high, as in 2010 a “Referentiekader Taal en Rekenen” [Literacy and Numeracy Framework (LaNF)] was passed as a law (Ministerie van OCW, 2009).

In this framework six levels are formulated (see Table 3.1), and four (out of the six levels) are assessed in nationwide examinations at the end of primary and secondary education. The F-levels focus mainly on so-called functional mathematics and are assessed with tests consisting of 45 to 55 mostly word problems (Cito, 2015). In the framework, four content domains are formulated: numbers; proportions; measurement & geometry; and relations (tables, diagrams, graphs, formulas, etc.). A diagnostic tool for these upcoming nationwide examinations would be supportive for teachers and students, and insights into the effect of changing the representation of the problem situation is likely to contribute to the body of knowledge on how to assess the Dutch LaNF in a research-informed way (Hoogland & Stelwagen, 2012).

Table 3.1
Overview of numeracy levels from Dutch literacy and numeracy framework and moments of examination.

Level	Description	Moment of Nationwide Examination
1F	basic level	end of primary education
1S	advanced level	end of primary education
2F	basic level	end of secondary general and vocational education, lower tracks
2S	advanced level	not tested
3F	basic level	end of secondary general and vocational education, higher tracks
3S	advanced level	not tested

Note. F stands for Fundamenteel Niveau [Basic Level], S stands for Streefniveau [Advanced Level]

3.3.2 General layout of the final instrument

The final instrument is a web-based numeracy test of 24 items, very similar in content and layout to the nationwide examination (Cito, 2015). The 24 items are diagnostic for the level 2F of the Dutch LaNF. Of these items, 21 items are randomly presented in one of two versions: word problem or image-rich numeracy problem, which means that the problems are equivalent regarding the content and the level of mathematical knowledge and skills needed to solve the problem. The versions differ only in the way the problem situation is presented to the participant (see Figure 3.2). This last feature of the instrument makes it possible to measure the effect of changing the representation of the problem situation on students' performance in a randomised controlled way. In the method section we described the activities undertaken to counter threats to the validity of both proposed uses.

3.4 Method

The activities undertaken in designing the instrument and the validation of its different uses are presented in Table 3.2. For the use as a measuring tool on the difference in students' performance related to the representations of the problem situation, we foremost focused on the equivalence of the word problems and the image-rich numeracy problem. For diagnostic use we focused on the construct and content validity of the items. In order to counter threats to reliability an important aspect was computer scoring of the students' solutions, to guarantee that for each participant the scoring was consistent.

Next to activities to provide evidence for construct and content validity we argued for criterion validity with measures obtained in the test run. However, we used these measures with caution, because we are aware that the use of criterion validity is also a topic of discussion among psychologists (Borsboom, Mellenbergh, & van Heerden, 2004). In the next paragraphs we explain the activities undertaken in more detail.

Table 3.2
Overview of activities undertaken in design of the instrument to counter threats to validity and reliability

Phase of development	Number of research activity	Description of research activity	To counter threats to:
Design	1	Selecting 40 existing items around level 2F of the LaNF.	content validity
	2	Designing 40 alternatives and gathering comments on quality of 40 paired problems by 13 experts.	content validity
	3	Estimation of levels 2F of 40 revised paired problems by eight experts.	construct validity
Validating the diagnostic use	4	Creating a web-based version similar to the nationwide examination in content and layout.	construct validity
	5	Relating the items to the LaNF and spreading the items evenly over the domains of the LaNF.	content validity
	6	Performing a test run with over 7000 participants.	feasibility
	7	Checking for internal consistency of the items with measures of the classical item response theory.	criterion validity
	8	Checking correlation of scores on both versions.	content validity
Validating the measurement of changing the representation of the problem situation	9	Checking 40 revised paired problems on equivalence of paired items by eight experts.	content validity
	10	Programming random representations of the problem situation in 21 items and presenting them in random order in the instrument.	construct validity and reliability
	11	Computer scoring students' solutions.	reliability
	12	After test run: checking for correlation between scores on both versions.	content validity
Composing the final instrument	13	Combine results from all above to construct the final instrument.	

3.4.1 The process of designing

The process of designing started with the selection of 40 relevant word problems that were used in recent years in Dutch textbooks and tests, which were developed to teach or assess the numeracy 2F level in the LaNF. Additional selection criteria were that the problems were dealing with a real, perceived as real, or at least imaginable problem from daily life. The selected problems were specifically aiming at level 2F of this framework, and evenly spread over the domains of numbers, proportions, and measurement & geometry. Items on the domain of relations were later added to the instrument to allow for the possible use as a diagnostic test. We limited ourselves to one version for those items because in the domain of relations hardly any word problems without a visual element could be found. The selected problems for the other domains had as common characterisation: the problem situation was described in words, without the use of illustrations or photos. The selected problems were all standardised in the same format: problem situation; problem question; and answer box (see Figure 3.2).

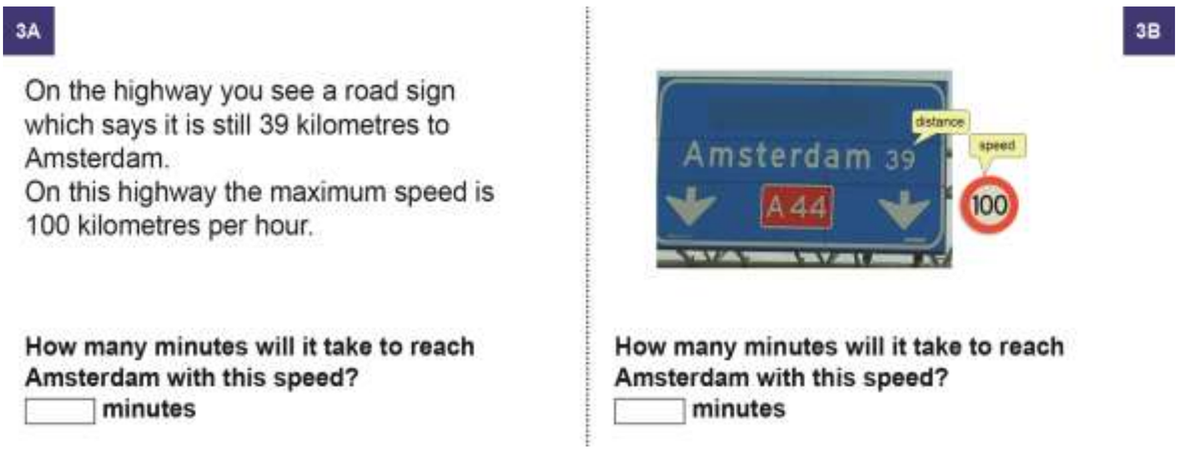


Fig 3.2 Example of word problem (left) and image rich numeracy problem (right).

For all 40 selected word problems an alternative version was designed, in which the descriptive representation of the problem situation was replaced, as much as possible, by a depictive representation. For the depiction we chose one or more images from reality, mostly photographs, with little redundancy, whereby as little language as possible was used. These image-rich numeracy problems were standardised in the same format (see Figure 3.2) as the word problems.

The problem question and the answer box were identical; only the representation of the problem situation was changed. The word problems with a descriptive representation of the problem situation we called the A-version; and the image-rich numeracy problems with a depictive representation of the problem situation we called the B-version, resulting in 40 paired problems. In Table 3.2 these research activities are numbered 1, 4 and 5.

3.4.2 Validation activities by experts

Experts in mathematics education played an important role in the process of content validation. The aim was to design paired problems that differed only in the representation of the problem situation, leaving all the other possible variables the same. For valid conclusions on the effects of changing the representation of the problem situation, the pairs of problems had to be, apart from representation of the problem situation, otherwise equivalent. In a first cycle to improve the items, the 40 paired problems were openly discussed by 13 experts on quality, relevance for the level 2F, and equivalence between A-version and B-version. This led to an improved set of 40 paired problems.

In a second cycle the improved set was presented to eight other Dutch experts in the field of mathematical literacy and numeracy, with the following questions:

(i) Do the two versions of the problem test the same mathematical knowledge and skills?
(ii) If the two versions test the same mathematical knowledge and skills, are they testing on the same mathematical level?
(iii) Give for each problem the estimated level on a five-point scale: too easy for level 2F, easier than level 2F, level 2F, more difficult than level 2F, too difficult for level 2F.

The experts were explicitly asked to disregard their estimation of the effect of changing the representation of the problem situation on the difficulty of the problem. This expert group consisted of teachers, teacher educators, researchers, and test constructors in the domain of mathematics and numeracy education. These experts sent in their results anonymously. We as designers decided after discussion that a pair of problems was acceptable for the instrument, if the first question was answered with “Yes” by a minimum of all but one of the experts. The answers to the second and the third question made it possible that in the final instrument the selected problems were spread evenly around level 2F of the LaNF. In Table 3.2 these research activities are numbered 2, 3, 5 and 9.

3.4.3 The test run of the instrument

The test run of the instrument took place in parallel (with respect to time) with the activities by the experts. From the list of 40 paired problems 21 paired problems were selected considering a good distribution over the domains of numbers, proportions, and measurement & geometry. The aforementioned three items of the domain of relations were added. In this way the test would qualify as a complete 2F test according to the LaNF. The 24 items (21 items in two versions, and 3 items in one version) were programmed as a web-based digital test. From experience with earlier pilot tests of the LaNF we estimated that the 24 problems could be solved in about 50 minutes. The selection of the problems for the test run took place in time before the validation by experts, so the selection was done by the designer with the following criteria: evenly spread over the domains of numbers, proportions, and measurement & geometry, and evenly spread over a bandwidth of levels around 2F.

3.4.4 The participants for the test run

The participants for the test run were selected by inviting schools to participate. We looked for participants from a broad range of tracks and levels to gather information on the feasibility of the items and the test as a whole. The distribution of the participants in the test run of the instrument around the levels of education is shown in Table 3.3.

In the test run of the instrument 7,434 students from 63 different schools in an age range of 11–18 years participated. The schools in the test run of the instrument were geographically spread around the Netherlands and delivered students from all levels of education, which gave a good representation of the proposed use of the instrument in the future.

Table 3.3
Number of participants in test run (N = 7434)

	Primary education (grade 5-6)	Secondary education (grade 1–6)			Total
		VMBO (pre-vocational)	HAVO/VWO (general)	MBO (vocational)	
n	172	3,796	2,838	348	7,434

Note. VMBO is the pre-vocational track in secondary education, HAVO/VWO is the general track in secondary education, MBO is the vocational track in secondary education. Participants are representative to proposed future use of the instrument.

3.4.5 Conducting the test

For each participant a test was generated by randomly choosing 12 items to be presented with a descriptive representation of the problem situation (A-version) and the other 12 to be presented with a depictive representation of the problem (the B-version). The order of the problems presented to the participants was again randomised. To explain, for instance a test could look like the following: 4B, 5B, 9B, 1A, 3B, 6A, 12B, 23A, 24B, 19A, 15A, 17B, and so on. Each problem in the web-based test was presented as a screen-filling problem. The question was posed at the bottom of the screen. Below the question the numerical solution to the problem could be entered. For solving the problems an online calculator was allowed. For the total test a time limit of 60 minutes was set, to make it possible for the students to answer a few short additional questions. All solutions to the problems were numerical values. Participants typed the numerical solutions into an empty answer box. The answers were evaluated and scored by the computer.

In Table 3.2 these research activities are numbered 6, 10 and 11. The outcomes of the remaining research activities in Table 3.2 are discussed in the results section below, to show how we countered threats to the validity of the two proposed uses of the instrument.

3.5 Results

3.5.1 Results of the design process

The resulting items of the first cycle of the design process can be found under open access (Hoogland & De Koning, 2013) and in Appendix A of this dissertation. The results on the three questions posed in the second cycle can be found in Table 3.4.

Table 3.4
Problems, expert judgments, average expectations, and selection for the test run and the final instrument

Problem	**Same knowledge**		**Same level**		**Estimated level**			
	N	Yes	N	Yes	M(A)	M(B)	test run	final instrument
AEX index	8	100%	8	100%	3.88	3.88	P17	v
Endive	8	100%	8	100%	3	3		v
Cough syrup	8	100%	8	100%	3.25	3.25		v
Coffee cups	8	100%	8	100%	3	3	P2	
Scale model	8	100%	8	100%	3.13	3.13	P20	v
Winter tyres	8	100%	8	100%	3.75	3.75		v
Chicory	8	100%	8	100%	2.25	2.25		
Bathroom	8	100%	8	88%	3.63	3.63		v
Baking tin	8	100%	8	88%	3	3		v
Budget cuts	8	100%	8	88%	3.63	3.63	P18	v
Public debt	8	100%	8	88%	3.88	4		v
Gas usage	8	100%	8	75%	2.75	2.75	P5	v
Driving time	8	100%	8	75%	3.25	3.38	P7	v
Apples	6	100%	7	71%	2.86	2.71	P1	v
Petrol	8	100%	8	63%	3.75	4		
Offer on buying	8	100%	8	50%	3.13	2.88	P10	v
Groceries	8	88%	8	88%	2.5	2.5	P13	v
Lawn fertilising	8	88%	8	88%	3.13	3.13		
Hamburger	8	88%	8	88%	3.13	3.13		v
Icelandic Crowns	8	88%	8	88%	3.63	3.63	P16	v
Bedroom tiles	8	88%	8	88%	3.38	3.38		v
Water bottles	8	88%	8	88%	2.88	2.88		v

Double glazing	8	88%	7	86%	3	3.29		v
Bank balance	8	88%	8	75%	2.13	2.25		
Weed control	8	88%	8	75%	3	3		
World cities	8	88%	8	75%	3.71	3.57		
Swimming pool	8	88%	8	75%	3.13	3.13		
Bonbons	8	88%	8	63%	3.13	3.25		v
Music songs	8	88%	8	63%	3.63	3.63	P14	
Water usage	8	88%	8	50%	3.75	3.38	P4	
Recipe	7	86%	7	86%	3.13	3.13	P15	v
Carpet	8	75%	7	71%	2.71	2.75	P6	
Purchase laptop	8	75%	8	63%	3.13	3	P8	
Cubes	8	75%	8	50%	4	4	P21	
Art work	8	75%	8	50%	3.14	3.13	P12	
Elephant	8	75%	8	38%	4.25	4.25	P3	
Book shelves	8	63%	8	63%	3.25	3.13	P9	
Packing box	8	63%	8	50%	3.71	3.43		
Thermometer	8	63%	7	43%	2.38	2.38	P19	
Photo	8	63%	8	25%	4	4	P11	

Note. *N* is number of experts, Yes is percentage of affirmative answers of the experts to question1 and question 2 respectively. M(A) is mean expert estimation of the word problem levels on a 5-point scale referenced to 2F level. M(B) is the mean expert estimation the image-rich numeracy problem levels on a 5-point scale referenced to 2F level. The column *test run* shows the problems used in the test run. The column *final instrument* shows the problems selected for the final instrument.

From the criterion that a pair of problems was acceptable for the instrument if the first question ("Do the two versions of the problem test the same mathematical knowledge and skills?") was answered with "Yes" by a minimum of all but one of the experts, 30 of the 40 paired items were confirmed as eligible for the final instrument. See in Table 3.4 the items above the bold line. To spread the items evenly over the domains another nine items were discarded.

For the 21 problems that were selected for the final instrument we found as mean estimation (with standard deviation in parentheses) of the level 3.23 (0.37) for the word problems, and 3.25 (0.39) for the image-rich numeracy problems. For an average estimation on a 5-point scale this is a good indication for equivalence. In Table 3.2 these research activities are numbered 5 and 8.

3.5.2 Results from the test run of the instrument

Table 3.5
Scores on both versions, item-rest-correlations (RiR), word count (A-version) and number of visual elements (B-version) from test run (*N* = 7434)

item	item score (A-version)	item score (B-version)	RiR(A+B version)	word count (A-version)	number of visual elements (B-version)
P1	.69	.75	.42	17	2
P2	.55	.56	.42	15	2
P3	.28	.14	.26	33	3
P4	.09	.13	.28	37	3
P5	.52	.49	.34	21	3
P6	.77	.82	.36	42	2
P7	.29	.26	.50	29	2
P8	.83	.84	.32	17	2
P9	.60	.49	.42	32	2
P10	.78	.83	.43	28	2
P11	.13	.22	.41	43	2
P12	.42	.48	.44	22	2
P13	.90	.89	.25	11	2
P14	.26	.27	.41	17	2
P15	.84	.86	.38	37	1
P16	.43	.39	.51	26	2
P17	.28	.28	.41	28	2
P18	.16	.18	.33	31	1
P19	.85	.76	.41	20	1
P20	.47	.51	.48	23	2
P21	.14	.10	.32	70	2

Note. Items P1, …, P21 can be found in Table 3.4. Item score is average good score. RiR is item-rest correlation. Words and number of visual elements are counted in the representation of the problem situation. Cronbach's α = .82 for these 21 items.

After the test run of the instrument the test as a whole was evaluated by analysing the average good scores of the items, the item-rest-correlations (RiR), and the correlation between the numbers of words, visual elements and test scores (see Table 3.5). The average good scores of the items range from .10 to .90, with the exception of P4, which was not included in the final instrument. The item-rest correlations are all positive and higher than .25. For the test of the 21 items we found Cronbach's α = .82 and for the test of the 24 items we found Cronbach's α = .83, which indicates a good internal consistency (Kline, 1999). For the diagnostic use of the instrument we considered that this was indicating sufficient criterion validity. We also checked the correlation of the scores on A-versions and B-versions and found r = .98, which contributes to both the content validity in measuring the same construct and the criterion validity in adding to the internal consistency of the instrument. In Table 3.2 these research activities are numbered 7 and 12.

Furthermore in the test run we investigated the relation between the number of words used in the representation of the problem situation in the A-version and the scores on the A-version of the items. We found a moderate correlation r= -.45 indicating that the word problems with more words were more difficult. Likewise we found r = -.34 for the correlation between the number of visual elements in the representation of the problem situation of the B-version and the scores on the B-version, indicating that the problems with more visual elements were more difficult. In this test run we found no correlation (r = .00) between the reduction of the number of words and the difference in performance between the two versions. In results obtained with the final instrument with validated items, this will be investigated in more depth.

3.5.3 The resulting final instrument

In constructing the final instrument another nine of the remaining 30 paired problems were discarded to meet the following criteria: evenly spread over the domains of numbers, proportions, and measurement & geometry; evenly spread over a bandwidth of levels around 2F; and discarding items for which the underlying mathematical structures were almost identical, for example “chicory” and “endive”, and “coffee cups” and “water bottles”.

Finally, with the combined results from the expert validation and the test run of the instrument the final instrument was composed and constructed as a web-based test. Ultimately 21 paired items on the domains of numbers, proportions, and measurement & geometry (see Table 3.4), and three additional items on the domain of relations were selected for the final instrument. The three additional items did not play a role in the eventual analysis of the results. In Table 3.2 this research activity is numbered 13.

The instrument was made available as an on-screen test on any computer connected to the internet. When a student participates in the test a personal activation code is provided to start the digital test of 24 problems and afterwards an optional additional short digital questionnaire is delivered to each participant to collect the following additional data: school level, grade level, gender, ethnicity, age, and last mathematics grade received (test or student report). All data are recorded anonymously in a research database. The design of the instrument makes it possible to compare the performances on word problems and the performances on image-rich numeracy problems in a controlled randomised way. In October and November 2011 the instrument was used in the Dutch context with a sample of over 32,000 students (see Chapter 4). With this article the English version of the instrument is made available under open access (Hoogland & De Koning, 2013).

3.6 Discussion

An important goal of mathematics education is to foster students' abilities to use mathematical knowledge and skills to solve problems from daily life. The practical use of mathematics is seen as one of the justifications for mathematics education in countries around the world (Kilpatrick, 1996). To reach this goal there is a need for adequate teaching materials and assessment tools. In those materials and tools quantitative problems from daily life are typically represented. For many decades, even centuries, those representations have been dominated by descriptive representations of reality; the resulting problems have been numeracy word problems that students had to read and make sense of. And although in educational research the use of word problems has been problematised (Gellert & Jablonka, 2009; Roth, 2009; Verschaffel et al., 2009), in classroom practice word problems are, without much discussion, seen as accepted parts of mathematics education, although the use of (too much) language is seen by many teachers as an additional difficulty, especially for low performing students. According to our findings, the dilemma of using word problems versus using real-life situations for teaching and assessing students in solving problems from daily life seems to have been "overcome"; a third way seems possible. The representation of a problem situation can get much closer to the real-life problem situation by using photographs. This is a potential way to bring more perceived authenticity into the mathematics classroom, without automatically encountering the practical constraints that occur when introducing real artefacts and simulations into the classroom. The aim of our research was to show a feasible alternative, by designing image-rich alternatives to existing word problems. Before we could conclude that such an alternative to word problems was feasible in the mathematics classroom, however, we had to gain more insights into the effect of such an alternative (in the way the problem situation is represented) on students' performance. To measure this effect, we needed an instrument that could validly measure such an effect. In this study we described the design of such an instrument that can measure the effect of changing the representation of the problem situation on the performance of the students in a controlled randomised way. This made it possible to analyse the effect also in relation to other possible intervening variables.

To validate the instrument's uses we took three pathways: firstly, we designed the alternative items and referenced them with the Dutch LaNF; secondly, two panels of experts commented on the problems used in the instrument, and validated the equivalence of the selected word problems and the alternative image-rich numeracy problems; thirdly, we undertook a test run with the instrument to see if the condition of a controlled randomised trial could be met. These activities around validations were carried out successfully, so that the final instrument can validly be used to measure the effect of changing the representation of the problem situation.

Our instrument also has limitations. The number of problems used for the instrument was limited. Further study is necessary to get a better view of which problems in particular were sensitive to the change of representation of the problem situation, which were not, and why.

In this article we contended that, despite the limitations, the instrument can reliably measure the difference in performance on two representations of the problem situation. There is, however, still a long way to go before we identify and understand all the underlying factors that actually could cause a difference in performance. In the analysis of the results the effect of task characteristics, like the content domain to which the task belongs, might be worth investigating. In the analysis we intend to take task characteristic variables such as the wordiness of the word problem or the number of images used for the image-rich representation as interdependent factors to identify the underlying patterns. This could lead to replications of the study with more and more specifically designed items.

At the same time more qualitative research should be done that focuses on the actual behaviour and thought patterns of students when they are solving these problems, for instance, by using thinking-aloud protocols or stimulated recall. This could shed more light on the intricate relation between use of language, problem-solving capabilities, and sense-making.

3.7 Conclusion

The instrument designed and validated in this study is likely to contribute to further knowledge and insights in terms of how the representation of reality in contextual mathematical problems can affect the performance of students, and as a consequence the outcomes of their assessment and the subsequent conclusions.

In follow-up studies this instrument was scheduled to be used in a large-scale test of students (over 30,000 participants), in which more data on the effects of changing the representation of the problem situation were gathered, also in relation to other measured characteristics of the students and of the items. By using a latent variable model and analysis, conclusions about the expected differences can be made with a good degree of validity. Apart from establishing the effect of changing the representation of the problem situation on students’ performance, we hope to find better and more concrete indications why the differences in effect appear, and why effects appear with certain types/genres of problems and not with other types/genres, for instance, related to the wordiness of the word problem or the number of pictures used in the image-rich tasks.

In our endeavour to get a better understanding of the effects of the representation of real-life problems in mathematical contextual problems, we have designed, trialled and tested an instrument that can give an indication of the effect of changing particular representations (from word to picture), and can also provide the opportunity to analyse this effect systematically in relation to other variables that influence students’ performances in solving problems from daily life. We invite researchers to use the English or other language translations of the instrument to investigate further the effects of representation of reality in mathematical contextual problems.

Chapter 4. Descriptive versus depictive representation of reality in contextual mathematical problems: The effect on students' performance

Abstract

Research on solving mathematical word problems suggests that students may perform better on problems with a close to real-life representation of the problem situation than on word problems. In this study we pursued real-life representation by a mainly depictive representation of the problem situation, mostly by photographs. The prediction that students perform better on problems with a depictive representation of the problem situation than on comparable word problems was tested in a randomised controlled trial with 31,842 students, aged 10–20 years, from primary and secondary education. A probit model was used to analyse the effects of the manipulated variable – descriptive versus depictive problem situation – against the background and non-manipulated variables, such as school type, grade level, gender, ethnicity, and age. The conclusion was that students scored significantly higher on problems with a depictive representation of the problem situation, but with a very small effect size of Cohen's $d = 0.09$. The chance of a fictitious participant answering an item correctly increased by approximately two percentage points, which with this number of participants and in this context may be considered a very small effect, comparable with the effect of ethnicity or gender.

This chapter has been submitted as a journal article and is under review.

Hoogland, K., Bakker, A., De Koning, J., Pepin, B., & Gravemeijer, K. (under review). Descriptive versus depictive representation of reality in contextual mathematical problems: the effect on students' performance.

4.1 Introduction

In mathematics education worldwide it is common practice (Verschaffel et al., 2000; Verschaffel et al., 2009) to use word problems to teach and assess students in the domain of solving quantitative problems from real life. Word problems can be defined as: “verbal descriptions of problem situations wherein one or more questions are raised, the answer to which can be obtained by the application of mathematical operations to numerical data available in the problem statement” (Verschaffel et al., 2000, p. ix)

Word problems can be seen as a special genre of contextual mathematics problems. Many studies and discourses (Gellert & Jablonka, 2009; Gerofsky, 2009, 2010; Gravemeijer, 1997; Greer, 1997; Roth, 2009), however, give rise to serious concerns as to whether word problems foster or accurately assess students’ potential to solve quantitative problems from everyday life. Students very often do not succeed in bridging the gap between their school mathematical knowledge and representations of real-life situations in word problems (Boaler, 1993). Studies show that students in a classroom setting persistently approach word problems with a calculational orientation (Thompson et al., 1994) and typically do not take into account realistic considerations about the problem situation (Cooper & Harries, 2002, 2003; Dewolf et al., 2014; Dewolf et al., 2015; Verschaffel et al., 1994). This behaviour leads to underachievement; the students do not make sense of the problem situation, and hence cannot show their full potential in solving problems from daily life. There are indications from several studies (Palm, 2002, 2006, 2008, 2009; Verschaffel et al., 2009) that this effect might be counteracted or avoided by making the problem situations or the representation of the problem situations more authentic for the students.

For the study reported here we designed an alternative to existing word problems whereby a descriptive representation of the problem situation, as is common in word problems, was replaced as much as possible with a depictive representation of the problem situation that we assumed to be closer to real life. In the design process we paid special attention to the real-life familiarity and relevance of the chosen images. We called the resulting problems “image-rich numeracy problems”. In the spirit of the aforementioned definition of word problems, we suggested the following definition for image-rich numeracy problems: Image-rich numeracy problems can be defined as visual representations of a problem situation wherein one or more questions are raised, the answer to which can be obtained by the application of mathematical reasoning to numerical data available in the problem representation.

In our study we compared students' performance on word problems with their performance on image-rich numeracy problems, which were mathematically equivalent to the word problems and in which images were used to make the problem situation closer to students' real-life experiences.

The primary aim of this study was to establish if the performance of students changed due to changing the representation of the problem situation. Given the difficulties that are reported with word problems, we predicted that the performances could be better, because some of the aforementioned difficulties with word problems and the resulting underachievement could be counteracted with the change of representation. The next section elaborates on the difficulties that are reported in using word problems in schools and how we explored ways to counteract them by changing the representation of the problem situation.

4.2 Theoretical background

4.2.1 Difficulties with word problems

Over the last 20 years the use of word problems in assessing mathematical potential in solving real-life problems has been intensively researched. Many research findings mention serious drawbacks of using straightforward word problems to assess students' mathematical potential (Verschaffel et al., 2000; Verschaffel et al., 2009). Numerous studies on student behaviour when solving word problems report various student blockages in the problem-solving process (Galbraith & Stillman, 2006). The most reported behaviour is that students base their analysis and calculations on a rather arbitrary association between certain salient quantitative elements of the problem situation and certain mathematical operations (Thompson et al., 1994; Verschaffel et al., 2000). Studies of word problems further show that students tend not to consider the possible constraints imposed by reality

(Caldwell, 1995; Cooper & Harries, 2003; Dewolf et al., 2015; Lave, 1992; Reusser & Stebler, 1997; Schoenfeld, 1992; Verschaffel et al., 1994; Verschaffel et al., 2009; Wyndhamn & Säljö, 1997). Students seem to value the bare outcome of the calculation more than the realism of the outcome. There is strong evidence that students approach word problems as "school maths" problems and not as problems from real life. This behaviour of students is reported to be reinforced by teachers' approaches that tend to be more on the mathematical structure of the problem than on the contextual aspects (Depaepe et al., 2010).

An explanation for this behaviour is the strong calculational orientation (Thompson et al., 1994) that for many decades has been, and to a great extent still is (Madison & Steen, 2008), dominant in mathematics classrooms. Thompson, Philipp, Thompson, and Boyd (Thompson et al., 1994) distinguish between a calculational approach in which the focus is primarily on procedures and operations with numbers, and a conceptual approach in which the focus is primarily on explaining, reasoning, and interpreting the quantitative situation. In the calculational approach, quantitative problems seem first and foremost to be used to train students in executing arithmetical operations. From this perspective, the representation of the problem situation is not an especially important aspect and the word problems are mainly straightforward. When teachers and students share a calculational orientation, they maintain a mathematics classroom culture based on implicit or explicit sociomathematical norms (Gravemeijer, 1997; Yackel & Cobb, 1995), in which solving word problems is limited to finding a calculation to perform without relating the outcomes to the original problems.

The use of superficial strategies and dissociation from reality in problem-solving with word problems has become known as "suspension of sense-making" (Reusser & Stebler, 1997; Schoenfeld, 1991; Verschaffel et al., 2000). This phenomenon is quite persistent in classroom situations and has generated serious concern as to whether students can show their full potential in solving quantitative real-life problems when confronted with word problems in a classroom or in an assessment setting.

4.2.2 Counteracting difficulties with word problems

Many attempts have been made to counteract or avoid the calculational approach and the suspension of sense-making, for instance, by adapting word problems into less straightforward problems; by adding instructions for students to take realistic considerations; and by changing the setting in which problem-solving takes place (Cooper & Harries, 2002, 2003; DeFranco & Curcio, 1997; Dewolf et al., 2011; Palm, 2009; Reusser & Stebler, 1997; Verschaffel et al., 2000; Wyndhamn & Säljö, 1997). To counteract suspension of sense-making, most of these studies suggest making the problem situation or the representation of the problem situation more authentic for students, hence helping them to make sense of it.

Other researchers and practitioners question the feasibility of counteracting suspension of sense-making by adapting word problems. They advocate the creation of real-life situations in mathematics lessons in order to teach and assess students' potential to deal with quantitative problems from everyday life (Bonotto, 2007, 2009; Frankenstein, 2009; Lave, 1992; Zevenbergen & Zevenbergen, 2009). Bonotto (2007) recommends encouraging students to analyse mathematical facts embedded in appropriate "cultural artefacts", such as supermarket receipts, bottle and can labels, railway schedules, or a weekly TV guide. Although many of the arguments for using real-life situations to teach students relevant problem-solving skills are convincing, there is no widespread dissemination of such practices. Practical constraints and the persistent classroom culture of "getting to the right answer as quickly as possible" are mentioned as the main reasons (Verschaffel et al., 2000).

4.2.3 Towards a more effective design of depictive representations

In our particular attempt to design problems using images from real-life situations to avoid suspension of sense-making we took into account other studies that investigated the effect of using pictorial elements in contextual mathematical problems. Some of these studies may indicate a mitigating effect on the decrease of suspension of sense-making that we predicted.

In the early stages of educational psychology there was a focus on individual characteristics. The Dual Coding Theory (Paivio, 1986), for instance, categorised people as visual learners (visualisers) or verbal learners (verbalisers). Massa and Mayer (2006), however, found no empirical evidence that in order to gain better results verbal learners should be given verbal instruction and visual learners should be given visual instruction. Instead they found that adding pictorial aids to an online lesson that was heavily text-based tended to help both visualisers and verbalisers. An earlier study by Plass et al. (1998) on learning a second language also concluded that a combination of text and pictures yielded better results in terms of learning outcomes than text alone. Research findings from this theoretical perspective add to the plausibility of our prediction. Mayer (2005) defined multimedia learning as learning from words (e.g., spoken or printed text) and pictures (e.g., illustrations, photos, maps, graphs, animation or video). Our image-rich numeracy problems fall into this definition because, next to the representation of the problem situation, the problem question is posed in words. Various studies from the perspective of multimedia learning suggest that, under the right conditions, the combination of text and pictures can promote better comprehension (Mayer, 2005).

There is, however, a variety of approaches in this domain. Kaminski, Sloutsky, and Heckler (2008) compared the effect of abstract and concrete representations of the cyclic operation in a mathematical group and concluded that transfer to a new abstract domain is better enhanced by abstract representations than by concrete representations. De Bock, Deprez, Dooren, Roelens, and Verschaffel (2011) replicated and extended that trial. They were able to show that this is only one side of the coin. They found that transfer to a new concrete domain is enhanced more by concrete representations than by abstract representations. These studies gave us valuable clues on the importance of choosing visual representations that are relevant to the task. In this respect the studies of Dewolf, Van Dooren, Ev Cimen, and Verschaffel (Dewolf et al., 2014; Dewolf et al., 2015) and Elia et al. (2007) are connected more closely to our research. They studied the effect of adding decorative, representational, and informational pictures to mathematical problems. They found no significant effects. These studies gave us valuable insights into the importance of using real-life images closely connected to the “real” situation that is to be represented.

We have also considered research investigating specifically the solving of (word) problems by students, for instance, students of mixed ability, and their use of pictorial elements in the problem-solving process (Hegarty, 2004; Krawec, 2014; Van Garderen, 2006; Van Garderen & Montague, 2003). The results in these studies corroborate our chain of reasoning. These studies showed that when students in the solving process use drawings that are structured and relevant to the mathematical model needed, the performance of students improves. However, when they make drawings that are merely pictorial or not relevant, their performance is worse, as is to be expected.

However, not all research on using text and pictures in lesson material points in the same direction. In recent years, research on performance in mathematical problems has been carried out from a Cognitive Load Theory perspective (Sweller, 2005, 2010). These studies show that redundancy of information, common when adding illustrations, can put extra cognitive load on students' working memories, and split-attention effects can occur when students have to jump between text and illustration elements (Berends & van Lieshout, 2009; Rasmussen & Bisanz, 2005; Scheiter et al., 2006). From this perspective, a mitigating effect could occur on the positive effects on student results that we predicted from using more real-life elements in the representation of problem situations.

An overall perspective is provided by research on the effect of depictive and descriptive representations on creativity and problem-solving (Schnotz, 2002; Schnotz et al., 2010; Schnotz & Bannert, 2003). Schnotz and Bannert (2003) concluded that task-appropriate graphics may support learning and task-inappropriate graphics may interfere with mental model construction. Schnotz et al. (2010) stated that to solve a quantitative problem, a task-oriented construction of a mental mathematical representation is necessary, provided that it is task-appropriate. The latter conclusion could also be an explanation of the results found, for instance, by Kaminski et al. (2008) in trying to represent operations in a mathematical group with partly filled cups. Schnotz's line of reasoning is that depictive representations can help students make a relevant mathematical mental model of the situation, and that depictive representations have a high inferential power because the information can "be read off more directly from the representation" (p. 21). This perspective also added to the plausibility of our prediction.

4.2.4 The images in our study

In our study we designed problems that can be used in regular teaching and assessment situations, and at the same time are closer to the real-life problem situations, by changing the description of the problem situation to a mainly depictive representation of the problem situation. We replaced the verbal – sometimes verbose – representation of the problem situation with a representation that is visually connected to the quantitative problem at hand. With technologies such as digital cameras and on-screen presentations, this has become a feasible option in regular teaching and assessment situations. We followed the reasoning in Palm’s Theory of Authentic Task Situations (Palm, 2009) which states: “that a strong argument can be made that the fidelity of the simulations (…) clearly has an impact on the extent to which students, when dealing with school tasks, may engage in the mathematical activities attributed to the real situation that are simulated” (p. 9).

In our depictive representations, the problem situation was represented with images from real-life situations, in the form of photographs, headlines from newspapers and “handwritten” notes (see Figure 4.1), and in that way augmenting the fidelity of the simulation, as advocated by Palm (2009). We distinguished our research from research on word problem solving that uses or adds images that are less realistic or sometimes possibly confusing for students (Berends & van Lieshout, 2009; Dewolf et al., 2014; Dewolf et al., 2015; Kaminski et al., 2008).

Considering the evidence from the abovementioned research, we expected that the use of images depicting real situations would increase the students’ association with real-life situations and problems and therefore would decrease suspension of sense-making and possibly the strong calculational orientation. In the trial, this was translated to the prediction that students scored better on our image-rich numeracy problems than on comparable word problems. This is elaborated on in the next paragraph.

4.2.5 The prediction

In this study we predicted that students would score better on image-rich numeracy problems than on comparable word problems. This prediction is based on the results of research on student behaviour in solving word problems. Theories and research from educational psychology do not seem to be conclusive on the positive effects for student results of changing from a descriptive representation to a depictive representation of the problem situation. In our research project we designed image-rich numeracy problems that were as close as possible to the realistic problem they represented (see Appendix A). Subsequently, we put our prediction to the test. Assessing students' potential to solve quantitative problems from daily life is a multifaceted problem in which many factors play a role. In this study one variable, namely type of representation of the problem situation, was systematically varied. In our analysis we therefore also took into account the interaction effects between the manipulated variable – description versus depiction – and background factors of the participants, such as school type, grade level, gender, ethnicity, and age. The number of collected data allowed for such an in-depth analysis, by which the possible intervening effects of these background variables could be ruled out. Considering the literature, we expected a positive but small effect. From our power analysis, therefore, we aimed at a number of participants well over 5,000, so that inferences about a small effect could be drawn (Ellis, 2010). This is elaborated on in the method section.

4.3 Method

Participants were presented with a web-based numeracy test, in which 21 problems came in either one of two versions: word problem or image-rich numeracy problem.

4.3.1 Design

The randomised controlled trial called for problems in two versions: a word problem version (WP) and an image-rich numeracy problem version (IRP). The design of these paired problems was conducted as follows. A batch of 40 paired problems in the mathematical domains of numbers, proportions, and measurement & geometry were constructed. A panel of eight independent experts in mathematics and numeracy education was asked to analyse the 40 paired problems and answer the following questions: Do the two versions of the problem test the same mathematical knowledge and skills? And if the two versions test the same mathematical knowledge and skills, are they testing on the same mathematical level? This led to 21 paired problems for which the first question was answered positively by all, or all but one, of the experts. The answers to the second question made it possible that in the final instrument (see Appendices B, and C) the selected problems were spread evenly around the domains of the LaNF.

The original 40 paired problems can be found in Appendix A and the final instrument with 21 paired problems can be found in Appendix B (in Dutch language), and in Appendix C (in English language). They are also available under open access (Hoogland & De Koning, 2013). To each participant, a test was delivered of 21 problems presented in either the IRP version or the WP version, with a minimum of 9 and a maximum of 12 of each version, randomly assigned. The order of the problems in the test situation was again randomised for each participant. By this design the trial fulfilled the conditions of a randomised controlled experiment – the characteristics of the participants answering the IRP version of a particular problem and the characteristics of the participants answering the WP version of that problem are highly likely to be the same. This holds for the measured characteristics as well as for the characteristics that were not measured. The manipulated variable is the version of the problem (WP or IRP); the dependent variable is the student's scores on the problems.

4.3.2 Participants

In a four-week period at the end of 2011, the test was available on the internet for schools to use as a numeracy test. In total, 31,842 participants, from 179 schools geographically spread across the Netherlands, took the test. Table 4.1 shows the number of participants from different age groups and different educational streams in the Dutch school system. From the total student population in the Netherlands (CBS, 2012) around 2% participated.

In the Dutch school system, primary education is for 4- to 12-year olds and runs over eight grades (K-6). In secondary education, the Netherlands has a highly streamed school system. VMBO is a (pre-)vocational education stream for12- to 16-year olds; HAVO and VWO are the general secondary and pre-university streams that prepare children for college and university respectively. MBO is a tertiary vocational stream that is a follow-up to VMBO and is intended for 16- to 20-year olds.

Table 4.1
Number of participants in school types and age groups

School Type	**Subtype**	**Age**	***n***
Primary (BO)		11–12	969
Pre-vocational (VMBO)	Low (VMBO-BB)	12–16	1,932
	Middle (VMBO-KB)	12–16	2,658
	High (VMBO-GT)	12–16	7,869
General secondary (HAVO)		12–17	8,918
Pre-university (VWO)		12–18	7,670
Vocational (MBO)		16–20	1,146
Unknown			680

Note. *n* is number of participants

Of the participants, 49.5% were male, 49.1% were female and for the remainder the gender is unknown. In line with the definition by CBS (CBS, 2012) a participant was considered to be of a migrant family if the participant or one of his/her parents was born outside the Netherlands. Of the participants, 23% were considered to be of a migrant family and 77% not. The percentages of gender and ethnicity in the various school types were close to the national percentages (CBS, 2012). Therefore, we assumed participating schools to be representative of Dutch schools in general.

4.3.3 Tasks

Each problem in the web-based test was presented as screen-filling. The question was posed at the bottom of the screen. Below the question, the numerical solution to the problem could be entered and was computer scored as right (1) or wrong (0). Figure 4.1 provides three examples of paired problems with a WP version and an IRP version.

4.3.4 Procedure

Each participant was assigned a personal activation code to begin the digital test with the 21 problems. Participants conducted the test on-screen at an internet-connected PC. For the total test, a time limit of 60 minutes was set, which was sufficient for the participants to complete the test. An online calculator was allowed in solving the problems. All answers to the problems were numerical values, to be entered into an answer field by the participants.

The participants' answers for each problem were scored and recorded. After finishing the test, a short digital questionnaire was administered to each participant to collect the following additional data: school type, grade level, gender, ethnicity, age and maths grade (test or student report) last received. All data were recorded anonymously in a research database.

1A

Apples are sold in bags of 2.5 kilograms. You weigh one apple and find it weighs 157 grams.

About how many apples are there in the bag?
[ ] **apples**

1B

About how many apples are there in the bag?
[ ] **apples**

6A

For a picnic you found a recipe for wraps. The recipe gives the ingredients for 5 persons: 2 packs of wraps, 250 grams of cream cheese, 1 sachet of garden herbs, 300 grams of fricandeau, green lettuce, pepper and salt to taste

How much cream cheese do you need for 12 persons?
[ ] **grams**

6B

How much cream cheese do you need for 12 persons?
[ ] **grams**

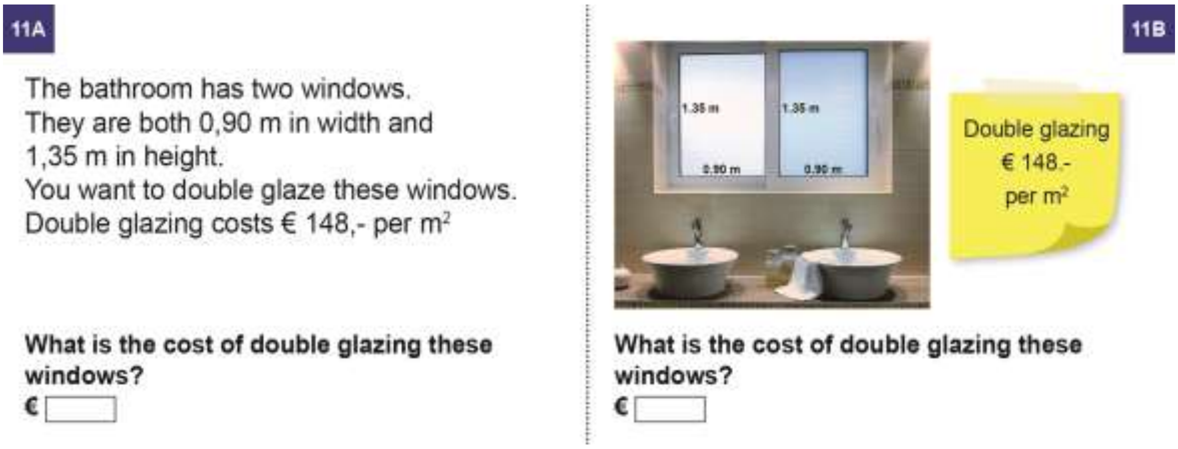


Fig. 4.1 Three examples of paired problems: Word Problem (WP) and Image-Rich Problem (IRP). Translated to English for readability.

After the experiment the test as a whole was evaluated by analysing the mean scores of the items and the item-rest-correlations. This analysis showed that none of the items failed to fulfil the requirements of an acceptable item. The mean scores of the items range from .04 to .84 and all items had positive item-rest correlations between .05 and .54. The Cronbach α for the test was .84, which indicated good internal consistency (Kline, 1999).

4.3.5 Statistical analysis

To test the prediction, we first carried out a paired-samples *t*-test on the mean scores of the students on their IRP version items and their WP version items. To get more in-depth insight into the effects of the background variables on the results and their possible interaction effect with the manipulated variable, we constructed a probit model, which is a limited dependent variable model (Long, 1997). In a probit model the following assumptions are common (Borooah, 2002): the probability that a student solves a given problem is determined by his or her "proven ability" of crossing a certain threshold. Proven ability may differ from true ability because students may not be able to show their true ability fully. Our prediction suggested that this is the case with word problems more than with image-rich numeracy problems, because students' abilities were impaired by characteristics of word problems that hinder the demonstration of their full potential. Proven ability *y* was not observed directly, but was only indirectly reflected in the student's results in solving the presented problems. We assumed that a problem is answered correctly if proven ability *y* crosses a threshold value δ. The dichotomous variable *z* took the value 1 when the problem was answered correctly and 0 when it was answered incorrectly. Hence $z = 1$ if $y \geq \delta$ and $z = 0$ otherwise. We further assumed that proven ability *y* was normally distributed and depended on several background variables that as non-manipulated independent factors contributed to the measured outcomes on the problems: school type, grade level, gender, ethnicity, age and maths grade last received. This led to the following model equation: $y = \alpha_0 + \alpha_1 v + \alpha_2 x + \varepsilon$. In this equation, *v* is the manipulated dummy variable regarding the version of the problem: IRP (depictive, $v = 1$) versus WP (descriptive, $v = 0$). The vector *x* can consist of one or more of the non-manipulated independent variables presented in Table 4.2.

Table 4.2
Overview of recorded non-manipulated independent variables

Variable		Explanation	Coding
BO	$\alpha_{2,1}$	primary education	dummy
VMBO-BB	$\alpha_{2,2}$	pre-vocational, lower level	dummy
VMBO-KB	$\alpha_{2,3}$	pre-vocational, intermediate level	dummy
VMBO-GT	$\alpha_{2,4}$	pre-vocational, higher level	dummy
HAVO	$\alpha_{2,5}$	secondary general	dummy
VWO	$\alpha_{2,6}$	pre-university education	Dummy
MBO	$\alpha_{2,7}$	secondary vocational	Dummy
Grade	$\alpha_{2,8}$		BO(5-6), VMBO(1- 4), HAVO(1-5), VWO(1-6), MBO(1-4)
Gender	$\alpha_{2,9}$		0 = female; 1 = male;
Ethnicity	$\alpha_{2,10}$		0 = not migrant family; 1 = migrant family;
Agedev[a]	$\alpha_{2,11}$	age relative to average age in grade level	0-3
Maths grade	$\alpha_{2,12}$	maths grade last received	1-10

Note. $\alpha_{i,,j}$ is the corresponding coefficient in the probit model. For all dummy variables the coding is: 1 = in this level, 0 = not in this level. [a]Agedev is the deviation of the age of a participant to the average age of all participants in the same grade level.

The error term ε represented unobserved variables affecting ability. In the model ε can be treated as a random variable with a normal distribution with mean 0 and y is scaled in such a way that the variance of ε is equal to 1. So, y defined a probit model from which consistent estimators for the coefficients α_1 and α_2 can be obtained by maximising the likelihood function. The participants were asked for their age, but since age is dependent on grade level, we changed the variable age to relative age (Agedev) as the deviation of a participant's age to the average age of all participants in the same grade level. Maths grade was the last test grade or report mark the participant received for his or her mathematics performance. It was an indication of the mathematical ability of a participant relative to the abilities of other participants in the same school level and grade level.

We can then from the model $y = \alpha_0 + \alpha_1 v + \alpha_2 x + \varepsilon$ formulate the probabilities of answering an item correctly:

$$P(z = 1) = P(y \geq \delta\ ) = P(\varepsilon \geq \delta - \alpha_0 + \alpha_1 \mathrm{v} + \alpha_2 \mathrm{x}\ ) \text{ and}$$

$$P(z = 0) = 1 - P(y \geq \delta\ ) = 1 - P(\varepsilon \geq \delta - \alpha_0 + \alpha_1 v + \alpha_2 x).$$

The unknown coefficients $\delta - \alpha_0$, α_1, and α_2 were estimated by maximum likelihood. In all our analyses these coefficients are calculated using STATA11 (probit, dprobit).

4.4 Results

We first conducted a classical analysis of the test results to get an overall idea of the average students score on the word problems and on the image-rich numeracy problems. Second, we conducted an in-depth analysis using a probit model, taking into account the possible effects of other variables and the possible interdependency of the involved variables.

4.4.1 Correct scores of the participants

We analysed the data as a set of 31,842 paired observations, where for each participant a pair of observations consisted of the mean score of that participant on the WP version items and the mean score of that participant on the IRP version items presented in his or her test. We conducted a paired samples *t*-test on the mean correct scores for items of both versions and found M(A) = .436 (.234), M(B) =.455 (.237), and a difference of .019(.202) with t = 16.84, which was statistically significant (p < .001), and we found as effect size Cohen's d = 0.09. This result indicated that the students' average performance on the IRP version items was a statistically significant two percentage points higher than on the WP version items. The data therefore supported our prediction. However, the effect size of Cohen's d = 0.09 could be considered very small.

In Table 4.3 we presented the results of the total correct scores which we split out for the main conditions: school type, gender, and ethnicity. For all these subgroups we found the same pattern. In the probit model in the next section we further investigated other interdependence of the variables involved in a model approach.

Table 4.3
Correct scores of participants on A-version and B-version items

		A-version ***M*** **(SD)**	**B-version** ***M*** **(SD)**	***N***
Overall		.44(.44)	.45(.45)	31,842
Condition	BO	.24 (.42)	.25(.43)	969
	VMBO_BB	.18(.38)	.19(.39)	1,932
	VMBO_KB	.26(.44)	.28(.45)	2,658
	VMBO_GT	.36(.48)	.37(.48)	7,869
	HAVO	.49(.50)	.50(.50)	8,918
	VWO	.60(.49)	.62(.48)	7,670
	MBO	.52(.50)	.54(.50)	1,146
	Gender = f	.42(.49)	.44(.50)	15,637
	Gender = m	.46(.50)	.48(.50)	15,766
	Ethnicity = not migrant family	.45(.50)	.47(.50)	24,183
	Ethnicity = migrant family	.39(.49)	.41(.44)	7,220

Note. N is number of participants. M is mean score over all participants and per condition with standard deviation in parenthesis.

4.4.2 Results from the probit analysis of three expanding models

The probit model was used to investigate how the background variables, alongside the manipulated variable (IRP version versus WP version), contributed to the correct scores on the items. In the probit model we considered the data as a set of 31,842 times 21, which is 668,682 item scores. In this approach we considered the trial as 668,682 observations of student behaviour on individual items. The probit model analysis related the probability that a student solves an item correctly to a set of independent variables.

In Table 4.4 we presented three runs of the probit analysis with an expanding number of non-manipulated independent variables. In model 1, VMBO-BB, VMBO-KB, VMBO-GT, HAVO, VWO were treated as an aggregated group.

Table 4.4
Coefficients of manipulated and non-manipulated variables in three models

Variable		**Coefficients model 1**	**Coefficients model 2**	**Coefficients model 3**
Version	α_1	0.05***(0.00)	0.05***(0.00)	0.05*** (0.00)
BO	$\alpha_{2,1}$	-0.40***(0.01)	-0.52***(0.01)	-0.60***(0.01)
VMBO-BB	$\alpha_{2,2}$		-0.92***(0.01)	-0.95***(0.01)
VMBO-KB	$\alpha_{2,3}$		-0.62***(0.01)	-0.64***(0.01)
VMBO-GT	$\alpha_{2,4}$	ref. cat.	-0.33***(0.00)	-0.34***(0.00)
HAVO	$\alpha_{2,5}$		ref. cat.	ref. cat.
VWO	$\alpha_{2,6}$		0.29***(0.00)	0.26***(0.00)
MBO	$\alpha_{2,7}$	0.35***(0.01)	0.23***(0.01)	0.26***(0.01)
Grade	$\alpha_{2,8}$	0.22***(0.00)	0.22***(0.00)	0.24***(0.00)
Gender	$\alpha_{2,9}$	.		0.13***(0.00)
Ethnicity	$\alpha_{2,10}$			-0.04***(0.00)
Agedev	$\alpha_{2,11}$			-0.01***(0.00)
Maths grade	$\alpha_{2,12}$			0.06*** (0.00)
Unknown var.	α_0^*	-0.66***(0.00)	-0.55***(0.01)	-1.02*** (0.01)
pseudo-R^2		.02	.07	.08

Note. αi,j is the coefficient in the probit model, $\alpha_0^* = \delta - \alpha_0$. Coefficients are calculated with maximum likelihood by STATA11, standard errors are in parentheses. Ref.cat is reference category. The measure of good fit pseudo-R^2 is McFadden's R^2. All variables are significant *p< .05, **p< .01, ***p< .001.

From Table 4.4 we concluded that the positive effect of the IRP version on a correct score on an item is significant and stayed significant if other variables were taken into account. As a caveat note that the values of the coefficients are not indicative of the magnitude of the effect, but the sign of the coefficients give its direction. Table 4.4 showed that adding variables hardly has any effect on the coefficients. This meant that the model and the outcomes are quite robust and only depended to a small extent on a particular choice of variables.

Table 4.5 presented the marginal effects calculated from model 3. In the probit model the marginal effect of a variable is the increase in probability of answering an item correctly under the condition that all other variables are at their mean. The marginal effects can be considered as a measure of effect size, but this must be done with great prudence (Hoetker, 2007).

Table 4.5
Marginal effects of manipulated and non-manipulated variables in probit model of participants correct scores

Variable		**Marginal effects**
Version	α_1	.02 ***(.00)
BO	$\alpha_{2,1}$	-.22*** (.00)
VMBO-BB	$\alpha_{2,2}$	-.32*** (.00)
VMBO-KB	$\alpha_{2,3}$	-.23*** (.00)
VMBO-GT	$\alpha_{2,4}$	-.13*** (.00)
HAVO	$\alpha_{2,5}$	ref. cat.
VWO	$\alpha_{2,6}$	.10*** (.00)
MBO	$\alpha_{2,7}$	.10*** (.00)
Grade	$\alpha_{2,8}$	.09*** (.00)
Gender	$\alpha_{2,9}$	.05*** (.00)
Ethnicity	$\alpha_{2,10}$	-.02*** (.00)
Agedev	$\alpha_{2,11}$	-.00***(.00)
Maths grade	$\alpha_{2,12}$	.02*** (.00)

Note. Marginal effects are calculated with dprobit in STATA11, standard errors are in parentheses. $\alpha_{i,j}$ is the coefficient in the probit model. Ref. cat. is reference category. All variables are significant **p*< .05, ***p*< .01, ****p*< .001.

The marginal effects presented in Table 4.5 can be interpreted as follows. The variable version had two values: 0 for the word problem version and 1 for the image-rich version. Changing the representation of the problem from the WP version to the IRP version resulted in an increase of two percentage points in overall students' performance. Let us compare this with the effect of gender. The variable gender has also two values: 0 for female and 1 for male. The conclusion is that male students scored overall five percentage points higher than female students on the problems presented.

From the marginal effects we concluded that when the problem is presented with the IRP version rather than with the WP version, the chance of a fictitious participant answering an item correctly increased by around two percentage points, assuming that all other variables are at their mean. This was consistent with what we found in Table 4.3. The variables regarding school type (BO, VMBO-BB, VMBO-KB, VMBO-GT, HAVO, VWO, MBO) were statistically significant and the sign of the coefficient was consistent with the increasing levels. The variable gender was statistically significant and the coefficient indicated that males had a higher chance than females of getting an item correct. The variable ethnicity was statistically significant and the coefficient indicated that the chance of answering an item correctly was lower for participants from migrant families than for participants from non-migrant families. The variable maths grade was statistically significant and the coefficient indicated that participants with a higher maths grade had a higher chance of answering an item correctly.

Overall we concluded that the variable version was significant, but with a very small effect. This effect was comparable to the effect of being from a non-migrant family or of having a higher maths grade. The effect was less than that of gender. In the next section we analysed in more depth the interaction effects of these variables.

4.4.3 Results from the probit analysis with variable interaction terms

A more in-depth analysis was carried out by investigating how the manipulated variable version interacts with the non-manipulated variables. Probit models are non-linear models in which the effect of each individual explanatory variable depends on the other variables. Hence, interaction effects are always non-zero even if we do not include interaction terms. Interaction terms can be used to infer whether the independent variables are interdependent (Norton, Wang, & Ai, 2004).

First, we added in the probit model the interaction terms of the variable version with the variables that are concerned with school type. Table 4.6 showed that the interaction terms were not significant, with the exception of the interaction term “VMBO-KB,” which was just significant. We do not have an explanation for why participants in this school track had a bigger interdependent effect for the effect of version. Although interpreting interaction terms in non-linear models is hazardous (Ai & Norton, 2003; Greene, 2010), the global picture was that the effect of version was not interdependent with school type.

Table 4.6
Probit model coefficients and marginal effects of manipulated and non-manipulated variables on school type

Variable	**Coefficient**	**Marginal effect**
Version	0.05 (0.01)	.02 (.00)
BO	-0.52***(0.01)	-.19*** (.00)
VMBO-BB	-0.93***(0.01)	-.31*** (.00)
VMBO-KB	-0.64***(0.01)	-.23*** (.00)
VMBO-GT	-0.33***(0.01)	-.13*** (.00)
HAVO	ref. cat.	ref. cat.
VWO	0.28***(0.01)	.11*** (.00)
MBO	0.23***(0.01)	.09*** (.00)
Grade level	0.22***(0.00)	.09*** (.00)
BO ∗ Version	-0.01 (0.02)	.00 (.01)
VMBO-BB ∗ Version	0.02 (0.02)	.01 (.01)
VMBO-KB ∗ Version	0.03* (0.01)	.01* (.01)
VMBO-GT ∗ Version	0.00 (0.01)	.00 (.00)
HAVO ∗ Version	ref. cat.	ref. cat.
VWO ∗ Version	0.01 (0.01)	.00 (.00)
MBO ∗ Version	0.01 (0.02)	.00 (.01)
Unknown Variables	-0.54*** (0.01)	

Note. Coefficients (and standard errors) and marginal effects (and standard errors) are displayed. Ref. cat. is reference category. Variables are significant *p< .05, **p< .01, ***p< .001, variables’ interactions are not significant, with the exception of VMBO-KB * Version.

Next, we added in the probit model the interaction terms of the variable version with the variables gender, ethnicity, age deviation, and maths grade last received. Table 4.7 showed that the gender $*$ version interaction is significant, which indicates that the effect of version is partly dependent on the effect of gender in the direction of higher differences with male participants. We specified the results for males and females from the general results from Table 4.3 and found that for males the difference in mean score on IRP versions and mean score on WP versions was M = .022 (SD = .202) and for females M = .016 (SD = .202), which is consistent with the finding from the interaction terms. This result calls for more profound research in the future.

Table 4.7
Probit model coefficients and marginal effects of manipulated and non-manipulated variables on participants' background

Variable	Coefficient	Marginal effect
Version	.05** (.02)	.02** (.01)
Gender	.10*** (.00)	.04*** (.00)
Ethnicity	-.14*** (.01)	-.05*** (.00)
Agedev	-.12*** (.00)	-.05*** (.00)
Maths grade	.02*** (.00)	.01*** (.00)
Gender $*$ Version	.02** (.01)	.01** (.00)
Ethnicity $*$ Version	-.00 (.01)	-.00 (.00)
Agedev $*$ Version	.01 (.00)	.00 (.00)
Maths grade $*$ Version	-.00 (.00)	-.00 (.00)
Unknown Variables	-.33 (.01)	

Note. Coefficients (and standard errors) and marginal effects (and standard errors) are displayed. Variables are significant *p< .05, **p< .01, ***p< .001, interactions are not significant, except for gender $*$ version.

In our analyses we have dealt with all forms of within and between variation given the variables included in the database. We checked especially for the between variation between school types and separate schools to investigate whether better scores on IRP versions could be explained by the fact that some schools specially trained students for these kinds of problems. This could be translated to the question whether the variable school was interdependent with the variable version. In the probit model we checked for this interclass effect by incorporating school as a different categorical class in the model and looking at the interaction terms with the variable version. This analysis produced the result that school as a variable was statistically significant for performance, which meant that schools make a difference. The interaction term school $*$ version was, however, not significant. There was no indication of a systematic interdependence of the variables version and school.

4.5 Discussion and conclusion

Students' difficulties with the genre of word problems is a serious educational challenge. These difficulties are becoming increasingly relevant since word problems are used more and more for high stakes testing and international comparison studies (Hoogland & Stelwagen, 2012; OECD, 2012; Palm & Burman, 2004; PIAAC Numeracy Expert Group, 2009). Several researchers (Gravemeijer, 1997; Greer, 1997; Palm, 2008, 2009) have theorised on improvement of word problems in assessment situations and the way they are used in classroom practice. In addition, there is ample empirical research on the effects on students' performance of changing the design of the problems or of changing the classroom setting in which the problem solving takes place (DeFranco & Curcio, 1997; Reusser & Stebler, 1997; Wyndhamn & Säljö, 1997). A common factor in the findings is that making the problems more real-life could reduce both the "suspension of sense making" (Schoenfeld, 1991) and the predominant calculational approach (Thompson et al., 1994), and could result in better student performance.

The present study extended that body of knowledge by investigating the effects of changing specifically one aspect of existing word problems – the way in which the problem situation is represented – from primarily descriptive or verbal, as is common in word problems, to primarily depictive, using photographs as representations of real-life situations. This is commonly used in contemporary multimedia lesson materials and multimedia assessment tools (Mayer, 2005). In our study we paid special attention to the authenticity and the relevance of the depictive elements. Using representations of authentic situations in classroom situations is one of the core elements of many educational practices, and is often considered to be a condition of knowledge transfer. At the same time, one can argue that every representation of "reality" creates its own reality or its own genre of realities (Gerofsky, 2009, 2010). Nevertheless, there are indications that more depictive representation of the problem situation can counteract to a certain extent the difficulties students have with word problems. From a cognitive psychology perspective it is argued (Schnotz et al., 2010) that a more depictive representation can help students make a relevant (mental) mathematical model of the situation, which can lead to more success in solving the problem.

In other studies on using more depictive representations in (word) problem solving, lower scores (Berends & van Lieshout, 2009; Kaminski et al., 2008) or no effect (Dewolf et al., 2014; Dewolf et al., 2015) have been reported. These studies differed from our study in the sense that images in our studies were close to a real-life problem situation, relevant and an integral part of the problem situation (see Appendix A). In our design of image-rich numeracy problems we tried to keep as close as possible to the actual problems to be solved, to keep students as much as possible in a "problem-solving mindset". The results of our study can also be contrasted with results of studies that focus on the use of drawings by students as part of the problem-solving process. Most of those studies found that when students used relevant and structured diagrams they performed better, and when they used more pictorial sketches they performed worse (Boonen, van Wesel, Jolles, & van der Schoot, 2014; Elia et al., 2007; Hegarty, 2004; Krawec, 2014; Van Essen & Hamaker, 1990; Van Garderen, 2006). In all these cases the relevance and appropriateness of the depictive elements were of major importance.

From the perspective of Cognitive Load Theory (Sweller, 2010) one can argue that redundancy and split-attention effects due to using visual representations of the problem situation in a classroom problem, could have a negative effect on better result. Weighing the arguments of the different perspectives, in the present study the prediction was that students would score better on numeracy problems where the problem situation was represented as closer to real life by using relevant and "realistic" depictive elements.

Reviewing the literature, we expected a small effect. This prediction was tested by a randomised controlled trial with over 32,000 students. Next to the more straightforward analysis, the data were analysed using a probit model so that possible interacting variables could be investigated.

The effect of a mainly depictive representation on participants' correct scores was statistically significant, but this was not surprising given the large sample. The effect size found was very small ($d = 0.09$). We elaborated on the found effect size; in this study the chance of solving a problem correctly was about two percentage points higher when the representation of the problem situation was of a depictive nature. Although the effect was small, the result of a two percentage-points increase is a noteworthy effect, in particular in large-scale assessments. In the context of this study we expected a small effect.

Changing the representation of the problem situation was not a teaching intervention, unlike the teaching programmes and long-lasting interventions that were compared, for instance, by Hattie (2009, 2015), who reported an average effect size of .40 for major educational interventions that ranged over several months. It would be surprising or even dubious if the changes we made should have an effect size that could compare with that. The effect size found is comparable with the findings of Slavin (2016) on large randomised studies in education. Because of the expected small effect size, we paid special attention to the power analysis. To be able to draw inferences about such small effect sizes a large sample is necessary. Ellis (2010) suggests a sample of well over 5,000 participants to be able to draw inferences about effect sizes around $d = 0.10$. The sample in our study was big enough to fulfil these conditions. The measured rise in performance of 1.9 percentage points in our study came with a 95% confidence interval of [1.7; 2.1]. Combining this with the fact that a two percentage-point rise in performance in large-scale testing could make the difference between passing and failing for many students – in our sample around 600 students –, we were inclined to conclude we had a noteworthy effect, which needed, however, further research and analysis to come to possible recommendations for practical use.

This conclusion was strengthened further by excluding other factors that could cause this effect. Because of the large number of participants, we were able to investigate whether background variables of the participants were interdependent with the variable regarding version. Only the variable gender had an interaction effect with the variable version, indicating that gender could to some extent play a role in the difference in scores on the versions, in the sense that the difference in performance between the two versions (IRP minus WP) is slightly greater for males than for females. Given the important issue of gender in mathematics education (Lindberg, Hyde, Petersen, & Linn, 2010), this was a result that should be investigated further. By using interaction variables with background variables we found that the measured result is robust, meaning that the variable version was the sole factor, except gender, that explained the increase in students' results.

In conclusion, we found that the students' better scores on our image-rich numeracy problems than on comparable word problems could be attributed with high certainty to the representation of the problem situation. This result supported our prediction that using visual elements in the representation of the problem situation close to a real-life problem situation could possibly prevent to a certain extent the suspension of sense-making and the calculational approach that occur so frequently when students are solving word problems in classroom and assessment situations.

Further research could investigate whether specific task characteristics had an influence on the measured effect. For instance, the domain of the tasks and the wordiness of the word problems could be of effect. Other future research to find possible explanations for the measured effect could be in the realm of solution times, strategy selection, problem-solving behaviour, or other explanations on the cutting edge of mathematics education and cognitive and neurosciences (De Smedt et al., 2010; Imbo & Vandierendonck, 2007)

The results of this study should also be interpreted with an eye on their limitations. One limitation was the small number of 21 paired items used in the trial. If we had anticipated that more than 30,000 students would participate, we could have pooled more items, a strategy used in international assessments such as PISA (OECD, 2012) and PIAAC (PIAAC Numeracy Expert Group, 2009). Furthermore, the study focused on students' scores and not on their actual behaviour. The actual behaviour of students when solving image-rich numeracy problems could be quite different from their behaviour in solving word problems. Investigating this behaviour, for example by observation or by eye-tracking, could offer more detailed explanations for the results in this study. It could also give further insights into the balance between the positive effect on the scores from decreasing suspension of sense-making and the possible negative effects on the scores caused by redundancy and split-attention effects.

In our view, our findings can act as a good starting point for further research into the effects of representation of problem situations on students' performances. Furthermore, the results could be an incentive for designers of teaching materials and assessments to (re)consider the chosen representations of reality in their products. Overlooking the body of literature on solving contextual mathematical problems and the role of depictive elements in the design of the problems and/or in the solving of the problems by the students, we suggested a short checklist regarding the images used:

- Are the images realistic and relevant for solving the problem (in contrast to cartoon-like and distracting)?
- Are the images an integral part of the representation of the problem situation (in contrast to mere illustration)?
- Are the images relevant and consistent with the envisioned mathematical concepts and mathematical models that are involved in the problem solving (in contrast to irrelevant or even confusing)?
- Using such a checklist could make studies on this topic and their results more understandable and comparable.

In conclusion, we may argue that, with better design and a better knowledge of the underlying factors that explain students' results in solving quantitative problems from daily life, the approach of using mainly depictive representations has the potential of offering a better way of fostering or assessing students' potential to solve quantitative problems from daily life.

Chapter 5. Word problems versus image-rich numeracy problems: An analysis of effects of task characteristics on students' performance

Abstract

The objective of this study was to gain insights into the effects of task characteristics on students' performance after changing the representation of the problem situation from solely descriptive (word problems) to mainly depictive. The empirical basis for this study consists of data gathered in a randomised controlled trial with 31,842 students in the Netherlands with an instrument consisting of 21 paired problems. Analysis of task characteristics indicated that wordiness and number of pictures were related to the performance of the students and to the difference in performance after changing the representation of the problem situation from descriptive to depictive. Likewise, differences in performance were found related to the content domain of the problems. One of the tentative conclusions was that for image-rich numeracy problems in the domain of measurement & geometry the inferential step from representation of the problem situation to the mathematical problem is smaller than for word problems.

This chapter has been submitted as a journal article and is under review.

Hoogland, K., Pepin, B., De Koning, J., Bakker, A., & Gravemeijer, K. (under review). Word problems versus image-rich problems: An analysis of effects of task characteristics on students' performance on contextual mathematics problems.

5.1 Introduction

Over the past decades problem solving has gained importance as one of the main goals of mathematics education (Santos-Trigo, 2014; Schoenfeld, 2007, 2014). For both teaching and assessment purposes, problem-solving tasks have been designed by mathematics educators worldwide, especially in the field of numeracy and mathematical literacy (Askew, 2015; Geiger et al., 2015). The aim of these problem-solving tasks is that students solve the posed problems using their mathematics knowledge and skills. In some cases, the problem will be a routine task for the solver, whereby adequate procedures must be executed in a correct way. In other cases, the problem will be a non-routine task for the solver, whereby higher order, problem solving and heuristic skills must be put to use. In a typical problem-solving task a problem situation is presented, followed by a problem question.

For centuries the predominant representation of a problem situation has been strictly verbal (Verschaffel et al., 2000; Verschaffel et al., 2009) – the typical consequence of using pen and paper, a type writer or a simple word-processor to design tasks. This has led to the genre of 'word problems'. According to Verschaffel et al. (2014) word problems can be defined as: "verbal descriptions of problem situations wherein one or more questions are raised the answer to which can be obtained by the application of mathematical operations to numerical data available in the problem statement" (p. 641). In word problems both the description of the problem situation and the actual problem statement are presented in words. The predominance of the verbal description of problem situations can be demonstrated most clearly by the intriguing fact that the concept of *word* problem in education almost exclusively means a *contextual mathematical* problem.

Even though there are many studies which report serious difficulties with the use of word problems in the mathematics classroom, word problems are used to a great extent. It is often reported that students are not inclined to adopt a problem-solving attitude, when tackling these problems. In many cases they fall back on an "answer-getting" attitude (Daro, 2013). They seem to restrict themselves to a merely calculational approach of pulling the numbers out of the verbal problem representation and performing operations (they know well or have used in a similar context), without making sense of the situation.

In our research project we are in search of an alternative to word problems. We took the approach of systematically replacing verbal representations of the problem situations by depictive ones, with the idea that depictive representations of the problem situations stay closer to the real problems that are represented, and that students are more likely to make sense of a pictorial situation. Subsequently, they may be more likely to adopt a problem-solving attitude, with more chances of solving the posed problem.

Chapters 3 and 4 showed in detail the design and validation of the instrument with which we measured the effect of changing the representation of the problem situation on students' performance in a randomised controlled way. The instrument contained 21 paired problems, and each pair consisted of a word problem and an equivalent image-rich numeracy problem. The equivalence regarded the content and the level of the problem; the only variation was the way in which the problem situation was represented and presented to the participants. This instrument will be explained further in the method section. Using this instrument, we found that the students performed significantly better on image-rich numeracy problems than on otherwise equivalent word problems, with an increase of good scores of about two percentage points, while the effect size was low (Cohen's d = 0.09). This result was a robust general overall result, in the sense that the measured effect was not interdependent with background variables like age of students, ability of students, ethnicity and (type of) school, and only to some extent with gender.

In the current study we conducted a more in-depth analysis of these results, investigating a set of task characteristics of the items used in the instrument. In this way we aimed at getting more insights into which task characteristics could explain the increase in students' performance on the equivalent image-rich numeracy problems. The item characteristics we explored were task complexity and the content domains of the tasks.

5.2 Theoretical and empirical background

This study is part of a larger research project to investigate alternatives to the persistent and problematic use of word problems to teach and to assess students' abilities to deal with numerical problems based on everyday life situations. In current classroom practice predominantly word problems are used to teach and assess these abilities, although many research findings since the 1990s have reported serious difficulties in using word problems for this purpose (Verschaffel et al., 2000; Verschaffel et al., 2009). The reported difficulties can be related to the steps the problem solver is expected to take to solve the task at hand. Figure 5.1 shows a diagram in which the most important steps of the problem-solving process are laid out, keeping in mind that the problem-solving process most certainly will not take place in the problem solver's mind as a linear or strictly circular process. Many similar diagrams are used in research on problem solving in mathematics education (Blum et al., 2007; Burkhardt, 2006; Lesh, Post, & Behr, 1987; Tout & Gal, 2015).

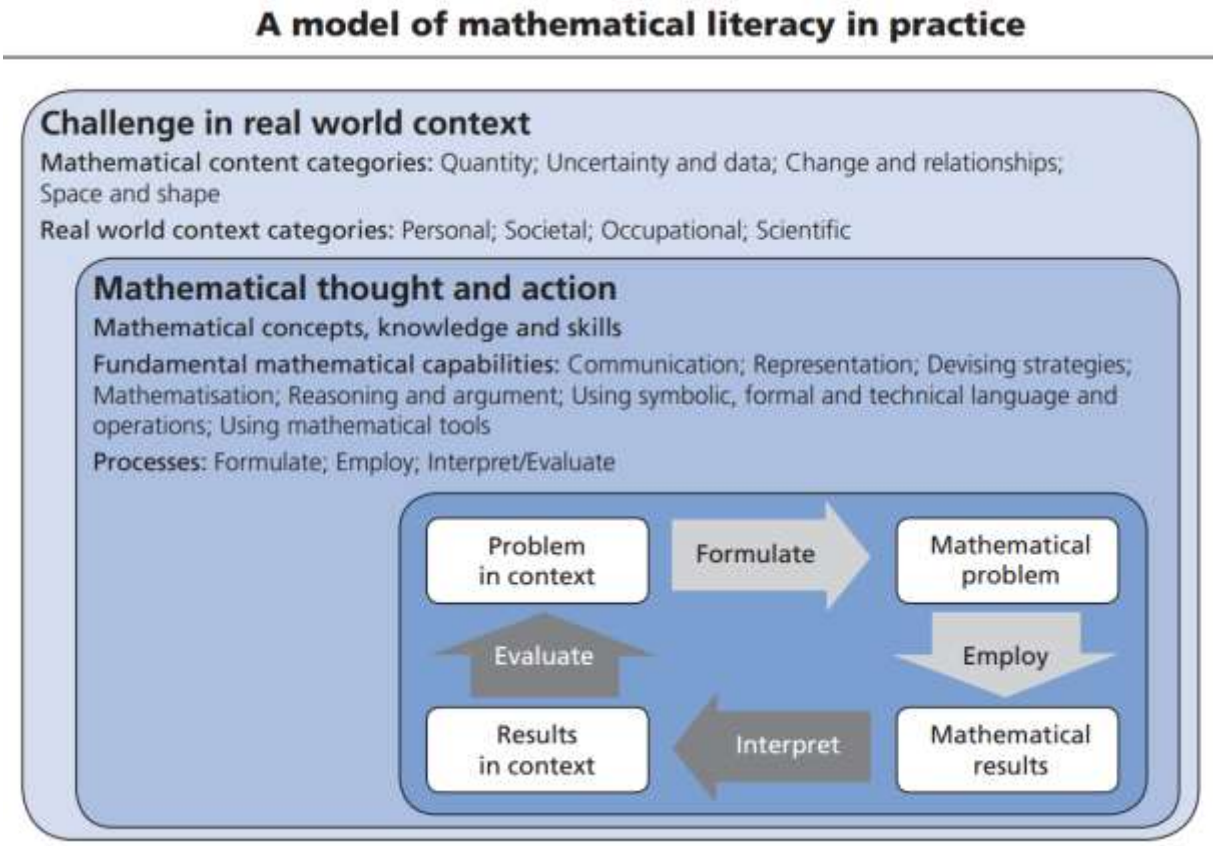


Fig. 5.1 A model of mathematical literacy in practice, according to: OECD (2013a) (p.26).

The most widely reported difficulty is that students do not take into account common-sense considerations about the problem (Greer, 1997; Verschaffel et al., 1994), which affects the processes described in both horizontal steps of the diagram, which are formulating the mathematical problem and interpreting the mathematical results. In many analyses of the difficulties encountered it is found that students look at these problems with a strong "answer-getting mindset" (Daro, 2013) and that they use calculational approaches (Thompson et al., 1994), as if the problem was only limited to the right-hand vertical step of the problem-solving process, which is employing operations. This answer-getting mindset is arguably a result of the mindsets of both students and teachers (Depaepe et al., 2010). There are persistent socio-mathematical norms (Gravemeijer, 1997; Yackel & Cobb, 1995) in many mathematics classrooms implying that solving problems of any kind means "getting the right answer" by conducting a series of operations on the numbers in the problem, and subsequently getting the "right answer".

In order to counteract these tendencies and the associated difficulties, we designed tasks that were more authentic by changing the representation of the problem situation from descriptive to mainly depictive. We expected that the use of images from real life would strengthen the association with real-world situations (Palm, 2009) and therefore decrease the suspension of sense-making (Schoenfeld, 1992) and the strong calculational focus (Thompson et al., 1994). The studies of Dewolf et al. (2014), Dewolf et al. (2015), Elia et al. (2007), and Elia and Philippou (2004) were of special relevance to our design, because they reported on the effect of decorative, representational, and informational pictures in contextual mathematical problems and found mixed effects. These studies encouraged us in using authentic images closely connected to the "real" situation that was to be represented.

The designed tasks were incorporated in an instrument to measure students' performance on both word problems and image-rich numeracy problems in a randomised controlled way. And although the prediction was confirmed in an earlier study (see Chapter 4), the results were not straightforward.

In this study we pursued a further analysis of these findings to shed more light on the intricate relation between task design and student performance in the realm of contextual mathematical problems. The trial with the aforementioned instrument delivered data on students' performance as well as on the relation of students' performance with the variable version (of the problem). More specifically, with these data the relation can be analysed between the independent variables and the difference in students' performance on both versions of the paired problems.

In this study we discuss two distinct types of task characteristics, which might have an effect. These task characteristics are: (1) the complexity of the task, and (2) the content domains of the tasks. We elaborate on this in the next section.

5.2.1 The complexity of the tasks

Several different perspectives and frameworks can be used to look at task complexity in mathematics education (Watson & Ohtani, 2015; Williams & Clarke, 1997) and in general educational settings (Gill & Hicks, 2006; Robinson, 2001). In this study we defined complexity of the word problem versions as the number of words used to represent the problem situation. This so-called "wordiness" is often used in research on task design, for example, by Piel and Schuchart (2014), albeit anchored in general research on the effect of text comprehension on students' performance. This research claims that wordiness often has a negative effect on students' performance (Boonen, 2015; Boonen et al., 2014; Fuchs, Fuchs, Compton, Hamlett, & Wang, 2015). We argued, likewise, that high levels of wordiness would predict a negative impact on students' performance, which would produce a positive effect on the difference in outcomes when changing the representation of the problem situation from descriptive to mainly depictive. In line with the definition of wordiness, we defined complexity of the image-rich numeracy problems as the number of visual elements used. We took an approach that is commonly used in research on cognitive load theory on element interactivity (Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014; Paas, Renkl, & Sweller, 2003). In this research a high number of visual elements is assumed to have a negative impact on students' performance. We argued likewise, and predicted that a high number of visual elements would limit the effect of changing the representation of the problem situation from descriptive to mainly depictive. In this particular analysis we took wordiness for the A-version tasks and the number of visual elements for the B-version of the tasks as independent variables.

5.2.2 The content domains of the tasks

We further assumed that the difference in students' performance on the two versions of the paired problems may work out differently on tasks in different content domains. In the Netherlands in 2010 a new numeracy framework (Hoogland & Stelwagen, 2012) was introduced as part of the "Referentiekader Taal en Rekenen" (Literacy and Numeracy Framework: LaNF). The content domains in this framework resemble the categories used in the international frameworks on numeracy and mathematical literacy, such as TIMSS, PISA and PIAAC (Mullis & Martin, 2013; OECD, 2013c; PIAAC Numeracy Expert Group, 2009). In Table 5.1 we show an overview of the content domains used in the various frameworks. In the Dutch framework there is an emphasis on proportions, fractions, and percentages, and topic areas such as uncertainty, chance and data (representation) are backgrounded.

Assuming the diagram of problem solving in Figure 5.1 contains essential steps for the solving process (going from the problem situation to the situation model and on to the mathematical model), we argued that the mental activity needed for the necessary steps in the process is interdependent on the mathematics domain of the task. In the domain of numbers the mathematical model is primarily computational and thus one dimensional. In that case a mainly depictive representation was presumed not to contribute considerably to the ease with which problem solvers make the situational or mathematical model.

However, in the domain of proportions the mathematical model is in general more complex than in the domain of numbers because there is always some activity of (relatively) comparing quantities or comparing a quantity to a whole. A mainly depictive representation was assumed to be beneficial here. At the same time a counter-effect is possible if the mental model and the depictive representations are not mutually beneficial. In that case, an addition of complexities is likely to be experienced by the students. Hence, for tasks from the domain of proportions one could not make a plausible straightforward prediction, whether a mainly depictive representation could help the solvers to construct the appropriate mental model and hence help them in solving the problem in a successful way.

Table 5.1
Overview of international frameworks on numeracy and their categories

Framework	Categories				
TIMSS 2015 – 8th grade	Number	Algebra	Geometry		Data & Chance
PIAAC 2016	Quantity & Number		Dimension & Shape	Pattern, Relationship & Change	Data & Chance
PISA 2015	Quantity		Space & Shape	Change & Relationships	Uncertainty & Data
Dutch LaNF 2010	Numbers	Proportions	Measurement & Geometry	Relations	

Note. Presented by similarity (horizontal).

In the domain of measurement & geometry the underlying problem situation is often in itself two- or three-dimensional. Hence, a mainly depictive representation of the problem was assumed to help the problem solver to create the appropriate mental and mathematical model.

5.3 Research method

The instrument used to collect the data was a web-based numeracy test (see Appendix B) very similar in content and layout to the nationwide examination (Cito, 2015). The items were diagnostic for level 2F of the Dutch LaNF. The 21 items, which dealt with the content domains of numbers, proportions, and measurement & geometry, were randomly presented in one of two versions: as word problem (A-version) or as image-rich numeracy problem (B-version). Table 5.2 provides an overview of the tasks used in the trial (see also Appendix C), supplemented with the quantification of the characteristics which we added as variables for this study. This instrument made it possible to measure the effect of changing the representation of the problem situation on students' performance in a randomised controlled way. Figure 5.2 shows two examples of paired problems used in the instrument, translated to English for readability. The original instrument and the English translation can be found in Appendices B and C. In our analysis we took other data of the participants into account, such as gender, ethnicity, school level, and mathematics level. By adding other characteristics of the used problems, such as complexity and cognitive/content domains, we were able to shed light on the interdependences of, for instance, problem complexity and effect of changing the representation of the problem situation.

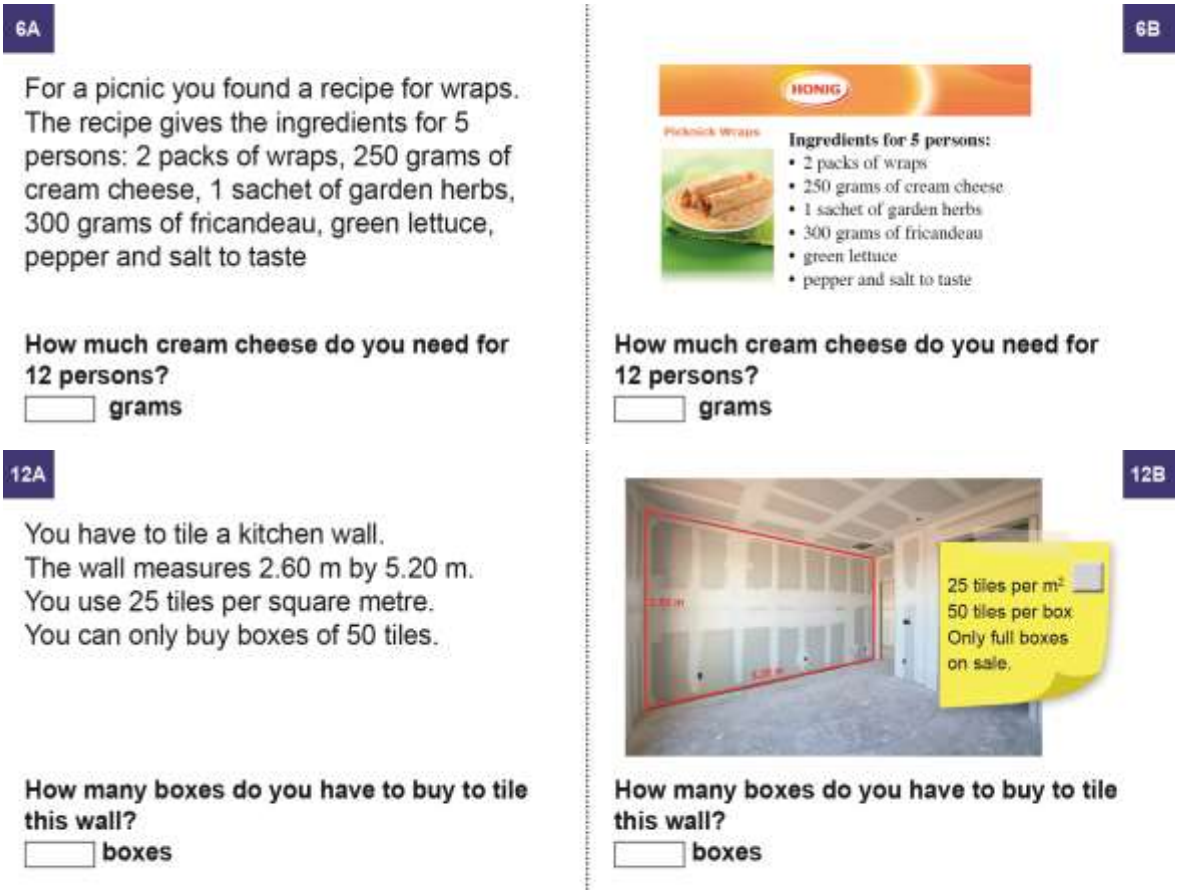


Fig. 5.2 Two examples of paired problems items 6 and 12.

In October and November 2011, the original test was made available through internet for all schools nationwide to use as a diagnostic test. In total, 31,842 students from 179 schools geographically spread across the Netherlands participated in the test. In terms of the total student population in the Netherlands (CBS, 2012) around 2% participated. The students came from grade 7 and 8 of primary education (969 participants), from general secondary education (29,067 participants), and from vocational education (1,556 participants) in an age range of 11–18 years. The distribution of participants over gender and ethnicity in the various school types were close to the national percentages (CBS, 2012). Therefore, we assumed the participants to be representative of Dutch students.

5.4 Statistical analysis

The statistical analysis on the separate items focused on the differences in scores on the A-version and the B-version of the 21 paired problems. We took two approaches. Firstly, we conducted a straightforward classical analysis using mean, standard deviation, *t*-tests, and Cohen's *d* as effect size to get a general idea of how the separate items contributed to the overall result we found. Secondly, we used a probit model which gave the opportunity to investigate in more depth how the difference between the scores on the A- and B-versions were interdependent with the variables we measured from the participants (gender, ethnicity, school level, grade level) or constructed for this study (wordiness and number of pictures, and content domain of task).

For analysing the overall results, in an earlier study a probit model was composed (see Chapter 4). A probit model is a limited dependent variable model which allows for a multivariate analysis (Long, 1997). The results on the items were scored with wrong/right (0/1) and the independent variables, measured or not, were assumed to have a normal distribution. Under these assumptions a probit model is a suitable model to conduct this analysis and present the results. In this approach each matched pair of items can be seen as a problem with 31,842 observations. The representation (A-version or B-version) can be regarded as the manipulated variable *v* which could have an effect on the participants' results. The question was then if the independent variable *v* (version of problem) was a significant factor in interpreting the score on the problem.

In summary the probit model built up as follows:

Ability $y = \alpha_0 + \alpha_1 v + \alpha_2 x + \varepsilon$ can cross a threshold δ which leads to answering an item correctly ($z = 1$). We can then formulate the probabilities:

$$P(z = 1) = P(y \geq \delta\ ) = P(\varepsilon \geq \delta - \alpha_0 + \alpha_1 v + \alpha_2 x\ ) \text{ and}$$

$$P(z = 0) = 1 - P(y \geq \delta\ ) = 1 - P(\varepsilon \geq \delta - \alpha_0 + \alpha_1 v + \alpha_2 x).$$

The unknown coefficients $\delta - \alpha_0$, α_1 and α_2 were estimated by maximum likelihood. In all our analyses these coefficients are calculated using STATA11 (probit, dprobit). The background variables we used were school level ($\alpha_{2,1}$–$\alpha_{2,7}$), grade ($\alpha_{2,8}$), gender ($\alpha_{2,9}$), ethnicity($\alpha_{2,10}$), age deviation ($\alpha_{2,11}$), and maths grade last received ($\alpha_{2,12}$). The task related variables wordiness

($\alpha_{2,13}$), number of pictures ($\alpha_{2,14}$), and content domain($\alpha_{2,15}$–$\alpha_{2,17}$), we used only in this study (see Table 5.2). All these variables, as non-manipulated independent factors, might contribute to the measured results.

Table 5.2
Item characteristics of 21 paired problems

Item	Theme	Content domain	Wordiness	Pictorial complexity
i01	Apples in bag	meas. & geom.	23	2
i02	Fuel usage	meas. & geom.	30	2
i03	Travel time	proportions	29	2
i04	TV + DVD	numbers	26	2
i05	Change	numbers	19	2
i06	Recipe	proportions	44	1
i07	Price magazine	proportions	36	2
i08	AEX index	proportions	30	3
i09	Money pile	numbers	36	2
i10	Scale model	proportions	23	2
i11	Double glazing	meas. & geom.	27	2
i12	Kitchen tiles	numbers	35	2
i13	Water bottles	meas. & geom.	19	2
i14	Bedroom tiles	meas. & geom.	38	3
i15	Endive	proportions	9	1
i16	Hamburgers	numbers	21	2
i17	Cough syrup	numbers	36	3
i18	Public debt	numbers	18	2
i19	Cake tin	meas. & geom.	17	1
i20	Winter tires	proportions	40	3
i21	Chocolate boxes	meas. & geom.	51	4

Note. Wordiness is number of words used in the description of the problem situation (A-version); Pictorial complexity is the number of visual elements in the depictive representation (B-version). meas. & geom. stands for measurement & geometry.

By this design, the characteristics of the participants answering the A-version of a particular item and the characteristics of the participants answering the B-version of that item had the largest likelihood of being the same. This should hold for the measured characteristics as well as for the characteristics that were not measured. The independent manipulated variable was the version of the problem (A or B). The dependent variable was the students' scores on the A-versions and the B-versions of the problems. More details of the probit model and the results of the overall conclusions were described extensively in Chapters 3 and 4.

5.5 Results

In earlier reports we showed that the overall effect was a significantly higher score of about two percentage points on image-rich numeracy problems, albeit with a very small effect size of Cohen's $d = 0.09$. For this study, first, we zoomed in for each paired problem on the results of the t-test on the difference in score on both versions. Second, we interpreted the results of the probit model to show the effects of the other not-manipulated independent relevant variables. The probit model also provided the opportunity to investigate whether additional variables related to item characteristics could shed further light on the effect that these variables had on the difference in students' performance on both versions of the paired problems.

5.5.1. The results on separate paired problems

We analysed the data by comparing the average scores of the participants on each paired problem, relating the average score on the A-version of a problem with the average score on the B-version of the problem. A two-sided t-test with pooled variances was conducted (see Table 5.3) to evaluate whether for each item the scores on the two versions differed significantly. We used a common effect size index, namely Cohen's d, for a first general conclusion.

We found in three paired problems (i5, i11, i21) that the scores on the B-version were significantly higher than the scores on the A-versions, with small effect sizes. Moreover, we found seven paired problems (i7, i10, i13, i16, i18, i19, i20) where the scores on the B-version were significantly higher than the scores on the A-versions, with less than small effect sizes. Furthermore, we found in four paired problems (i2, i3, i12, i17) that the scores on the A-versions were significantly higher than the scores on the B-versions, with less than small effect sizes (Cohen, 1988).

Table 5.3
Testing the difference in students' performance on item level

Item	***N***		**Mean (SD)**		***t*-test**	**effect size *d***	
	version A	version B	version A	version B	*p*(|*T*|>|*t*|)	B > A	A > B
i1	15,878	15,964	.716(.004)	.720(.004)	.424		
i2	15,986	15,856	.525(.004)	.483(.004)	.000 ***		0.08
i3	15,785	16,057	.314(.004)	.290(.004)	.000 ***		0.05
i4	15,835	16,007	.826(.003)	.833(.003)	.131		
i5	16,038	15,804	.720(.004)	.828(.003)	.000 ***	0.26	
i6	15,775	16,067	.631(.004)	.640(.004)	.102		
i7	16,065	15,777	.404(.004)	.416(.004)	.042 *	0.02	
i8	16,298	15,544	.303(.004)	.299(.004)	.420		
i9	16,069	15,773	.221(.003)	.213(.003)	.085		
i10	15,882	15,960	.495(.004)	.525(.004)	.000 ***	0.06	
i11	15,850	15,992	.145(.003)	.310(.004)	.000 ***	0.39	
i12	15,871	15,971	.466(.004)	.438(.004)	.000 ***		0.06
i13	15,931	15,911	.619(.004)	.641(.004)	.000 ***	0.05	
i14	15,889	15,953	.040(.002)	.046(.002)	.080		
i15	15,793	16,049	.394(.004)	.388(.004)	.264		
i16	15,921	15,921	.803(.003)	.815(.003)	.005 **	0.03	
i17	15,986	15,856	.803(.003)	.787(.003)	.000 ***		0.04
i18	15,847	15,995	.153(.003)	.168(.003)	.000 ***	0.04	
i19	15,932	15,910	.247(.003)	.284(.004)	.000 ***	0.08	
i20	15,925	15,917	.130(.003)	.164(.003)	.000 ***	0.10	
i21	16,044	15,798	.188(.003)	.256(.003)	.000 ***	0.16	
All	334,600	334,082	.436(.001)	.455(.001)	.000***	0.09	

Note. *N* is number of items tested. M is mean score on items (with standard deviation in parentheses) $P(|T|>|t|)$ is result of *t*-test, unpaired, unequal with hypothesis that difference in score is = 0; *p< .05, **p< .01, ***p< .001. Cohen's *d* is effect size. Version A is the word problem; Version B is the image rich numeracy problem.

These results were in line with the small effect size we found in the overall results. In some items the effect of higher scores on the B-version occurred with a small or less than small effect size, in other items it did not. In a few items the effect was even the opposite. In the remainder of the section we analysed whether this can be attributed to specific characteristics of the items.

5.5.2. The results of the probit model

The results of the basic probit model with variables α_1 and $\alpha_{2,1}$–$\alpha_{2,12}$ were reported earlier (see Chapter 4). The overall conclusion of the generic increase of two percentage points as a result of changing the representation of the problem situation from descriptive to depictive can also be seen in Table 5.4 as the marginal effect of the variable version. In a probit model the marginal effect can be considered as a measure of effect size, but this must be done with great prudence (Hoetker, 2007).

5.5.3 The probit model expanded with variables wordiness and number of pictorial elements.

Two variables were used to indicate the complexity of the problems: wordiness and number of pictorial elements. Table 5.4 shows per item the values of these variables. We expanded the probit model with these data and analysed the effect of these variables on the students' performance and the interdependent effect of these data with the variable version.

Table 5.4 shows that the wordiness of the word problem had a significant relation to the students' performance, and that the effect was negative, as was expected: higher wordiness led to lower performance on the A-versions of the problem. From Table 5.4 we also concluded that the number of pictorial elements was significant for the students' performance and that the effect was negative: a higher number of pictorial elements led to lower performance on the B-versions of the problem. This indicated also that the number of pictorial elements seems to be a valid measure of complexity of image-rich numeracy problems. From the marginal effects we can get an indication of the effect, for each paired set of tasks, of changing from a verbal description to a pictorial description. In item 1, for instance, 23 words were used to describe the problem in the A-version of the item, and two pictorial elements were used in the B-version of the problem. According to the model, changing the representation would give an increase of the chance of students' success of $23 \times -0.0678 - 2 \times -0.0923 = 0.028$, about three percentage points. We can get such an indication for all the items and see an added overall effect of 0.038, almost four percentage points. The list of such indications on each item had a positive but only weak correlation of r = .17 with the measured difference in students' performance on the A-version and B-version for each item, hence this indication must be used with prudence.

Table 5.4
Probit model expanded with item characteristics: wordiness and number of depictive elements

Variable		**Coefficient**	**Marginal effect**
Version (A = 0, B = 1)	α_1	0.050***(0.011)	0.020***(0.004)
Other variables	$\alpha_{2,1}$–$\alpha_{2,12}$	[a]) ***	[a]) ***
Gender × Version	cross term	0.018**(0.007)	0.007**(0.003)
Number of words | version A	$\alpha_{2,14}$	-0.017***(0.000)	-0.007***(0.000)
Number of pictorial elements | version B	$\alpha_{2,15}$	-0.234***(0.003)	-0.092***(0.001)
Constant	α_0*	0.538***(0.013)	
Pseudo-R^2		0.087	

Note. N = 605,421, $\alpha_{i,j}$ is the coefficient in the probit model, α_0*= $-\delta + \alpha_0$. Coefficients are calculated with maximum likelihood by STATA11, standard errors are in parentheses. The measure of good fit pseudo-R^2 is McFadden's R^2. All variables are significant *p< .05, **p< .01, ***p< .001. The effect of Number of words is only analysed for the word problems (version A); The effect of Number of pictorial elements is only analysed for the image-rich numeracy problems (version B). [a]) Other variables $\alpha_{2,1}$–$\alpha_{2,12}$ are left out for easier readability.

5.5.4 The probit model expanded with the variable content domain.

In Table 5.5 the results can be found of expanding the probit model with the variables regarding the content domains of the tasks. The used content domains from the Dutch LaNF – numbers, proportions, and measurement & geometry – were represented with seven items each (see Table 5.2).

In Table 5.5 in the probit analysis the domain of numbers was used as reference category for the effect of the domain categories. The marginal effects regarding the domain of proportions and measurement & geometry were negative compared with the domain of numbers.

Table 5.5
Probit model expanded with variables regarding content domain

Variable		Coefficient	Marginal effect
Version (A = 0, B = 1)	α_1	0.030***(0.007)	0.012***(0.003)
Other variables	$\alpha_{2,1}$–$\alpha_{2,12}$	[a]) ***	[a]) ***
Gender × Version	cross term	0.019**(0.007)	0.008**(0.003)
Domain numbers	$\alpha_{2,16}$	ref. cat.	ref. cat.
Domain proportions	$\alpha_{2,17}$	-0.522***(0.006)	-0.201***(0.002)
Domain meas. & geom.	$\alpha_{2,18}$	-0.596***(0.006)	-0.228***(0.002)
Domain proportions × Version	cross term	-0.016(0.008)	-0.006(0.003)
Domain meas.& geom.× Version	cross term	0.062***(0.008)	0.024***(0.003)
Constant	α_0*	-0.947*** (0.189)	
Pseudo-R^2		0.111	

Note. *N* = 605,421, $\alpha_{i,j}$ is the coefficient in the probit model, $\alpha_0^* = -\delta + \alpha_0$. Coefficients are calculated with maximum likelihood by STATA11, standard errors are in parentheses. Ref. cat. is reference category The measure of good fit pseudo-R^2 is McFadden's R^2. All variables are significant *p< .05, **p< .01,***p< .001, with the exception of proportions × version. [a]) Other variables $\alpha_{2,1}$–$\alpha_{2,12}$ are left out for easier readability.

This meant that in this instrument the tasks from the domains of proportions and measurement & geometry were harder than the tasks from the domain of numbers for the participating students. For the aim of this study, however, we looked at the interdependent effects of the variables regarding the content domains and the variable version, and we found a significant and negative effect of the cross term proportion × version, and a significant and positive effect of the cross term measurement & geometry × version. This meant that the effect of changing the representation of the problem situation had the least effect in the domain of proportions, and the most effect in the domain of measurement & geometry. The effect in the domain of numbers was in between these effects, but closer to the effect of proportions. The effect on tasks in the domain of measurement & geometry was in line with our expectations: the depictive representation in these kinds of tasks was likely to be beneficial in the problem-solving process. The result of the tasks in the domain of proportions led us to believe that in these tasks the mental model and the depictive representations were not mutually beneficial, which might have led to an increasing complexity.

5.6 Discussion

This study was part of a larger project to research alternatives to word problems in representing the problem situation in mathematical contextual problems. In earlier analyses of the data collected with the described instrument we found that changing the representation of the problem situation from descriptive to mainly depictive had an overall positive effect on students' performance that was small but robust. In our design, if the difference was positive, students performed better on image-rich problems than on otherwise equivalent word problems. By robust we indicated that the measured effect was not interdependent with background variables like age of students, ability of students, ethnicity and (type of) school, although there was a small interdependency with gender.

In the overall findings we were able to discern some patterns in the differences in performance we found. These patterns revealed that task characteristics could have an interdependent effect on the effect of changing the representation of the problem situation. And those task characteristics could also give indications of possible explanations of the overall results. We focused on specific task characteristics and the interdependence of those task characteristics with the effect of changing the representation of the problem situation, indicated by a difference in students' performance on word problems and image-rich numeracy problems. From the analysis reported in this study we found that the effect of changing representation of the problem situation on the students' performance is interdependent with the following item characteristics:

- the number of words used in the descriptive representation and the number of pictorial elements used in the depictive representation;
- the content domain of the task.

To elaborate on each of these characteristics, the literature on use of language in contextual mathematical problems indicated that wordiness of the representation of the problem could have a negative effect on the performance of students, and therefore a positive effect on changing the representation of the problem situation from descriptive to depictive (Boonen et al., 2014; Fuchs et al., 2015). This indication was corroborated by this study. Likewise, the literature on the use of pictorial elements in classroom problems indicated that the number of pictorial elements could have a negative effect on the performance of students (Leppink et al., 2014). This indication was also corroborated by this study. However, the effect of changing from words to pictures in one trial was under-researched, with the exceptions of work by Lowrie and colleagues (Lowrie, Diezmann, & Logan, 2012; Lowrie & Diezmann, 2009). Our study provided indications of the magnitude of the effect related to the number of words and the number of pictorial elements. If a description of, for instance, 25 words is replaced with one pictorial element, then the model analysis in this study indicated an increase in the chance of answering the task correctly by around eight percentage points. We are far from suggesting that this is a general result that will occur in all comparable tasks and task domains (see also next paragraph), but with the magnitude of the sample used in this study it is an effect that deserves further investigation.

Regarding the domain of the task, we found hardly any research on the interdependent effect of the domain of the task and the representation of the problem situation. One of the findings of our study was that the effect of changing the representation of the problem situation from descriptive to depictive was least strong with tasks from the domain of proportions, slightly stronger for the domain of numbers, and clearly stronger for the domain of measurement & geometry. From our experience in mathematics education we would anticipate that the effect would be the largest in the domain of measurement & geometry. The most likely explanation is that in geometrical or measurement contexts pictorial elements support students in making a relevant model of the situation, and as a consequence support them in formulating the mathematical problem correctly, which was one of the crucial steps in the problem-solving cycle we described in the theoretical and empirical background section. In other words, as exemplified in Figure 5.3, in the domain of measurement & geometry the information can “be read off more directly from the representation” (Schnotz et al., 2010). For tasks in the domain of measurement & geometry word problems can in many cases be considered as quite an inadequate representation of the problem situation: the designer of the tasks describes a geometrical situation in words, which must be “undescribed” and interpreted by the students to make a geometrical mental model and formulate the mathematical problem. Especially in cases where many words are necessary to describe the geometrical situation, the student is likely to be seriously hampered in demonstrating his or her geometrical problem-solving knowledge and skills. Figure 5.3 shows an example of an item with a significantly higher score on the depictive representation of the problem situation.

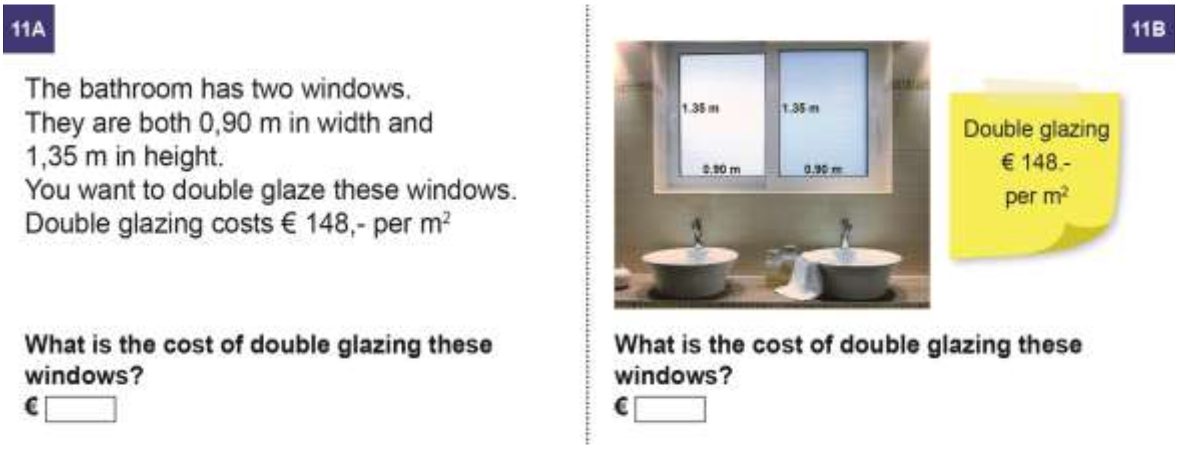


Fig. 5.3 Paired problem item 11: double glazing.

The limitations of this study should be kept in mind, however. First, the variable gender seemed to have, at least in part, an interdependence with the variable version (i.e., representation of the problem situation), in the sense that the performance of male students was enhanced slightly more by the change to image-rich numeracy problems (see Table 5.4). The effect of gender on mathematics performance has been researched extensively (Lindberg et al., 2010). However, this gender effect regarding the representation of mathematical problems needs more systematic research. Second, nowadays students are trained in word problems and are familiar with them. The effect of this on their performance on rather "new" image-rich numeracy problems is still under-researched. It needs more investigation to determine the precise effect of complexity and representation of the problem on performance and the problem-solving mindset of the students. Third, the variables wordiness and number of pictures are rather crude measures of problem complexity. Another approach could be to take the empirical complexity as an independent variable in the analysis of the effect of the task characteristics. This a posteriori way of analysis was outside the design of the original controlled randomised study, however.

5.7 Conclusion

In educational settings where contextual mathematical problems are used, the effect on students' performance of changing the verbal representation of the problem situation to a mainly depictive representation of the problem situation is still under-researched. This study is one of very few that have investigated the effect on students' performance of changing the representation of the problem situation in contextual mathematical problems in a randomised controlled trial. The results can be of importance to mathematical task designers, especially in cases where assessments are high stakes for students, or where achievement results are being used for comparing countries, like TIMSS, PISA, and PIAAC (Mullis & Martin, 2013; OECD, 2012; PIAAC Numeracy Expert Group, 2009; Tout & Gal, 2015). In many countries large-scale assessments have had, and often still have, a large influence on decision making concerning educational policies and curriculum development. In these cases, an increase of a few percentage points in results is significant for decision makers.

In this study we did not expect to find an overall explanation for the measured effects on performance of students of changing the representation of reality in contextual mathematical problems. There most probably are more factors than task characteristics – like wordiness and domain of the task – that could offer possible explanations of the measured results. More quantitative and qualitative research is needed to gain a better understanding of the relation between problem representation and students' behaviour, for instance research that investigates the ways students are actually reasoning during the problem-solving process.

Furthermore, we see broader developments in education, media, and communication in which visual representations of reality become increasingly common in daily communications: YouTube, Instagram, and so forth. This makes the search for alternatives in the direction of depictive representation of problem situations even more relevant. An increasing number of topics on websites and in newspapers are explained, illustrated, or illuminated by depictive messages with quantitative aspects. The use of depictive representations of reality in contemporary classrooms is lagging behind the students' experiences outside school. This is a serious point for concern. Too big a gap between reality and the representation of that reality in classroom situations can threaten the motivation of students by diminishing the perceived relevance of the task (Seegers & Boekaerts, 1993) and their sense of personal relatedness (De Brabander & Martens, 2014).

We suggest that students' difficulties with word problems, in particular in assessments, have to be taken seriously, and that further research should be encouraged to shed more light on the relations between task characteristics and students' performance in dealing with mathematical contextual problems.

Chapter 6. A replication study using the instrument with adult participants

Abstract

Word problems are often used to assess numeracy, despite the growing number of reports on difficulties students encounter with this genre of mathematical problems. These reports contend that a large number of difficulties are influenced by the way the problems are presented, that is, with verbal representations of the problem situations. These difficulties are said to be associated with a form of suspension of sense-making. In this study, conducted in the Netherlands, we investigated the effect on adult participants' performances of changing the representation of the problem situation from verbal to image-rich. A controlled randomised trial was the main part of this investigation. In a recent study a similar trial was held with students from primary and secondary education. The trial is now replicated with adult participants. The study showed that participants' performances on tasks in the content domain of measurement & geometry particularly were improved when changing from word problems to image-rich numeracy problems. These results could be of interest for all practitioners involved in the work of numeracy task design.

This chapter is accepted as a journal article and in press.

Hoogland, K., & Pepin, B. (in press). The intricacies of assessing numeracy: Investigating alternatives to word problems. *Adults Learning Mathematics - An International Journal*, *11*(2).

6.1 Introduction

In most recent approaches in adult numeracy research, adult numeracy is defined as a complex, multifaceted, and sophisticated construct, incorporating the mathematics, communications, cultural, social, emotional and personal aspects of each individual in context (American Institutes for Research, 2006; Geiger et al., 2015). As a consequence, learning and assessing numeracy in authentic situations is often advocated (Frankenstein, 2009).

A closer look at lesson or test materials used in numeracy education in many countries, however, reveals that most assessment materials consist of word problems or of items assessing procedural arithmetic skills. The same is the case in the Netherlands where, despite the country's high rankings in international comparisons, there are persistent complaints about the literacy and numeracy levels of young adults in vocational education and in the workplace. As a result, in 2010 a Literacy and Numeracy Framework (LaNF) was introduced in the Netherlands, with a compulsory numeracy examination at the end of the vocational educational tracks (Hoogland & Stelwagen, 2012). After a lively debate on the assumed value of procedural skills for (young) adult learners, a compulsory numeracy examination has been implemented which consists of 45 mathematical problems of which 15 are strictly procedural problems and 30 are contextual problems. Many teachers and mathematics educators have questioned the relevance of assessing vocational students this way, and they perceive a gap between the numeracy used by their young adult students in everyday life and (future) work, and numeracy as operationalised in the final examinations (see also Chapter 2).

A study in 2011 and 2012 in the Netherlands focused on the idea that using image-rich numeracy problems contributes to bridging the gap between common classroom practice in numeracy and more sophisticated numeracy concepts (see Chapter 4). Part of this study was a controlled randomised trial with almost 32,000 primary, secondary, and vocational students to investigate the effect on students' performance of changing the representation of the problem situation from verbal (word problem) to image-rich (mixture of picture and words). The trial revealed that students performed better on image-rich numeracy problems than on otherwise equivalent word problems (see Chapter 4), indicating that students are less hampered by the many difficulties with word problems that are frequently reported (Verschaffel et al., 2000; Verschaffel et al., 2009).

In an experiment in 2013 in the Netherlands the results of this trial were replicated with adult participants. It revealed which types of tasks particularly benefitted from the change from a verbal description of the problem situation to a mainly depictive description of the problem situation.

6.2 Theoretical background

This study is part of a larger research project to investigate alternatives to the persistent and problematic use of word problems to teach and assess students' ability to deal with numerical problems originating in everyday life. This ability of students is often labelled as numeracy or mathematical literacy, although these concepts have been and are still evolving (Geiger et al., 2015; Ruthven, 2016). The sometimes superficial use of the concept is also criticised (Jablonka, 2015).

In current classroom practice, word problems are used predominantly to teach and assess these abilities. Many researchers, however, report serious difficulties in using word problems to assess these abilities (Verschaffel et al., 2000; Verschaffel et al., 2009). The reported difficulties can be related to the steps the problem solver is expected to take to solve the task at hand.

Figure 6.1 shows the diagram used in PISA as a schema for the relevant steps in the problem-solving process. Similar diagrams (see Chapter 1) are used in related research on problem solving and modelling in mathematics education (Blum et al., 2007; Burkhardt, 2006; Lesh et al., 1987).

The reported difficulties seem to appear mainly in the two horizontal steps in the diagram: "formulate the mathematical problem" and "interpret the mathematical results". In the first step (formulate) students are reported to look at these problems with a strong "answer-getting mindset" (Daro, 2013) and a calculational approach (Thompson et al., 1994), as if the problem was limited to the right-hand vertical step of the problem-solving process and that solving problems of any kind means getting the "right answer" by conducting a series of operations on the numbers in the problems. In the third step (interpret) students are reported not to take common-sense considerations about the problem into account (Greer, 1997).

In order to counteract these tendencies and the associated difficulties, in an earlier study we designed tasks that were more "authentic" by changing the representation of the problem situation from descriptive to mainly depictive. We conjectured that the use of images from real life would strengthen the association with real-world situations (Palm, 2009) and therefore decrease the suspension of sense-making (Schoenfeld, 1992) and the strong calculational focus (Thompson et al., 1994). As a paraphrase of the most used definition of word problems (Verschaffel et al., 2014), we suggested the following definition for such image-rich problems: "Image-rich numeracy problems can be defined as visual representations of a problem situation wherein one or more questions are raised, the answer to which can be obtained by the application of mathematical reasoning to numerical data available in the problem representation".

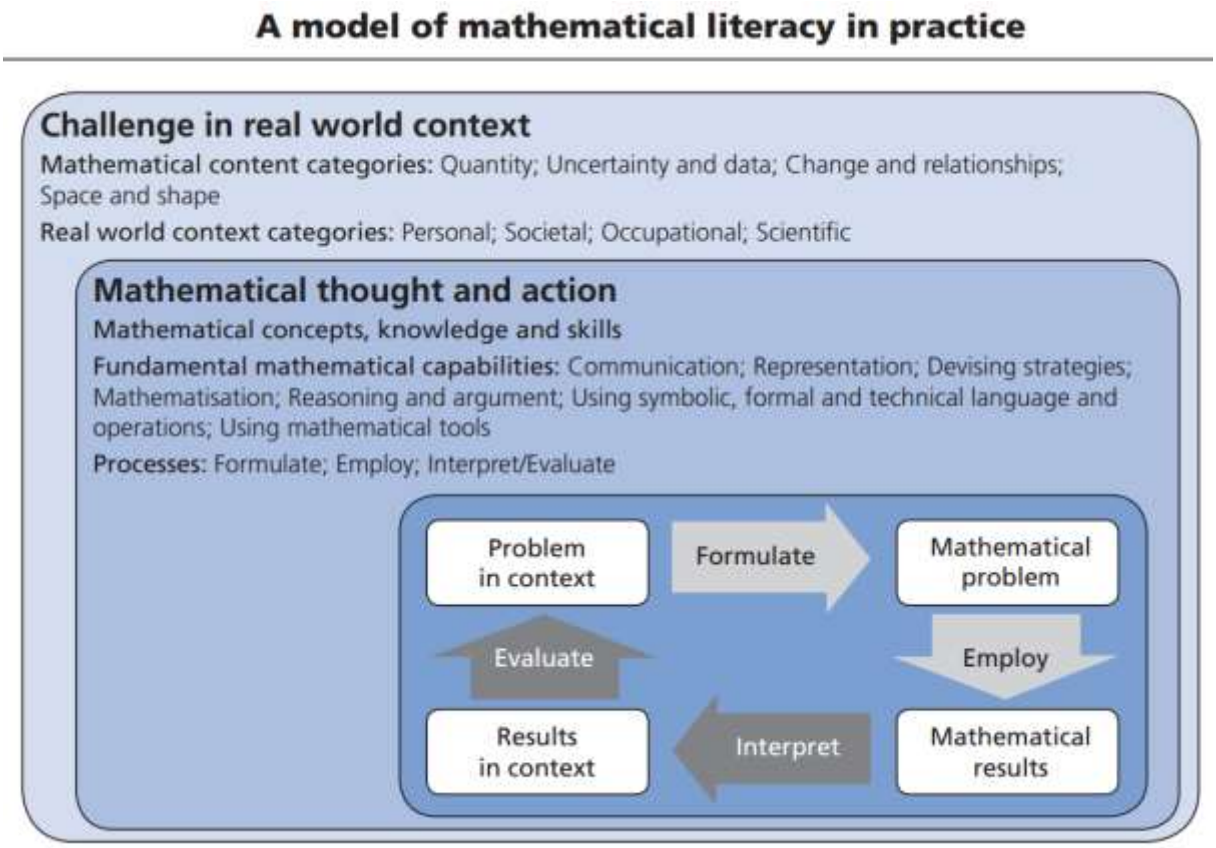


Fig. 6.1 A model of mathematical literacy in practice, according to: OECD (2013a) (p.26).

Those tasks were incorporated in an instrument to measure students' performance on both word problems and image-rich problems in a randomised controlled way. In an earlier study we reported the overall result of using the instrument (see Chapters 3 and 4). Although the prediction was confirmed, the results were not straightforward. The students' scores on image-rich problems were slightly higher (2%), which was significant, but with a small effect size (Cohen's d = .09) and the effect of better performance was most noticeable in tasks in the domain of measurement & geometry (see Chapter 5).

The research question for the study reported here is: Does a replication of the original trial with adult participants show the same patterns and results as the original trial with primary and secondary students? In this paper we report on that replication of the original study with adults who participated in the "Groot Nederlands Rekenonderzoek (GNRO)" [Great National Numeracy Survey], a research initiative by the public broadcasting organisations VPRO and NTR, supported by the Netherlands Organisation for Scientific Research (NWO). Individuals of all ages and of all places in the Netherlands could register as participants on the GNRO website and could engage in a series of mathematical tests.

6.2.1 The Dutch context

For the international reader, we provide some information on the Dutch educational context. In the Netherlands in 2010 the "Referentiekader Taal en Rekenen" [Literacy and Numeracy Framework (LaNF)] was introduced as a guideline for Literacy and Numeracy education in the age range of 4–18 years (Hoogland & Stelwagen, 2012; Ministerie van OCW, 2009), followed by a very similar version for adult education.

Table 6.1
Overview of international frameworks on numeracy and their content domains

Framework	**Categories**				
TIMSS 2015 – 8th grade	Number	Algebra	Geometry		Data & Chance
PIAAC 2016	Quantity & Number		Dimension & Shape	Pattern, Relationship & Change	Data & Chance
PISA 2015	Quantity		Space & Shape	Change & Relationships	Uncertainty & Data
Dutch LaNF 2010	Numbers	Proportions	Measurement & Geometry	Relations	

Note. Presented by similarity (horizontal).

The content domains in these frameworks resemble the categories used in the international frameworks on numeracy and mathematical literacy, such as TIMSS, PISA and PIAAC (Mullis & Martin, 2013; OECD, 2013c; PIAAC Numeracy Expert Group, 2009). Table 6.1 gives an overview of the content domains used in the various frameworks. It is noteworthy that in the Dutch framework there is more emphasis on proportions, including fractions and percentages, and an absence of focus on uncertainty, chance and data (representation).

6.3 Method

We will describe the instrument that was used in both the original trial and in this replication study with adult participants. The trials were held with Dutch language items (see Appendix B); English translations of these items are available (see Appendix C) under open access (Hoogland & De Koning, 2013).

6.3.1 The instrument

The instrument consisted of 24 items of which 21 items were designed in two versions: word problem and image-rich problem. For every participant a test was composed randomly with 10 or 11 items in each version. The randomly selected items were presented in random order for each participant. In this case a randomised controlled trial was built into the test. Both versions of each item had an equal chance of being selected, independent of any other variables, measured or not.

Fig. 6.2 An example of an item in two versions: word problem and image-rich problem.

In Figure 6.2 we show an example of two versions of an item. The items are translated into English for better readability. In the test, each item was presented as a screen-filling problem with an open numerical answer field. The tasks in the research instrument were validated and tested in earlier research activities (see Chapter 3). The complete set of tasks can be found under open access via the Dutch institute DANS/NWO (Hoogland & De Koning, 2013).

In Table 6.2 we give an overview of the items in the instrument, evenly distributed across three domains of the LaNF: numbers, proportions, and measurement & geometry. Three tasks in the instrument were in the domain of relations, and were only presented in one version, because of the already visual nature of the items.

Table 6.2
Overview of tasks in the used instrument

item	**Domain numbers**	**item**	**Domain measurement & geometry**	**item**	**Domain proportions**
i04	TV + DVD	i01	Apples in bag	i03	Travel time
i05	Change	i02	Fuel usage	i06	Recipe
i09	Money pile	i11	Double glazing	i07	Price magazine
i12	Kitchen tiles	i13	Water bottles	i08	AEX index
i16	Hamburgers	i14	Bedroom tiles	i10	Scale model
i17	Cough syrup	i19	Cake tin	i15	Endive
i18	Public debt	i21	Chocolate boxes	i20	Winter tires

6.3.2 Participants

The research conducted for this paper was a trial with 420 participants from the GNRO research. Table 6.3 shows the distribution of gender and age categories of these participants. The GNRO was, after registration, an open access public test held in 2013. We cannot consider these participants as a representative sample of Dutch adults. However, we consider the distribution over age and gender diverse enough to draw some tentative conclusions on the results in comparison with the results of the original trial.

Table 6.3
Number of participants from GNRO according to age groups and gender.

Age	*n* (%)	Gender	*n* (%)
10–19	15 (5.3%)	male	115 (40.4%)
20–29	61 (21.4%)	female	170 (59.6%)
30–39	66 (23.2%)	not stated	135
40–49	64 (22.5%)		
50–59	31 (10.9%)		
60–69	39 (13.7%)		
70–79	9 (3.2%)		
not stated	135		

Note. Total sample is $N = 410$. n is number with percentages in parentheses, taken on stated age and gender.

The original trial was conducted in October and November 2011. In that trial 31,842 students from 179 schools geographically spread across the Netherlands, participated. Table 6.4 shows the number of participants from the educational streams in the Dutch school system.

Table 6.4
Number of participants in original trial according to age groups and gender

Age	*n* (%)	Gender	*n* (%)
11–12	969 (3.1%)	male	15,310 (49.7%)
12–19	30,222 (96.9%)	female	15,465 (50.3%)
Not stated	680	not stated	1,067

Note. Total sample is $N = 31{,}842$ participants. Age group 11–12 is primary education, age group 12–19 is secondary education. *n* is number with percentages in parentheses, taken on stated age and gender.

In this original trial we assumed the sample to be representative of Dutch students in the age group 11–19 years.

6.3.3 Statistical analysis

The statistical analysis focused on the difference in scores on the A-version and the B-version of the 21 paired problems. We conducted a classical analysis using mean, standard deviation, *t*-tests, and Cohen's *d* as effect size to get a general idea of how the separate items contributed to the overall result we found (Cohen, 1988).

6.4 Results

We present for both trials the results in the same table format for easier comparison. In this study we focus on the overall test and the results at item level.

6.4.1 Results of the GNRO – adult participants

For the overall result on the test on the data collected in the GNRO we conducted a *t*-test on the mean scores on the A- and B-version items for each participant. We found that the difference in mean was .011 with standard error .001 and $p = .184$ (n.s.).

On item level we conducted a two-sided *t*-test with pooled variances to evaluate whether for each item the scores on the two versions differed significantly. We used a common effect size index, namely Cohen's *d*, for a first general conclusion. The results are shown in Table 6.5. We found in four paired problems that the scores on the B-version were significantly higher than the scores on the A-versions with small to moderate effect sizes. Furthermore, we found in one pair of problems that the scores on the A-version were significantly higher than the scores on the B-version with a small effect size. In the 16 other items the differences between the scores were not significant. The results in this replication trial were in most aspects in line with the results in the earlier large-scale student trial, which is discussed in more detail below.

6.4.2 Results of the large scale school trial

The results of the large scale trial have been published before (see Chapters 4 and 5). In this chapter, Table 6.6 highlights those results that are necessary to make the comparison with the population of this study. For this comparison only we incorporated $p<.10$ as a category – it is not used for further statistical inferences.

Table 6.5
Results of the GNRO trial, mean and *t*-test results

Item	N		Mean (SE)		*t* -test	effect size *d*	
	version A	version B	version A	version B	$p(\|T\|>\|t\|)$	B>A	A>B
i1	218	202	.899(.020)	.872(.024)	.374		
i2	215	205	.823(.026)	.800(.028)	.544		
i3	220	200	.836(.025)	.800(.028)	.337		
i4	205	215	.951(.015)	.916(.019)	.150		
i5	196	224	.898(.022)	.772(.028)	.000 ***		.04
i6	209	211	.895(.021)	.929(.018)	.218		
i7	223	197	.749(.029)	.751(.031)	.955		
i8	207	213	.662(.033)	.690(.032)	.537		
i9	209	211	.593(.034)	.540(.034)	.274		
i10	212	208	.821(.026)	.870(.023)	.162		
i11	203	217	.493(.035)	.774(.028)	.000 ***	.59	
i12	212	208	.811(.027)	.789(.028)	.559		
i13	202	218	.896(.021)	.858(024)	.233		
i14	212	208	.472(.034)	.433(.034)	.423		
i15	226	194	.774(.028)	.825(.027)	.198		
i16	219	201	.872(.022)	.866(.024)	.845		
i17	205	215	.971(.012)	.954(.014)	.355		
i18	194	226	.418(.035)	.540(.033)	.012 **	.24	
i19	203	217	.611(.034)	.691(.031)	.085*	.17	
i20	216	204	.533(.034)	.520(.035)	.794		
i21	220	200	.682(.031)	.755(.030)	.096*	.16	

Note. N is number of items tested. M is mean score on items (with standard error in parentheses), $P\|(\|T\|>\|t\|)$ is result t-test, unpaired, unequal with hypothesis that difference in score is 0; *p< .10, **p< .05,***p< .01. Cohen's *d* is effect size. Version A is the word problem; Version B is the image-rich numeracy problem.

For the overall results on the test on the data collected in the large-scale school trial, we conducted a *t*-test on the mean scores on the A- and B-version items for each participant. We found that the difference in mean was .019 with standard error .001 and p< .001. On item level we conducted a two-sided *t*-test with pooled variances to evaluate whether for each item the scores on the two versions differed significantly. We again used the effect size index, Cohen's *d*, for similar conclusions. The results are shown in Table 6.6.

Table 6.6
Results of the large-scale school trial, mean and t-test results.

Item	***N***		***M* (SE)**		***t* -test**	**effect size *d***	
	version A	version B	version A	version B	*p*(|*T*|>|*t*|)	B>A	A>B
i1	15,878	15,964	.716(.004)	.720(.004)	.424		
i2	15,986	15,856	.525(.004)	.483(.004)	.000***		.08
i3	15,785	16,057	.314(.004)	.290(.004)	.000***		.05
i4	15,835	16,007	.826(.003)	.833(.003)	.131		
i5	16,038	15,804	.720(.004)	.828(.003)	.000***	.26	
i6	15,775	16,067	.631(.004)	.640(.004)	.102		
i7	16,065	15,777	.404(.004)	.416(.004)	.042**	.02	
i8	16,298	15,544	.303(.004)	.299(.004)	.420		
i9	16,069	15,773	.221(.003)	.213(.003)	.085*		.02
i10	15,882	15,960	.495(.004)	.525(.004)	.000***	.06	
i11	15,850	15,992	.145(.003)	.310(.004)	.000***	.39	
i12	15,871	15,971	.466(.004)	.438(.004)	.000***		.06
i13	15,931	15,911	.619(.004)	.641(.004)	.000***	.05	
i14	15,889	15,953	.040(.002)	.046(.002)	.080*	.02	
i15	15,793	16,049	.394(.004)	.388(.004)	.264		
i16	15,921	15,921	.803(.003)	.815(.003)	.005***	.03	
i17	15,986	15,856	.803(.003)	.787(.003)	.000***		.04
i18	15,847	15,995	.153(.003)	.168(.003)	.000***	.04	
i19	15,932	15,910	.247(.003)	.284(.004)	.000***	.08	
i20	15,925	15,917	.130(.003)	.164(.003)	.000***	.10	
i21	16,044	15,798	.188(.003)	.256(.003)	.000***	.16	

Note. *N* is number of items tested. *M* is mean score on items (with standard error in parentheses) $P(|T|>|t|)$ is result of *t*-test, unpaired, unequal with hypothesis that difference in score is 0; *p< .10, **p< .05,***p< .01. Cohen's *d* is effect size. Version A is the word problem; Version B is the image-rich numeracy problem.

In the large-scale school trial we found with $p < .10$ in 11 paired problems that the scores on the B-version were significantly higher than the scores on the A-versions with effect sizes ranging from small to very small. Furthermore, we found in five paired problems that the scores on the A-versions were significantly higher than the scores on the B-versions with a very small effect size.

6.4.3 Comparing results

The overall result on performance in this study with adult participants was 1.1 percentage point higher scores on image-rich problems. This was in line with the overall results we found in the large school trial (1.9 percentage point higher scores on image-rich problems).

11A

The bathroom has two windows.
They are both 0,90 m in width and
1,35 m in height.
You want to double glaze these windows.
Double glazing costs € 148,- per m^2

What is the cost of double glazing these windows?
€ ______

11B

What is the cost of double glazing these windows?
€ ______

19A

A cake tin is 25 centimetres long and 25 centimetres wide. The height is 8 centimetres.

How many litres of cake mix do you need to fill this cake tin half full?
______ **liter**

19B

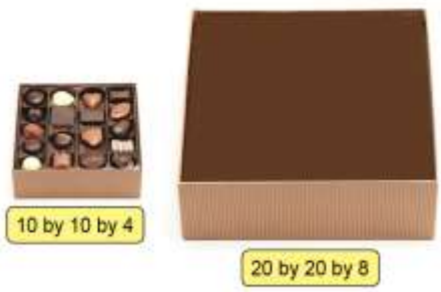
How many litres of cake mix do you need to fill this cake tin half full?
______ **liter**

21A

Afra designs packaging materials.
She made a box for luxury chocolates.
The bottom is a square with sides of 10 cm.
The height is 4 cm.
The manufacturer asks her to design a similar box, but with as a bottom a square with sides of 20 cm and a height of 8 cm.

The volume of the second chocolate box is, in comparison to the first,
______ **times as big.**

21B



The volume of the second chocolate box is, in comparison to the first,
______ **times as big.**

Fig. 6.3 Three examples from the domain of measurement & geometry with a significantly higher score on the image-rich version.

In some items the effect of higher scores on the B-version occurred with a small or less than small effect size, in other items it did not occur. In one item the effect was even opposite.

The remainder of the section presents analysis whether this can be attributed to some characteristics of the items. We found four tasks that in both trials showed a significant better performance for the image-rich versions. Three of them are from the domain of measurement & geometry. They are shown in Figure 6.3. This finding corroborates our earlier findings that the change in representation of the problem situation has the greatest positive influence on the performance of the participants in tasks from the domain of measurement & geometry. We found one item (see Figure 6.4), which gave an opposite effect in each trial. This task is from the domain of numbers, and regards a money question.

Fig. 6.4 One example from the domain of numbers with a opposite result in both trials comparing the two versions.

A possible explanation is that adults are more used to descriptive problems where money is involved than are primary and secondary school students.

6.5 Discussion

Assuming the diagram of problem solving in Figure 6.1 contains essential steps for the solving process (going from the problem situation to the situation model and on to the mathematical model), we argue that the mental activity needed for the necessary steps in the process is interdependent on the mathematics domain of the task. In the domain of numbers the mathematical model is primarily computational and thus one dimensional. In that case a mainly depictive representation was presumed not to contribute considerably to the ease with which problem solvers make sense of the situational or mathematical model. In this domain most items gave no significant difference, even one opposite effect. In the domain of proportions the mathematical model is in general more complex than in the domain of numbers, because there is always some activity of (relatively) comparing quantities or comparing a quantity with a whole. A mainly depictive representation was assumed to be beneficial here. At the same time a counter-effect is possible if the mathematical model and the depictive representations are not mutually beneficial, which might lead to an increasing complexity. So for tasks from the domain of proportions one could not make a plausible straightforward prediction, whether a mainly depictive representation could help the solvers to construct the appropriate mathematical model and hence help them in solving the problem in a successful way. The results supported this line of reasoning.

In the domain of measurement & geometry the underlying problem situation in itself is two- or three-dimensional. So, a mainly depictive representation of the problem was assumed to help the problem solver to create the appropriate (mental) mathematical model. We saw in both trials that of the four items that significantly favour the image-rich numeracy problem, three are in the domain of measurement & geometry, so this assumption is supported by the data.

In this replication, we found fewer tasks with a significant difference between the A- and the B-versions. With smaller samples, small increases in performance cannot be labelled as statistically significant. Nevertheless the findings give enough incentive for further research in the design of numeracy tasks and the way reality is (re)presented in those tasks.

6.6 Conclusion

Word problems are a dominant feature of both classroom teaching and assessment of numeracy worldwide, and also of large-scale international assessments like TIMSS, PISA, and PIAAC (Mullis & Martin, 2013; OECD, 2013c; PIAAC Numeracy Expert Group, 2009). Lessons learned from these assessments have been brought together recently, see, for instance, Tout and Gal (2015). Despite these efforts and despite the dominant use of word problems to teach and assess people's ability to solve practical numerical problems, not much research has been conducted that systematically focuses on the effect on students' performance of changing the verbal representation of the problem situation to a mainly depictive representation or a more authentic representation of the problem situation.

The original trial and this replication have limitations. The participants in the adult sample were not representative of all adults in the Netherlands. And although the replication strengthened some of the conclusions from the earlier large-scale school trial, the conclusions were still based on a limited number of items. More research is necessary to establish whether the results hold for other sets of problems that are paired in the same way as in these trials. The overall difference in results is small and effect sizes related to those results are in most cases very small. The change in representation of the problem situation can thus be labelled as hardly relevant. At the same time, the results of the change were significant and consistent, and not influenced by other variables, so there is, arguably, enough justification to speak of a small but robust effect.

Our suggestion is that in the task design of future assessments the representation of the problem situation should be taken into account as a factor when interpreting the results.

Chapter 7. Conclusions and discussion

7.1 Introduction

In this chapter the most important findings and conclusions of our studies are synthesised to answer the research question as formulated in Chapter 1:

> In presenting contextual mathematical problems, what is the effect on student performance of changing a descriptive representation of the problem situation to a mainly depictive one?

Improvement in student performance could be an indicator that by changing the representation of the problem situation to mainly depictive, students' chances of making more sense of the problem increase. In other words, it could be an indicator that students will suffer less from the suspension of sense-making that is often witnessed, and reported in the literature, when students engage in mathematical (word) problem solving.

Furthermore, here I discuss what our studies have added to the body of knowledge on contextual mathematical problem solving, sense-making, and students' performance on contextual mathematical problems with different representations of problem situations. Special attention has been given to the relevance of the studies by discussing possible spin-offs of the findings in classroom practice, and by making suggestions for future research.

7.2 Main findings

7.2.1 An exploratory study

Before the actual design and validation of the instrument a small exploratory study was conducted to find indications on which factors might have a positive effects on students' sense-making when dealing with mathematical problems in daily life and work. This was a first step towards a better understanding of which characteristics of the tasks to be designed were appropriate to improve the chance that the students made sense of the problem situation, and showed a suitable problem-solving attitude.

In this study, semi-structured interviews were conducted with students on situations in daily life or in work related settings. The actual behaviour of individuals in numeracy situations, as described in Chapter 2, is a complex amalgam of skills, knowledge, and personal attributes. This study highlighted one small part of that large domain: the way young adults spoke about their behaviour in numeracy situations. We found that in the interviews we could identify a large amount of so-called “numeracy incidents”, and we found we were able to categorise these by the content categories from the international PISA framework (quantity, shape & space, change & relationship, and uncertainty & data) as well as by the content categories of the Dutch Literacy and Numeracy Framework (LaNF) (numbers, proportions, measurement & geometry, and relations). This gave a good indication that the content domains of these numeracy frameworks could be related to the numeracy practices of vocational students and that tasks from these content domains might give an adequate impression of the reasoning, knowledge and skills needed for future demands regarding numeracy tasks in real life.

Moreover, we found that selected students, when interviewed on work related situations, in particular on artefacts they produced themselves, showed better understanding of mathematical concepts when the produced artefacts were literally at hand. In other words, in this context they showed that they could better make sense of the prevailing situation regarding the mathematical aspects involved.

7.2.2 The validation of the instrument

The results of the exploratory study encouraged us to design tasks with authentic visual representations of problem situations by using photographs of real situations from daily life or work situations (see Appendix A for examples). We conjectured that the change would help to avoid the suspension of sense-making that is often witnessed when students engage in solving contextual mathematical problems.

The instrument was designed to measure the difference in students’ performance after changing the representation of the problem situation from descriptive to mainly depictive. To measure this effect and make plausible inferences, the instrument was programmed as a digital test that could validly measure such an effect. For the validation of this instrument three activities were undertaken. First, alternative items for existing word problems were designed and referenced with the content domains of the LaNF; second, two panels of experts from the fields of mathematics and numeracy education were asked to comment on the problems used in the instrument, and to give their expert opinions on the degree of equivalence of the selected word problems and the alternative image-rich numeracy problems; third, a test run with a tentative instrument was held to see whether the condition of a controlled randomised trial could be met. In this way the major threats to validity – content validity, construct validity, criterion validity, reliability, and feasibility – were countered. These activities around validations were carried out successfully, so that the final instrument could be used to measure validly the effect on students' performance of changing the representation of the problem situation.

7.2.3 The main result of the trial

After the design of the instrument, it was programmed as an online test. In a four-week period at the end of 2011, the test was made available by internet for students aged 10–20 years, to be used as a numeracy test. In total, 31,842 students from 179 schools geographically spread across the Netherlands participated in the test. The students were assumed to be representative of Dutch students in general.

To test the prediction that students would likely score higher on image-rich numeracy problems, we carried out the relevant statistical analyses (see Chapter 4). To get more in-depth insight into the effects of the background variables on the results and their possible interaction effect with the manipulated variable, we used for that analysis a probit model, which is a limited dependent variable model. In both analyses, the overall result was a 1.9 percentage points rise when the representation was changed to an image-rich problem situation. This was statistically significant ($p < .001$) with a very small effect size of Cohen's $d = 0.09$. Our prediction was supported, but this statement should be regarded with great caution given the small effect size. Thus, the result needs more research (e.g., replication studies), whether it was to be of practical value and regarded as significant for educational practice.

However, through the design, the number of participants, and the use of a probit model to analyse the results, we were able to make more detailed inferences about the results.

First, due to the large sample the result of a 1.9 percentage-point rise came with a 95% confidence interval of [1.7 ; 2.1]. Second, from the probit analysis we were able to check for the interdependent effects of the other measured and not-measured background variables, to see if they could explain the found effect. The measured background variables were gender, ethnicity, educational track, grade and age, and most recent mathematics grade. As expected, all these variables had a significant effect on the scores of the students. The educational track (primary education, vocational track, general track) was the most influential variable in the overall result. Ethnicity had a two percentage-point effect (in favour of participants from non-migrant families) and gender a five percentage point effect in favour of male participants. Third, we looked for interaction effects of the variable version with the other variables. The other variables – when taken as cross terms with version in the model – showed no significant interaction with the variable version, with the exception of the variable gender. This means that the measured effect was not dependent on the background variables, except for gender. Gender explained about 35% of the measured effect. In other words, for male participants the difference in scores on both versions of the problems was slightly higher than for female participants. Similar results were earlier found in the PISA 2012 results on gender effects in solving contextual problems (Stacey & Turner, 2015).

7.2.4 An in-depth analysis into tasks characteristics.

A further analysis was carried out regarding some specific characteristics of the tasks. Two variables were used to indicate the complexity of the tasks: wordiness for the A-versions of the problems and the number of pictorial elements for the B-version of the problems. The statistical probit model, used earlier in the overall analysis, was also used to make inferences on the effect of the task characteristics on students' performance, but also on the interdependent effect of these characteristics with the variable version. The conclusion was that the wordiness of the word problem had a significant negative relation to the students' performance: higher wordiness indicated lower performance on A-versions of the problem. We also concluded that the number of pictorial elements was significant for the students' performance and that the effect was negative: a higher number of pictorial elements indicated lower performance on B-versions of the problem.

Furthermore, the marginal effects found in the statistical model approach gave an indication of the effect on performance of changing from a descriptive representation to a depictive representation. We found a tentative measure on the magnitude of the increase in in students' performance for each word reduced, also on the magnitude of the decrease in students' performance for each depictive element added (see Chapter 5 for details). These results gave some indications on the relation between number of words (A-version), numbers of pictures (B-version), and complexity of the tasks.

Another task characteristic examined was the content domains of the tasks. The content domains used from the Dutch LaNF – numbers, proportions, and measurement & geometry – were represented with seven items each in the instrument. From cognitive psychology research (Schnotz et al., 2010) we expected that for the domain of measurement & geometry the effect of changing the representation of the problem situation would probably have a bigger effect than for tasks in the domains of numbers and proportions. We examined the interdependent effects of the variables regarding the content domains and the variable regarding version, and we found a significant and negative effect of the cross term proportion × version, and a significant and positive effect of the cross term measurement & geometry × version. This meant that the effect of changing the representation of the problem situation had the least effect in the domain of proportions, and the most effect in the domain of measurement & geometry. The effect in the domain of numbers was in between these effects, but closer to the effect in the domain of proportions. The effect on tasks in the domain of measurement & geometry was in line with our expectations: the depictive representation in these kinds of tasks was likely to be beneficial in the problem-solving process. The result of the tasks in the domain of proportions led us to believe that in these tasks the mathematical model needed for the solution and the depictive representations were not mutually beneficial, which possibly led to an additional experienced complexity. So, overall our expectation was corroborated by the analysis.

7.2.5 A small replication study

We replicated the randomised trial in a different target group, namely adults. The data of the replication study were obtained in the Groot Nationaal Reken Onderzoek [Great National Numeracy Survey, GNRO], a research initiative by the public broadcasting organisations VPRO and NTR, supported by the Netherlands Organisation for Scientific Research (NWO). Individuals of all ages and of all places in the Netherlands could register as participants on the GNRO website and could engage in a series of mathematical tests.

The overall results on performance in this study with adult participants was 1.1 percentage point higher scores on image-rich numeracy problems. This was in line with the overall results we found in the large school trial, but slightly lower. In some items the effect of higher scores on the B-version occurred with a small or less than small effect size, in other items there was no effect at all. In one item the effect was even opposite. We found four tasks that in both trials showed a significantly better performance on the image-rich versions. Three of them were from the domain of measurement & geometry. This finding corroborates our earlier findings that with tasks from the domain of measurement & geometry the change in representation of the problem situation has the greatest positive influence on the performance of the solver.

To establish whether the effect on students' performance of changing the representation of the problem situation in contextual mathematics problems is robust and generic, more replications or expansion of the instrument with more (pooled) items could shed more light and corroborate, nuance or contradict these findings. The instrument is translated to English and available through open access, which gives the opportunity for replication in English-speaking countries. Replications can be conducted with other items or with more items, for example, specifically to test whether the effect holds true for the various content domains of the items.

7.2.6 General conclusion

Overall, we found a positive result on students' performance from changing the representation of the problem situation from descriptive to mainly depictive, as an indicator that students were able to make more sense of the depictive version of the tasks. We were able to exclude most of the other background variables (gender, ethnicity, level of student, educational track) as possible interdependent variables. Due to the large size of the sample we could make valid inferences despite the small effect size. These findings are not conclusive. In our considerations we regard the findings as a valuable starting point for future research into the more precise effects of changing the representation of the problem situation in solving contextual mathematical problems, and the underlying processes that may cause this effect. In the next section I will discuss the findings in more detail.

7.3 Further discussion

We chose four themes to discuss more elaborately. We identified these themes as the most essential elements in our attempt to answer the main research question. The themes were: (1) the connection between the "real world" and the "world of mathematics"; (2) the specific design of the tasks; (3) the effects of the change in representation on students' behaviour; and (4) the interpretation of the results of the trials conducted.

7.3.1 The "real world" and "the world of mathematics"

The connection between the real world and the mathematical world is a central theme in our studies and an essential ingredient in answering our research question. Many researchers have theorised on this connection (Blum et al., 2007; Gerofsky, 2010; Schoenfeld, 1992, 2014; Stacey, 2015; Verschaffel et al., 2009; Wake, 2014), as it is a crucial ingredient in problem solving, as illustrated in the problem-solving models and modelling diagrams (see Chapter 1). The "world of mathematics" is arguably part of the real world, but the distinction is useful when discussing modelling and problem solving (see Chapter1, Figures 1.1, 1.2, and 1.3). The exploratory study, which is described in Chapter 2, shed some more light on a possible aspect of the connection between the real world and the mathematical world, which is essential to the concepts of numeracy and mathematical literacy (Stacey, 2015). We found a clear discrepancy between students talking about numerical issues in a narrative or in dialogue with the interviewer and students talking about numeracy aspects with an artefact at hand (e.g., one made by themselves). In the latter situation the students showed more confidence and more mathematical reasoning, often accompanied by supporting gestures. This gave us an indication that concrete materials and artefacts from personal experiences contributed to the sense-making in mathematical dialogue. It also corroborates earlier research results (Frankenstein, 2009; Lave & Wenger, 1991; Zevenbergen & Zevenbergen, 2009).

The limitations identified in the exploratory study (described in Chapter 2) were the rather rough categories under which the numeracy incidents were categorised, and the lack of inter-subjectivity in the classifications. Moreover, the conclusion that students were better at expressing themselves mathematically with the artefacts at hand were not systematically collected and analysed. Nevertheless, the study gave useful information for the design of the tasks with mainly depictive representations of the problem situation.

7.3.2 The design of tasks

The next step in answering the research question was the design of relevant image-rich numeracy problems. The design of tasks that reflect the connection between the real world and the mathematical world in a meaningful way is neither easy nor straightforward (Gellert & Jablonka, 2009; Tout & Spithill, 2015). For the design of the instrument, existing word problems were transformed into image-rich numeracy problems, while keeping the problem question the same (see Appendices 1 and 2). Alternative, image-rich numeracy problems were designed by using photographs of concrete and practical situations as representation of problem situations. We anticipated that the students would be better able to make sense of the more realistic problem situations. Stacey (2015), in reflecting on the design of PISA 2012 items, described it as follows: “Items should not just be authentic; they should appear to be authentic so that students feel they are engaged in a sensible endeavour" (p.77).

The limitations of the instrument have already been discussed in Chapters 4, and 5. One limitation was the small number of 21 paired items used in the trial. Using such a small number of items meant that there was an increased risk that the measured differences were coincidental, especially where we further broke down the number of items in terms of the domains of the tasks. If we had anticipated that more than 30,000 students would participate, we could have pooled more items, a strategy used in international assessments such as PISA and PIAAC.

Another limitation was the assumption that the results of digital assessments, however sophisticated their design, give an indication of the transfer of knowledge and skills to actual authentic problems. This assumption can be rightfully questioned. That is why many educators advocate assessment in more authentic situations (Bonotto, 2007; Frankenstein, 2009), but practical constraints in schools impedes broad implementation of such assessments. Well-designed image-rich numeracy problems could possibly fill this gap to some extent.

7.3.3 The effects on students' behaviour

In designing the instrument for the trial, the choice was made to measure students' scores, and not to investigate their actual problem-solving behaviour (e.g., in classrooms). This was a consequence of the desire to conduct a large-scale trial with a web-based set of tasks. We noticed such a focus on students' performance as indicator in more studies in related fields (Berends & van Lieshout, 2009; Piel & Schuchart, 2014). We realised that the actual behaviour of students when solving image-rich numeracy problems could be quite different from their behaviour in solving word problems. Using students' performance as the single indicator of the degree in which the students made sense of a task has been a limitation of this study.

An important aspect of students' behaviour is the possibility that they are trained in solving word problems, and that this training induced such a strong attitude towards a calculational approach in solving numeracy problems that more realistic problem situations could cause confusion, and become more burdensome than helpful. This could have a mitigating effect on the prediction that students could make more sense of more realistic represented numeracy problems with an improvement of their scores as a consequence. Dewolf et al. (2014) found in their studies no impact of illustrations and warnings on solving word problems more realistically, and they concluded that there findings: “provides thus further strong empirical evidence for the very strong and persistent nature of traditionally schooled students’ tendency to approach and handle school word problems in a non-realistic way” (p.116).

Yet another aspect that could have a mitigating effect on better performance on image-rich numeracy problems was discussed in the studies of Piel and Schuchart (Piel & Schuchart, 2014; Schuchart, Buch, & Piel, 2015). They discussed the effect of social class and everyday knowledge on students’ performance on problems that are more realistic than straightforward word problems. They predicted that lower-class students would perform worse on more realistic problems because of a more limited experience with a broad array of real-life problems, but in the end could not find evidence for that, which was a similar result to earlier studies by Cooper et al. (Cooper & Dunne, 2000; Cooper & Harries, 2009).

Investigating the actual behaviour of students, for example by observation, by thinking-aloud protocols, or by eye tracking, could offer more detailed insights into the changes in actual student behaviour when solving the problems with different representations of the problem situations. More data on students’ characteristics, for instance, their familiarity with word problems, could also shed more light on the underlying mechanisms that could cause the results we found.

7.3.4 The interpretation of the results of the trials

The overall result of the trial with 31,842 participants (as reported in Chapter 3) was as follows. On average students scored approximately two percentage point higher on tasks with depictive representations of the problem situation than on the word problems, with an effect size of Cohen’s $d = 0.09$. Although these inferences are statistically correct, the overall result is of little value unless communicated with the complete context of the instrument used, including detailed knowledge of the items used. Effect size as such should never be the only aspect to evaluate outcomes of a study (Ellis, 2010). Ellis (2010) emphasised that simple lists and simple rankings of effect sizes are not a good way to compare educational studies and their outcomes. Only the complete context of the study, including the circumstances and the nature of the intervention, should be taken into account to make valid and useful inferences when interpreting results from research – even more so when research results are used to invoke new policies in education.

Slavin (2016) has recently stated, in a Huffington Post blog “What is a Large Effect Size?”, that in educational studies using a randomised controlled trial, effect sizes are seldom found over 0.2, and that with research populations over 5000 participants valid inferences could be made with effect sizes as low as 0.1. Following Slavin (2016), we may argue that inferences were justified in our study.

To conclude, by doing further statistical analysis we were able to show that most background variables had no significant interaction effects with the variable version. Our conclusion was that changing the representation of the problem situation to a mainly depictive representation of the problem situation is most likely an independent phenomenon, with different magnitudes depending on the content domains of the tasks.

7.3.5 Answering the research question

After synthesising the conclusions our answer to the research question was: Changing the problem statement from descriptive to mainly depictive by using photographs improves the chance that students will correctly solve the stated problem. The improvement is small and depends to some extent on the content domain of the tasks.

Limitations of the studies were identified, so the results should be interpreted with great care. Nevertheless, our findings add to the body of knowledge on contextual mathematical problem solving and sense-making, as it is to our knowledge the first study that has systematically, and on a large scale, compared students’ performance on different representations of the problem situation in contextual mathematical problem solving.

7.4 Relevance of the studies

7.4.1 General relevance

The reported studies have different layers of relevance: (1) theoretical and (2) practical. The studies provide a small contribution to one of the more fundamental issues in mathematics education, and in problem solving in particular – how to represent or simulate "real life" in the classroom for educational purposes. Engaging in problems from real life in the classroom is widely considered an important aspect of preparing students for their future life.

In mathematics education problems from real life are often represented as contextual mathematical problems. Word problems, in which the problem situation is described by language, are the staple and dominant form of such problems. Research on word problems, as a specific genre in mathematics education, is extensive (Dewolf et al., 2014; Dewolf et al., 2015; Palm, 2006, 2008, 2009; Verschaffel et al., 2000; Verschaffel et al., 2009). Nevertheless, we found a dearth of research that investigated the effect of changing the representation of the problem situation from descriptive to mainly depictive. In that sense, the findings from our studies can lead to further reconsiderations in future research on mathematical contextual problems, for instance, what kind of depiction could help the students to make sense of the problem, to avoid suspension of sense-making, and hence to improve students' performance. This could be of relevance for designers of numeracy problems who look for more authentic, more realistic, and likely more meaningful, numeracy problems.

7.4.2 Relevance for future research

Better understanding of the actual behaviour of students in contextual mathematical problem solving, and the way they make sense of the problem situation, is worth further investigation. Here I will sketch a few directions.

Comparable trials in other countries and cultures or trials with other items. The studies in this investigation were carried out in the Netherlands. Countries differ on aspects of curriculum, experience with the use of numeracy tasks, the degree of realism used in mathematics lessons, and more. Follow-up research questions in this direction could be:

- Would the difference in students' performance depend on the degree of realistic contexts being used in a country's curriculum?
- Would the difference in students' performance depend on their exposure to formal mathematics? (Exposure to formal mathematics as a variable was also measured in PISA 2012 results (OECD, 2013b).)
- Would the difference in students' performance depend on teacher quality? (Teacher quality as a variable was also measured in PISA 2012 results (OECD, 2013b).)

These differences could be identified and taken into account as variables in the statistical model to establish their effects on the difference in students' performance on descriptive and depictive presented problems. Thus, conducting similar studies in other countries could give indications on the effect of cultural aspects on the results. Researchers from countries other than the Netherlands are invited to use the instrument described in this study. For that purpose, an English translation of the instrument has been made available under open access (Hoogland & De Koning, 2013), and (on request) the web-based test can be adapted for use in other countries.

Another way to improve the plausibility of our findings is by repeating the trial with a larger amount of (pooled) items from the content domains in the LaNF or from the content domains in the international frameworks, like PISA and PIAAC.

Investigating actual student behaviour in contextual mathematical problem solving. Investigating the actual behaviour of students in solving contextual mathematical problems could shed more light on the ways they make sense of the posed problems. Future research could find answers to specific research questions, such as:

- Does it take longer for students to solve image-rich numeracy problems? This could give indications of the cognitive load involved in solving specific kinds of problems. Data for this is already available in our research database (Hoogland & De Koning, 2013) and will be part of a future study;

- What is the typical "reading behaviour" of students in solving the numeracy tasks? This can be measured, for example, by eye tracking (e.g., Hegarty & Mayer, 1992). Some interesting results in comprehensive reading and processing mechanisms are reported by Van Silfhout (2014), who found a clear connection between comprehension and time needed to process the necessary information;
- What are the typical problem-solving considerations made by the students in both versions of representing the problem situation? This can be investigated by analysing students' work (e.g., Dewolf et al., 2014, 2015), by thinking-aloud protocols, or by stimulated recall with semi-structured interviews, comparable with the method used in our exploratory study described in Chapter 2.

Incorporating Cognitive Load Theory in the theoretical framework. In our research project we conducted a small study to explore to what extent the effects we found in changing the representation of the problem situation could be explained by insights from Cognitive Load Theory (CLT) (Paas, Van Gog, & Sweller, 2010; Sweller, 2010; Sweller, van Merrienboer, & Paas, 1998; Van Merriënboer & Sweller, 2005). An expert on CLT was willing to look at the 21 paired problems and to make an expert estimation on the relative performance of the students on both versions. The CLT expert estimated students' performance on image-rich problems systematically lower than the results from the trial indicated. However, there was a strong correlation between differences estimated by the CLT expert and the differences measured in the trial. This could give an indication that CLT could explain to some extent the difference in performance between the two versions, but that the effect of using mainly depictive representations of the problem situation next to a verbally posed question is less negative than expected from the CLT perspective. It seems that the reduction in suspension of sense-making by using depictive representations of problem situations, could arguably have an opposite effect to that which is predicted from the CLT perspective, and so mitigate the prediction from the CLT perspective. A similar observation was stated in a study by Schüler, Arndt, and Scheiter (2015), namely that students were very well capable of combining both sorts of information, textual and pictorial, into one mental representation. Further and more systematic research is necessary to shed light on these seemingly contradictory theories. A research design could be chosen that incorporates both perspectives by comparing students’ performance on a multitude of different representations of problem situations: descriptive, mainly depictive, completely depictive with the question in voice-over, with short video-clips, or with augmented reality.

Multimedia representation of the problem situation. Technological developments provide new possibilities to represent the problem situations of quantitative problems from real life even closer to the actual problem. Instead of using photographs and other static information carriers, there are now more possibilities to represent the problem situations with short video-clips, dynamic animations, and augmented reality (Bower, Howe, McCredie, Robinson, & Grover, 2014). In the near future, representations of problem situations could evolve to instances of augmented reality for educational purposes. Nowadays, we are witnessing the rapid emergence of augmented reality systems, which allow designers to overlay recording of real-life situations in real time with extra information or questions. Examples are Aurasma (http://aurasma.com), Layar (http://layar.com), and Junaio (http://junaio.com). Some educational uses are already being explored (Bower et al., 2014; Sommerauer & Müller, 2014). This gives rise to inspiring new perspectives on how to bring problems from real life into classrooms to prepare students for mathematical problem solving in work situations (Wake, 2014), and in practical day-to-day real-life problems.

7.4.3 Relevance for classroom practice

The practical relevance of this study has a broad scope. Teachers can benefit from the findings in their design of numeracy learning materials and numeracy assessment tasks, taking into account that the way in which the problem situation is represented matters for students’ performance and students’ sense-making.

The findings from our studies could be an incentive for teachers and other designers of teaching materials to (re)consider the chosen representations of real-life situations in their products. Synthesising our findings on the role of depictive elements in the design of the problems, we suggest a short checklist regarding the used images:

- Are the images realistic and relevant for solving the problem (in contrast to cartoon-like and distracting)?
- Are the images an integral part of the representation of the problem situation (in contrast to merely illustrative)?
- Are the images relevant and consistent with the envisioned mathematical concepts and mathematical models that are involved in the problem solving (in contrast to irrelevant or even confusing)?

Due to new technologies, the barriers to implementing findings from our studies in classroom practice are getting lower. It is nowadays technically much easier to incorporate photographs in portfolio or digital learning materials, and the opportunities to show situations from real life with interactive whiteboards are rising, although some inertia on these developments has been reported (De Vita, Verschaffel, & Elen, 2014). In current education, describing the context in words as a representation of the problem situation seems to be more a habit than an explicit choice. For the further development of image-rich numeracy problems a databank of photographs of everyday problem situations would be a great stimulus to implement a greater use of such problems.

7.4.4 Relevance for assessment

Professional test designers might benefit from the findings of our studies in further validating the ways in which they translate goals of numeracy frameworks to actual test items. For instance, in the international PISA ranking on countries' mathematics performances a rise and fall of a few percentage points has at times led to far-reaching policy decisions on mathematics and numeracy programmes and curricula. In that light it is important to know that changing the representation of the problem situation to a more authentic and image-rich form can itself have an effect of a few percentage points, or even more depending on the domain of the tasks.

In the Dutch context in 2016, the Ministry of Education, Culture and Sciences decided to make a separate final numeracy examination for low-achieving students in the vocational tracks, called Rekentoets 2A [Numeracy test level 2A]. We show two examples (See Figures 7.1 and 7.2) from the series items that were publicly released (College voor Toetsen en Examens, 2015) as part of the information campaign. In these exemplary assessment items for low-achieving students in the vocational tracks we saw a strong resemblance to the type of items we used in our instrument. This gives a small indication that the kind of research described in this dissertation can find, and probably to some extent has found, its way to institutions that are responsible for the nationwide assessment of numeracy of Dutch students.

Fig. 7.1 Example from Rekentoets 2A [Numeracy test – level 2A], released by the College voor Toetsen en Examens (2015).

Fig 7.2 Example from Rekentoets 2A [Numeracy test – level 2A], released by the College voor Toetsen en Examens (2015).

To conclude, we may argue that with a better knowledge of the underlying factors that explain students' behaviour in solving quantitative problems from daily life, using mainly depictive representations in lesson materials has the potential to offer a better way of fostering students' abilities to solve quantitative problems from daily life.

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters, 80*(1), 123-129. doi:10.1016/s0165-1765(03)00032-6

American Institutes for Research. (2006). *A review of the literature in adult numeracy: Research and conceptual issues*. Washington DC.

Askew, M. (2015). Numeracy for the 21st century: A commentary. *ZDM – Mathematics Education, 47*(4), 707-712. doi:10.1007/s11858-015-0709-0

Bakker, A. (2014). Characterising and developing vocational mathematical knowledge. *Educational Studies in Mathematics, 86*(2), 151-156. doi:10.1007/s10649-014-9560-4

Berends, I. E., & van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction, 19*(4), 345-353. doi:10.1016/j.learninstruc.2008.06.012

Biesta, G. J. J. (2010). *What is education for? Good education in an age of measurement: Ethics, politics, democracy*. London, UK: Taylor & Francis.

Biesta, G. J. J. (2013). *The beautiful risk of education*. London, UK: Taylor & Francis.

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

Blum, W. (2015). Quality teaching of mathematical modelling: What do we know, what can we do? In S. J. Cho (Ed.), *The Proceedings of the 12th International Congress on Mathematical Education: Intellectual and attitudinal challenges* (pp. 73-96). Seoul, Korea: Springer International.

Blum, W., Galbraith, P. L., Henn, H.-W., & Niss, M. (Eds.). (2007). *Modelling and applications in mathematics education – the 14th ICMI study*. New York, NY: Springer International.

Blum, W., & Leiß, D. (2007). How do students and teachers deal with modelling problems? In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering and economics.* (pp. 222-231). Chicester, UK: Horwood.

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

Bonotto, C. (2007). How to replace the word problems with activities of realistic mathematical modeling. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education – the 14th ICMI study* (pp. 297-314). New York, NY: Springer International.

Bonotto, C. (2009). Working towards teaching realistic mathematical modelling and problem posing in Italian classrooms. In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 297-314). Rotterdam, The Netherlands: Sense.

Boonen, A. J. H. (2015). *Comprehend, visualize & calculate - Solving mathematical problems in contemporary math education.* (Doctoral thesis), Vrije Universiteit, Amsterdam, The Netherlands.

Boonen, A. J. H., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research, 68*(2014), 15-26. doi:http://dx.doi.org/10.1016/j.ijer.2014.08.001

Borooah, V. K. (2002). *Logit and probit: Ordered and multinomial models*. London, UK: Sage.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 4*(111), 1061-1071.

Bower, M., Howe, C., McCredie, N., Robinson, A., & Grover, D. (2014). Augmented reality in education – cases, places and potentials. *Educational Media International, 51*(1), 1-15. doi:10.1080/09523987.2014.889400

Burkhardt, H. (2006). Modelling in mathematics classrooms: Reflections on past developments and the future. *ZDM – Mathematics Education, 38*(2), 178-195. doi:10.1007/BF02655888

Caldwell, L. (1995). *Contextual considerations in the solution of children's multiplication and division word problems.* (Master's thesis), Queen's University, Belfast, Nothern Ireland.

CBS. (2012). *Jaarboek Onderwijs in cijfers 2011 [Yearbook Education in numbers 2011]*. Heerlen, The Netherlands: CBS.

Cito. (2015). Voorbeeldtoets 2F 2014 [Example numeracy test level 2F 2014]. Retrieved from http://www.cito.nl/~/media/cito_nl/files/voortgezet%20onderwijs/cito_voorbeeldtoets_2f_2014.ashx

Coben, D. (2003). *Adult numeracy: review of research and related literature*. Retrieved from London, UK: www.nrdc.org.uk

Cohen, J. (1988). *Statistical power analysis for the behavorial sciences (second edition)*. Hillsdale, NJ: Lawrence Erlbaum.

College voor Toetsen en Examens. (2015). Voorbeeldexamen rekenen 2A [Example of numeracy exam - level 2A].

Cooper, B., & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex and problem-solving*. Buckingham, UK: Open University Press.

Cooper, B., & Harries, T. (2002). Children's responses to contrasting 'realistic' mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics, 49*(1), 1-23. doi:10.1023/a:1016013332659

Cooper, B., & Harries, T. (2003). Children's use of realistic considerations in problem solving: some English evidence. *The Journal of Mathematical Behavior, 22*(4), 449-463. doi:10.1016/j.jmathb.2003.09.004

Cooper, B., & Harries, T. (2009). Realistic contexts, mathematics assessment, and social class: Lessons for policy from an English research programme. In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 93-110). Rotterdam, The Netherlands: Sense.

Cuendet, S., Bonnard, Q., Do-Lenh, S., & Dillenbourg, P. (2013). Designing augmented reality for the classroom. *Computers & Education, 68*, 557-569. doi:http://dx.doi.org/10.1016/j.compedu.2013.02.015

D'Ambrosio, U. (1997). Ethnomathematics and its place in the history and pedagogy of mathematics. In A.B.Powels & M.Frankenstein (Eds.), *Ethnomathematics: Challenging Eurocentrism in mathematics education*. Albany, NY: State University of New York.

Daro, P. (Producer). (2013). Phil Daro - against answer getting. [Video] Retrieved from https://vimeo.com/79916037

De Bock, D., Deprez, J., Dooren, W. V., Roelens, M., & Verschaffel, L. (2011). Abstract or concrete examples in learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal for Research in Mathematics Education, 42*(2), 109-126.

De Brabander, C. J., & Martens, R. L. (2014). Towards a unified theory of task-specific motivation. *Educational Research Review, 11*, 27-44. doi:http://dx.doi.org/10.1016/j.edurev.2013.11.001

De Smedt, B., Ansari, D., Grabner, R. H., Hannula, M. M., Schneider, M., & Verschaffel, L. (2010). Cognitive neuroscience meets mathematics education. *Educational Research Review, 5*(1), 97-105. doi:http://dx.doi.org/10.1016/j.edurev.2009.11.001

De Vita, M., Verschaffel, L., & Elen, J. (2014). Interactive whiteboards in mathematics teaching: A literature review. *Education Research International, 2014*, 16. doi:10.1155/2014/401315

DeFranco, T. C., & Curcio, F. R. (1997). A division problem with remainder embedded across two contexts: children's solutions in restrictive versus real world settings. *Focus on Learning Problems in Mathematics, 19*(2), 58-72.
Deloitte. (2014). *Mathematical sciences and their value for the Dutch economy* Retrieved from Amsterdam, The Netherlands: www.platformwiskunde.nl
Depaepe, F., De Corte, E., & Verschaffel, L. (2010). Teachers' approaches towards word problem solving: Elaborating or restricting the problem context. *Teaching and Teacher Education, 26*(2), 152-160. doi:http://dx.doi.org/10.1016/j.tate.2009.03.016
Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY: Simon & Schuster.
Dewolf, T., Van Dooren, W., Ev Cimen, E., & Verschaffel, L. (2014). The impact of illustrations and warnings on solving mathematical word problems realistically. *The Journal of Experimental Education, 82*(1), 103-120. doi:10.1080/00220973.2012.745468
Dewolf, T., Van Dooren, W., Hermens, F., & Verschaffel, L. (2015). Do students attend to representational illustrations of non-standard mathematical word problems, and, if so, how helpful are they? *Instructional Science, 43*(1), 147-171. doi:10.1007/s11251-014-9332-7
Dewolf, T., Van Dooren, W., & Verschaffel, L. (2011). Upper elementary school children’s understanding and solution of a quantitative problem inside and outside the mathematics class. *Learning and Instruction, 21*(6), 770-780. doi:10.1016/j.learninstruc.2011.05.003
Elia, I., Gagatsis, A., & Demetriou, A. (2007). The effects of different modes of representation on the solution of one-step additive problems. *Learning and Instruction, 17*(6), 658-672. doi:10.1016/j.learninstruc.2007.09.011
Elia, I., & Philippou, G. (2004). The functions of pictures in problem solving. In M. Johnsen Hoines & A. Berit Fuglestad (Eds.), *Proceedings of the 28th conference of the international group for the psychology of mathematics education* (Vol. 2, pp. 327-334). Bergen, Norway: Bergen University College.
Ellis, P. D. (2010). *The essential guide to effect sizes*. Cambridge, UK: Cambridge University Press.
Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 117 -133). New York, NY: Cambridge University Press.
Forman, S. L., & Steen, L. A. (1999). *Beyond eighth grade: Functional mathematics for life and work*. Berkeley, CA: National Centre for Research in Vocational Education.
Frankenstein, M. (2009). Developing a criticalmathematical numeracy through real real-life word problems. In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 111-130). Rotterdam, The Netherlands: Sense.
Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, The Netherlands: Reidel.
Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures* Dordrecht, The Netherlands: Reidel.
Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading, 19*(3), 204-223. doi:10.1080/10888438.2015.1005745
Gal, I., Groenestijn, M. v., Manly, M., Schmitt, M. J., & Tout, D. (2003). *Numeracy in the Adult Literacy and Lifeskills (ALL) Survey: An overview and sample items*. Retrieved from http://www.ets.org/all:
Galbraith, P., & Stillman, G. (2006). A framework for identifying student blockages during transitions in the modelling process. *Zentralblatt fuer Didaktik der Mathematik, 38*(2), 143-162. doi:10.1007/BF02655886

Geiger, V., Goos, M., & Forgasz, H. (2015). A rich interpretation of numeracy for the 21st century: A survey of the state of the field. *ZDM – Mathematics Education, 47*(4), 531-548. doi:10.1007/s11858-015-0708-1

Gellert, U., & Jablonka, E. (2009). "I am not talking about reality.": Word problems and the intracies of producing legitimate text. In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 39-54). Rotterdam, The Netherlands: Sense.

Gellert, U., & Jablonka, E. (Eds.). (2007). *Mathematisation and demathematisation – Social, philosophical and educational ramifications*. Rotterdam, The Netherlands: Sense.

Gerdes, P. (1994). Reflections on ethnomathematics. *For the Learning of Mathematics, 14*(2), 19-22.

Gerofsky, S. (2009). Genre, simulacra, impossible exchange, and the real: How postmodern theory problematises word problems In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 21-38). Rotterdam, The Netherlands: Sense.

Gerofsky, S. (2010). The impossibility of 'real-life' word problems (according to Bakhtin, Lacan, Zizek and Baudrillard). *Discourse: Studies in the Cultural Politics of Education, 31*(1), 61-73. doi:10.1080/01596300903465427

Gill, G. T., & Hicks, R. C. (2006). Task complexity and informing science: A synthesis *Informing Science Journal, 9*, 1-30.

Gravemeijer, K. (1994). *Developing realistic mathematics education*. Utrecht: CD-ß Press / Freudenthal Institute.

Gravemeijer, K. (1997). Solving word problems: A case of modelling? *Learning and Instruction, 7*(4), 389-397. doi:10.1016/s0959-4752(97)00011-x

Gravemeijer, K. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning, 1*(2), 155-177. doi:10.1207/s15327833mtl0102_4

Gravemeijer, K. (2004). Local instruction theories as means of support for teachers in reform mathematics education. *Mathematical Thinking and Learning, 6*(2), 105-128. doi:10.1207/s15327833mtl0602_3

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters, 107*(2), 291-296. doi:10.1016/j.econlet.2010.02.014

Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction, 7*(4), 293-307. doi:http://dx.doi.org/10.1016/S0959-4752(97)00006-6

Haines, C., Galbraith, P., Blum, W., & Khan, S. (Eds.). (2007). *Mathematical modelling: Education, engineering and economics.* Chichester, UK: Horwood.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyes relating to achievement*. Oxfordshire, UK: Routledge.

Hattie, J. (2015). *What doesn't work in education: The politics of distraction*. London, UK: Pearson.

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction, 14*(3), 343-351. doi:10.1016/j.learninstruc.2004.06.007

Hegarty, M., & Mayer, R. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84*(1), 76.

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal, 28*(4), 331-343. doi:10.1002/smj.582

Hoogland, K. (2006). Mind and gesture: The numeracy of a vocational student. In M. Horne & B. Marr (Eds.), *Connecting voices in adult mathematics and numeracy: Practitioners, researchers and learners. Proceedings of the adults learning mathematics 12th annual international conference (ALM)* (pp. 150-158). Melbourne, Australia: ACU.

Hoogland, K., & De Koning, J. (2013). *Dataset: Rekenen in beeld [Dataset: Images of numeracy]*. http://dx.doi.org/10.17026/dans-za6-5q6c.
Hoogland, K., Pepin, B., Bakker, A., de Koning, J., & Gravemeijer, K. (2016). Representing contextual mathematical problems in descriptive or depictive form: Design of an instrument and validation of its uses. *Studies in Educational Evaluation, 50*, 22-32. doi:10.1016/j.stueduc.2016.06.005
Hoogland, K., & Stelwagen, R. (2012). A new Dutch numeracy framework. In T. Maguire, J. J. Keogh, & J. O'Donoghue (Eds.), *Mathematical eyes: A bridge between adults, the world and mathematics. Proceedings of the 18th international conference of adults learning mathematics - a research forum (ALM)* (pp. 193-202). Tallaght, Ireland: ITT.
Hoyles, C., Noss, R., Kent, P., & Bakker, A. (2010). *Improving mathematics at work. The need for techno-mathematical literacies*. London, UK: Routledge.
Imbo, I., & Vandierendonck, A. (2007). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: Differences between children and adults? *Psychological Research, 72*(3), 331-346. doi:10.1007/s00426-007-0112-8
Inoue, N. (2009). The issue of reality in word problem solving: Learning from students' justifications of "unrealistic" solutions to real wor(l)d problems. In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 195-210). Rotterdam, The Netherlands: Sense.
Jablonka, E. (2003). Mathematical literacy. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second International Handbook of Mathematics Education* (pp. 75-102). Dordrecht: Kluwer Academic Publishers.
Jablonka, E. (2015). The evolvement of numeracy and mathematical literacy curricula and the construction of hierarchies of numerate or mathematically literate subjects. *ZDM – Mathematics Education, 47*(4), 599-609. doi:10.1007/s11858-015-0691-6
Jablonka, E., & Niss, M. (2014). Mathematical literacy. In S. Lerman, B. Sriraman, E. Jablonka, Y. Shimizu, A. M., R. Even, R. Jorgensen, & M. Graven (Eds.), *Encyclopedia of mathematics education* (pp. 391-396). Dordrecht: Springer.
Joint Committee on the Standards for Educational and Psychological Testing of the American Educational Research Association the American Psychological Association and the National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
Kaiser, G., Blomhøj, M., & Sriraman, B. (2006). Towards a didactical theory for mathematical modelling. *ZDM – Mathematics Education, 38*(2), 82-85. doi:10.1007/BF02655882
Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science, 320*(5875), 454-455. doi:10.1126/science.1154659
Kilpatrick, J. (1996). Introduction to section 1: Curriculum, goals, contents, resources. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 7-9). Dordrecht, The Netherlands: Kluwer.
Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London, UK: Routledge.
Krawec, J. L. (2014). Problem representation and mathematical problem solving of students of varying math ability. *Journal of Learning Disabilities, 47*(2), 103-115. doi:10.1177/0022219412436976
Lave, J. (1992). Word problems: A microcosm of theories of learning. In P. Light & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 74-92). New York, NY: Harvester Wheatsheaf.
Lave, J., & Wenger, E. (1991). *Situated learning, legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
Leonardo Pisano (Fibonacci). (1202). Liber Abaci.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*(0), 32-42. doi:http://dx.doi.org/10.1016/j.learninstruc.2013.12.001

Lesh, R., Post, T., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics*. Hillsdale, NJ: Lawrence Erlbaum.

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning*. Charlotte, NC: Information Age Publishing.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, November 2010*(136(6)), 1123-1135. doi:10.1037/a0021276

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Lowrie, T., Diezmann, C., & 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. doi:10.1007/s13394-012-0036-5

Lowrie, T., & Diezmann, C. M. (2009). National numeracy tests: A graphic tells a thousand words. *Australian Journal of Education, 53*(2), 141-158. doi:10.1177/000494410905300204

Madison, B. L., & Steen, L. A. (2008). *Calculation vs. context: Quantitative literacy and its implications for teacher education (1st edition)*. Washington, DC: Mathematical Association of America.

Maguire, T., & O'Donoghue, J. (2002). A grounded approach to practitioners training in Ireland: Some findings from a national survey of practitioners in Adult Basic Education. In L.O.Johansen & T. Wedege (Eds.), *Proceedings of the 8th International Conference of Adults Learning Mathematics (ALM)* (pp. 120-132). London, UK: Avanti Books.

Massa, L. J., & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences, 16*(4), 321-335. doi:10.1016/j.lindif.2006.10.001

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning (2nd edition)*. New York, NY: Cambridge University Press.

Ministerie van OCW. (2009). Referentiekader taal en rekenen [Literacy and numeracy framework]. Retrieved from http://www.taalenrekenen.nl/downloads/referentiekader-taal-en-rekenen-referentieniveaus.pdf/

Mullis, I. V. S., & Martin, M. O. (Eds.). (2013). *TIMSS 2015 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Niss, M. (1996). Goals of mathematics teaching. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal, 4*(2), 154-167.

OECD. (1999). *PISA mathematics framework*. Paris, France: OECD.

OECD. (2012). *Literacy, numeracy and problem solving in technology-rich environments*. Paris, France: OECD Publishing.

OECD. (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD Publishing.

OECD. (2013b). *PISA 2012 results: Excellence through equity. Giving every student the chance to succeed (volume II)*. Paris, France: OECD Publishing.

OECD. (2013c). *PISA 2015 draft mathematics framework*. Paris, France: OECD Publishing.

OECD. (2014). *PISA 2012 results: Creative problem solving. Students' skills in tackling real-life problems (volume V)*. Paris, France: OECD Publishing.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*(38), 1-4.

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review, 22*(2), 115-121. doi:10.1007/s10648-010-9133-8

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Palm, T. (2002). *The Realism of Mathematical School Tasks: Features and Consequences*: Umea Universitet.

Palm, T. (2006). Word problems as simulations of real-world situations: A proposed framework. *For the Learning of Mathematics, 26*(1), 42-47.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics, 67*(1), 37-58. doi:10.1007/s10649-007-9083-3

Palm, T. (2009). Theory of authentic task situations. In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 3-20). Rotterdam, The Netherlands: Sense.

Palm, T., & Burman, L. (2004). Reality in mathematics assessment: An analysis of task-reality concordance in Finnish and Swedish national assessment. *Nordic Studies in Mathematics Education, 9*(3).

PIAAC Numeracy Expert Group. (2009). *PIAAC numeracy: A conceptual framework*. Retrieved from Paris, France: http://dx.doi.org/10.1787/220337421165

Piel, S., & Schuchart, C. (2014). Social origin and success in answering mathematical word problems: The role of everyday knowledge. *International Journal of Educational Research, 66*(0), 22-34. doi:http://dx.doi.org/10.1016/j.ijer.2014.02.003

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90(1)*(March 1998), 25-36.

Platform Onderwijs 2032. (2016). *Ons onderwijs2032 - eindadvies*. Retrieved from The Hague, The Netherlands: http://onsonderwijs2032.nl/

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*(2), 137-157. doi:10.1016/j.jecp.2005.01.004

Reusser, K., & Stebler, R. (1997). Every word problem has a solution: The suspension of reality and sense-making in the culture of school mathematics. *Learning and Instruction, 7*(4), 309-327. doi:10.1016/s0959-4752(97)00014-5

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27-57. doi:10.1093/applin/22.1.27

Roth, W.-M. (2009). On the problematic of word problems – language and the world we inhabit. In L. Verschaffel, B. Greer, W. Van Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 55-69). Rotterdam, The Netherlands: Sense.

Ruthven, K. (2016). Numeracy in, across and beyond the school curriculum. In D. Wyse, L. Hayward, & J. Pandya (Eds.), *The SAGE handbook of curriculum, pedagogy and assessment* (pp. 638-654). Thousand Oaks, CA.

Säljö, R., Riesbeck, E., & Wyndhamn, J. (2009). Learning to model: Coordinating natural language and mathematical operations when solving word problems. In L. Verschaffel, B. Greer, W. V. Dooren, & S. Mukhopadhyay (Eds.), *Words and worlds – modelling verbal descriptions of situations* (pp. 177-194). Rotterdam, The Netherlands: Sense.

Santos-Trigo, M. (2014). Problem solving in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 496-501). Dordrecht, The Netherlands: Springer.

Scheiter, K., Gerjets, P., & Catrambone, R. (2006). Making the abstract concrete: Visualizing mathematical solution procedures. *Computers in Human Behavior, 22*(1), 9-25. doi:10.1016/j.chb.2005.01.009

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14*(1), 101-120. doi:10.1023/A:1013136727916

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge, UK: Cambridge University Press.

Schnotz, W., Baadte, C., Müller, A., & Rasch, R. (2010). Creative thinking and problem solving with depictive and descriptive representations. In L. Verschaffel, E. De Corte, T. De Jong, & J. Elen (Eds.), *Use of representations in reasoning and problem solving – analysis and improvement* (pp. 11-35). London, UK: Routledge.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141-156. doi:10.1016/s0959-4752(02)00017-8

Schnotz, W., & Kürschner, C. (2008). External and internal representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instructional Science, 36*(3), 175-190. doi:10.1007/s11251-007-9029-2

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 311-343). Hillsdale, NJ: Lawrence Erlbaum.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York, NY: McMillan.

Schoenfeld, A. H. (2007). Problem solving in the United States, 1970–2008: Research and theory, practice and politics. *ZDM – Mathematics Education, 39*(5-6), 537-551. doi:10.1007/s11858-007-0038-z

Schoenfeld, A. H. (2014). Reflections on learning and cognition. *ZDM – Mathematics Education, 46*(3), 497-503. doi:10.1007/s11858-014-0589-8

Schuchart, C., Buch, S., & Piel, S. (2015). Characteristics of mathematical tasks and social class-related achievement differences among primary school children. *International Journal of Educational Research, 70*, 1-15. doi:http://dx.doi.org/10.1016/j.ijer.2014.12.002

Schüler, A., Arndt, J., & Scheiter, K. (2015). Processing multimedia material: Does integration of text and pictures result in a single or two interconnected mental representations? *Learning and Instruction, 35*(0), 62-72. doi:http://dx.doi.org/10.1016/j.learninstruc.2014.09.005

Seegers, G., & Boekaerts, M. (1993). Task motivation and mathematics achievement in actual task situations. *Learning and Instruction, 3*(2), 133-150. doi:http://dx.doi.org/10.1016/0959-4752(93)90012-O

Sepeng, P. (2013). Issues of language and mathematics: Contexts and sense-making in word problem-solving. *Mediterranean Journal of Social Sciences, 4*(13), 51-61. doi:10.5901/mjss.2013.v4n13p51

Sepeng, P., & Webb, P. (2012). Exploring mathematical discussion in word problem-solving. *Pythagoras, 33*(1), 1-8. doi:http://dx.doi.org/10.4102/pythagoras.v33i1.60

Sigler, L. (2002). *Fibonacci's Liber Abaci – a translation into modern English of Leonardo Pisano's book of calculation*. New York, NY: Springer Verlag.

Slavin, R. E. (2016). What is a large effect size? Retrieved from http://www.huffingtonpost.com/robert-e-slavin/what-is-a-large-effect-si_b_9426372.html

SLO. (2014). *Advies over de uitwerking van de referentieniveaus 2F en 3F voor rekenen in toetsen en examens - commissie Bosker [Advice on the realisation of the numeracy levels 2F and 3F in examinations - Committe Bosker]*. Retrieved from Enschede:

Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & Education, 79*, 59-68. doi:http://dx.doi.org/10.1016/j.compedu.2014.07.013

Sriraman, B., Kaiser, G., & Blomhøj, M. (2006). A brief survey of the state of mathematical modeling aroung the world. *ZDM – Mathematics Education, 38*(3), 212-213. doi:10.1007/BF02652805

Stacey, K. (2015). The real world and the mathematical world. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy: The PISA experience* (pp. 57-84). Basel, Switzerland: Springer International Publishing.

Stacey, K., & Turner, R. (Eds.). (2015). *Assessing mathematical literacy: The PISA experience*. Basel, Switzerland: Springer International Publishing.

Steen, L. A. (Ed.) (2001). *Mathematics and democracy: The case for quantitative literacy*. Princeton, NJ: The Woodrow Wilson National Fellowship Foundation.

Straesser, R. (2015). “Numeracy at work”: A discussion of terms and results from empirical studies. *ZDM – Mathematics Education, 47*(4), 665-674. doi:10.1007/s11858-015-0689-0

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge, U.K. ; New York: Cambridge University Press.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138. doi:10.1007/s10648-010-9128-5

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296. doi:10.1023/a:1022193728205

Thompson, A. G., Philipp, R. A., Thompson, P. W., & Boyd, B. A. (1994). Calculational and conceptual orientations in teaching mathematics. In A. Coxford (Ed.), *1994 Yearbook of the NCTM* (pp. 79-92). Reston, VA: NCTM.

Toner, P. (2011). *Workforce skills and innovation (OECD education working papers).* Retrieved from Paris, France: https://search.oecd.org/innovation/inno/46970941.pdf

Törner, G., Schoenfeld, A. H., & Reiss, K. M. (2007). Problem solving around the world: summing up the state of the art. *ZDM – Mathematics Education, 39*(5-6), 353-353. doi:10.1007/s11858-007-0053-0

Tout, D., & Gal, I. (2015). Perspectives on numeracy: reflections from international assessments. *ZDM – Mathematics Education, 47*(4), 691-706. doi:10.1007/s11858-015-0672-9

Tout, D., & Spithill, J. (2015). The challenges and complexities of writing items to test mathematical literacy. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy: The PISA experience* (pp. 145-171). Basel, Switzerland: Springer International Publishing.

Treffers, A. (1987). *Three dimensions. A model of goal and theory description in mathematics instruction – the Wiskobas Project*. Dordrecht, The Netherlands: Kluwer.

Tsatsaroni, A., & Evans, J. (2014). Adult numeracy and the totally pedagogised society: PIAAC and other international surveys in the context of global educational policy on lifelong learning. *Educational Studies in Mathematics, 87*(2), 167-186. doi:10.1007/s10649-013-9470-x

Van den Heuvel-Panhuizen, M. (Producer). (2000). Mathematics education in the Netherlands: A guided tour. *Freudenthal Institute Cd-rom for ICME9*.

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2014). Realistic mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 521-525). Dordrecht, The Netherlands: Springer.

Van Essen, G., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *Journal of Educational Research, 83*(6), 301-312.

Van Garderen, D. (2006). Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. *Journal of Learning Disabilities, 39*(6), 496-506. doi:10.1177/00222194060390060201

Van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice, 18*(4), 246-254. doi:10.1111/1540-5826.00079

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177. doi:10.1007/s10648-005-3951-0

Van Silfhout, G. (2014). *Fun to read or easy to understand? Establishing effective text features for educational texts on the basis of processing and comprehension research.* (Doctoral thesis), Utrecht University, Utrecht, The Netherlands.

Verschaffel, L., De Corte, E., & Borghart, I. (1997). Pre-service teachers' conceptions and beliefs about the role of real-world knowledge in mathematical modelling of school word problems. *Learning and Instruction, 7*(4), 339-359. doi:10.1016/s0959-4752(97)00008-x

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*(4), 273-294. doi:10.1016/0959-4752(94)90002-7

Verschaffel, L., Depaepe, F., & Van Dooren, W. (2014). Word problems in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 641-645). Dordrecht, The Netherlands: Springer.

Verschaffel, L., Greer, B., & De Corte, E. (Eds.). (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger.

Verschaffel, L., Greer, B., Van Dooren, W., & Mukhopadhyay, S. (Eds.). (2009). *Words and worlds – modelling verbal descriptions of situations*. Rotterdam, The Netherlands: Sense.

Vox - Norwegian Agency for Lifelong Learning. (2013). *Competence goals for basic skills - examples of local curricula in reading, writing, numeracy, digital skills and oral skills*. Oslo, Norway: VOX.

Wake, G. (2014). Making sense of and with mathematics: The interface between academic mathematics and mathematics in practice. *Educational Studies in Mathematics, 86*(2), 271-290. doi:10.1007/s10649-014-9540-8

Wake, G. (2015). Preparing for workplace numeracy: A modelling perspective. *ZDM – Mathematics Education, 47*(4), 675-689. doi:10.1007/s11858-015-0704-5

Watson, A., & Ohtani, M. (Eds.). (2015). *Task design in mathematics education – an ICMI study 22*. Cham, Switzerland: Springer International.

Williams, G., & Clarke, D. J. (1997). The complexity of mathematics tasks. In N. Scott & H. Hollingsworth (Eds.), *Mathematics: Creating the future* (pp. 451-457). Melbourne, Australia: AAMT

Wyndhamn, J., & Säljö, R. (1997). Word problems and mathematical reasoning – a study of children's mastery of reference and meaning in textual realities. *Learning and Instruction, 7*(4), 361-382. doi:10.1016/s0959-4752(97)00009-1

Yackel, E., & Cobb, P. (1995). Classroom sociomathematical norms and intellectual autonomy. In L. Meira & D. Carraher (Eds.), *Proceedings of the nineteenth International conference for the psychology of mathematics education* (Vol. 3, pp. 264-271). Recife, Brazil: Program Committee of the 19th PME conference.

Zaslavsky, C. (1999). *Africa counts: Number and pattern in African cultures - 1999 edition*. Chicago, IL: Lawrence Hill.

Zevenbergen, R. (2000). Ethnography and the situatedness of workplace numeracy. In A. Bessot & J. Ridgway (Eds.), *Education for Mathematics in the Workplace* (pp. 208 - 224). Dordrecht, the Netherlands: Kluwer Academic Publishers

Zevenbergen, R., & Zevenbergen, K. (2009). The numeracies of boatbuilding: New numeracies shaped by workplace technologies. *International Journal of Science and Mathematics Education, 7*(1), 183-206. doi:10.1007/s10763-007-9104-9

Appendices

Appendix A: Expert consultation

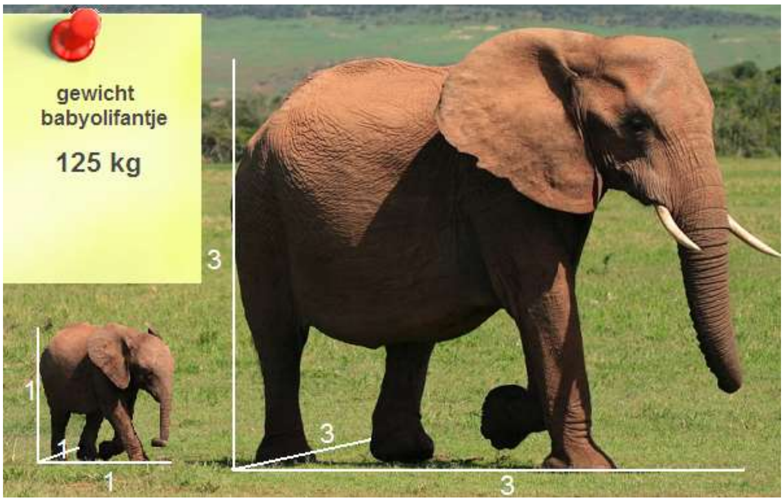
<table>
<tr>
<td>Appels van het ras Elstar worden verkocht in zakken van 1,5 kg.
Eén Elstar weegt gemiddeld 197 gram.

Hoeveel appels zitten er ongeveer in een zak?
[ ] appels</td>
<td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
<tr>
<td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td>
<td></td>
</tr>
<tr>
<td>

Hoeveel appels zitten er ongeveer in een zak?
[ ] appels</td>
<td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
</table>

In de docentenkamer staat een koffieketel met een maximale inhoud van 25 L. In een koffiekopje kan ongeveer 18 cL koffie. Hoeveel kopjes koffie kun je uit een volle koffieketel halen? [ ] kopjes	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 **25 L**  18 cL Hoeveel kopjes koffie kun je uit een volle koffieketel halen? [ ] Kopjes	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Een grote olifant is vergeleken met een babyolifantje 3x zo groot in de breedte, 3x zo groot in de hoogte en 3x zo groot in de lengte. Een babyolifantje weegt ongeveer 125 kilogram. Hoeveel weegt de grote olifant ongeveer? [ ] kg	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel weegt de grote olifant ongeveer? [ ] kg	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr><td>
Per persoon wordt in Nederlandse huishoudens ongeveer 125 liter water per dag gebruikt, vooral voor toilet, douchen en de was.
In 2010 kostte 1 m^3 water € 1,65.
Neem aan dat een gemiddeld gezin uit 3 personen bestaat.

Wat kost het waterverbruik van dit gemiddelde gezin per jaar in 2010?
€ [   ]
</td><td>
A

Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++
</td></tr>
<tr><td>
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want
</td><td></td></tr>
<tr><td>

125 liter water per dag
125 liter water per dag
125 liter water per dag
Water kost in 2010
€ 1,65 per m^3
Wat kost het waterverbruik van dit gemiddelde gezin per jaar in 2010?
€ [   ]
</td><td>
B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++
</td></tr>
</table>

Meneer Kremers woont in Groningen. Hij gaat met de auto naar Maastricht. De afstand van Groningen naar Maastricht is 350 km. De auto verbruikt 1 liter benzine voor elke 16 km. Hoeveel liter benzine gebruikt de auto voor de heen- en terugreis samen? [ ] liter	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel liter benzine gebruikt de auto voor de heen- en terugreis samen? [ ] liter	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr>
<td>Er is een aanbieding van mooi bruin kamerbreed tapijt van zeer stevige kwaliteit. De garantietermijn is 12 maanden en de levertijd is 10 werkdagen. Het tapijt is afgeprijsd van € 19,95 naar € 9,98 per meter. Voor de huiskamer is 8,5 meter tapijt nodig.

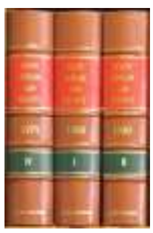
Hoeveel kost de vloerbedekking voor de kamer ongeveer?
€ [ ]</td>
<td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
<tr>
<td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td>
<td></td>
</tr>
<tr>
<td>


Hoeveel kost de vloerbedekking voor de kamer ongeveer?
€ [ ]</td>
<td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
</table>

Op de snelweg kom je een verkeersbord tegen waarop staat dat het nog 39 kilometer is naar Amsterdam. Op deze snelweg mag je maximaal 100 kilometer per uur rijden. Hoeveel minuten duurt de autorit naar Amsterdam met deze snelheid? [ ] minuten	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel minuten duurt de autorit naar Amsterdam met deze snelheid? [ ] minuten	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Je hebt op je bankrekening een saldo(€) van -53,00 Er wordt 75 euro naar deze rekening overgemaakt. Wat wordt het nieuwe saldo? € [ ]	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Wat wordt het nieuwe saldo? € [ ]	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Voor een bibliotheek worden boekenkasten besteld met 6 planken. Deze boekenkasten zijn 210 centimeter hoog en 60 centimeter breed. De boeken die in deze boekenkast moeten staan zijn gemiddeld 4 cm dik. Hoeveel van deze boeken kunnen er ongeveer in deze kast staan? [ ] boeken	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Gemidddeld 4 cm dik  Hoeveel van deze boeken kunnen er ongeveer in deze kast staan? [ ] boeken	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

A

Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++

Een kleurentelevisie met gratis DVD speler is afgeprijsd van € 799,- naar € 719,-
Je kunt de TV ook op afbetaling kopen.
Dan kost het 39 keer € 20,- per maand.

Hoeveel duurder is het als je per maand betaalt?
€ [ ]

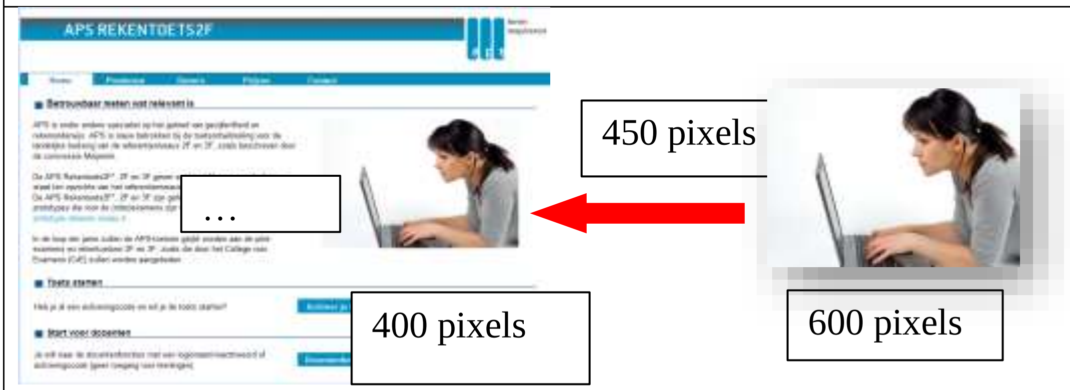
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want

B

Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++

Hoeveel duurder is het als je per maand betaalt?
€ [ ]

Je hebt een foto nodig voor je website. De foto heeft de afmeting van 450 pixels hoog en 600 pixels breed, maar voor je website mogen foto's maximaal 400 pixels breed zijn. Je wilt dat de verhouding tussen hoogte en breedte dezelfde blijft. Hoeveel pixels wordt de hoogte van de foto voor de website? [ ] pixels	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel pixels wordt de hoogte van de foto voor de website? [ ] pixels	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Een kunstwerk bestaat uit 6 lagen. Elke laag bestaat uit een vierkant van 6 bij 6 stalen bollen. Eén bol weegt 5 kilogram. Hoeveel weegt dit kunstwerk? [ ] kg	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
  Hoeveel weegt dit kunstwerk? [ ] kg	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr><td>Je koopt boodschappen voor € 21,30.
Je betaalt met een biljet van 50 euro en
twee munten van een euro.

Hoeveel krijg je terug?
€ [   ]</td><td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
<tr><td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td><td></td></tr>
<tr><td>Je moet betalen:

SUPERMARKT
Daliastraat 4
5707 SJ Helmond
0492-527384
----
15 blik cola 330ml 0.90 13.50
13 chips flav. pnt. light 0.60 7.80
aantal art. 28 subtotaal 21,30
TOTAAL 21.30

Je betaalt met:


Hoeveel krijg je terug?
€ [   ]</td><td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
</table>

Je wilt een Apple iPod nano met een flashgeheugen van 16 GB aanschaffen. Een muzieknummer is ongeveer 4,3 MB Hoeveel muzieknummers kunnen er ongeveer op de iPod? [ ] nummers	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
  Hoeveel muzieknummers kunnen er ongeveer op de iPod? [ ] nummers	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

In een recept voor Honig picknick wraps staat dat je de volgende ingrediënten nodig hebt voor 5 personen: 2 pakken Honig wraps, 250 g verse roomkaas, 1 zakje tuinkruiden, 300 gram fricandeau, groene krulsla en peper en zout naar smaak. Hoeveel gram verse roomkaas heb je nodig voor 12 personen? [ ] gram	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel gram verse roomkaas heb je nodig voor 12 personen? [ ] gram	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Harry ging in voorjaar 2008 op vakantie naar IJsland. In IJsland gebruikt men de IJslandse Kroon (ISK). Tijdens de vakantie gold ongeveer: € 100 = ISK 13 400 en ISK 100 = € 0,74627. Een IJslands tijdschrift kostte ISK 670. Hoeveel euro kost dit tijdschrift uit IJsland? € [ ]	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 **€ 100 = ISK 13 400** **100 ISK = € 0,74627** Hoeveel euro kost dit tijdschrift uit IJsland? € [ ]	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Een herenpak kost normaal € 399,- Er is een speciale aanbieding waarbij geldt dat de AEX-index van vandaag de prijs is die je betaalt. De AEX index is vandaag 342,40. Hoeveel procent korting krijg je als je vandaag dit pak koopt? [ ] %	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel procent korting krijg je als je vandaag dit pak koopt? [ ] %	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr><td>
Het kabinet wil 18 miljard bezuinigen.
Dat zijn 360.000.000 biljetten van 50 euro.
Die biljetten kun je opstapelen tot een hele hoge stapel.
Een stapel van 1000 biljetten van 50 euro zijn ongeveer 11 cm dik.

Hoe hoog wordt de stapel?
[ ] km
</td><td>
A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++
</td></tr>
<tr><td>
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want
</td><td></td></tr>
<tr><td>
Het kabinet wil 18 miljard bezuinigen.
Dat zijn 360.000.000 biljetten van 50 euro.
Die biljetten kun je opstapelen tot een hele hoge stapel.

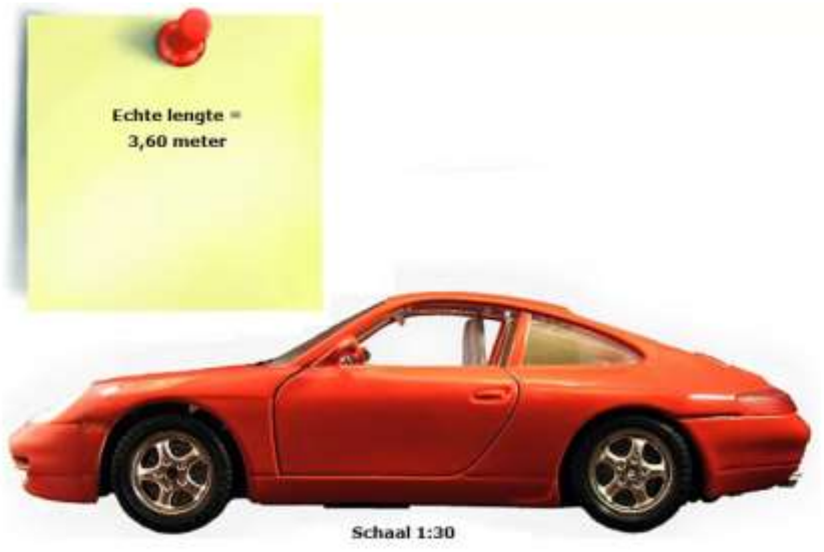

Hoe hoog wordt de stapel?
[ ] km
</td><td>
B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++
</td></tr>
</table>

<table>
<tr>
<td>Met een grondthermometer kun je de bodemtemperatuur meten.
Het bereik van de thermometer gaat van -20 tot +60 graden Celsius.

Hoe groot is het verschil tussen de hoogste en de laagste temperatuur die je met deze thermometer kunt meten?
[ ] graden</td>
<td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
<tr>
<td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

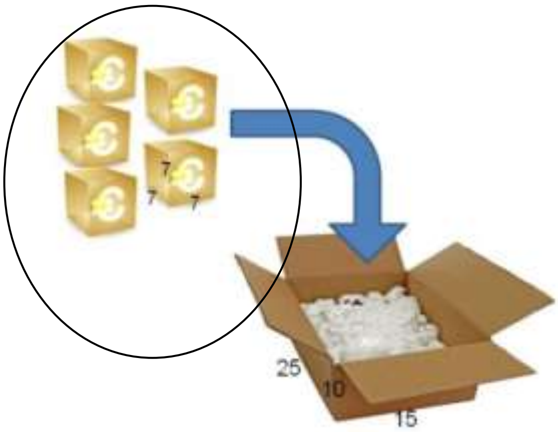
Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td>
<td></td>
</tr>
<tr>
<td>

Hoe groot is het verschil tussen de hoogste en de laagste temperatuur die je met deze thermometer kunt meten?
[ ] graden</td>
<td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
</table>

Van een sportauto wordt een schaalmodel gemaakt. De echte lengte van de auto is 3,60 meter. De schaal die gebruikt wordt is 1: 30. Wat is de lengte van het schaalmodel? [ ] cm	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Wat is de lengte van het schaalmodel? [ ] cm	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr><td>Irene heeft 5 glazen kubussen.
De ribben van deze kubussen zijn 7 cm.
Irene verpakt de kubussen in een kartonnen doos.
De doos is 25 cm lang, 15 cm breed en 10 cm hoog.
Irene wil de kubussen tegen breken beschermen. Daarom vult ze de lege ruimte rondom de kubussen met piepschuim bolletjes.
Bij deze vraag hoef je geen rekening te houden met de dikte van het glas en het karton.

Hoeveel cm^3 piepschuim bolletjes zijn er nodig om de doos af te vullen?
[ ] cm^3</td><td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
<tr><td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td><td></td></tr>
<tr><td>

Hoeveel cm^3 piepschuim bolletjes zijn er nodig om de doos af te vullen?
[ ] cm^3</td><td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
</table>

<table>
<tr><td>Het voorraam in je huis is 1,80 breed en 1,35 m hoog. Je wilt hier dubbelglas in laten zetten.
Dubbelglas kost € 148,- per m^2

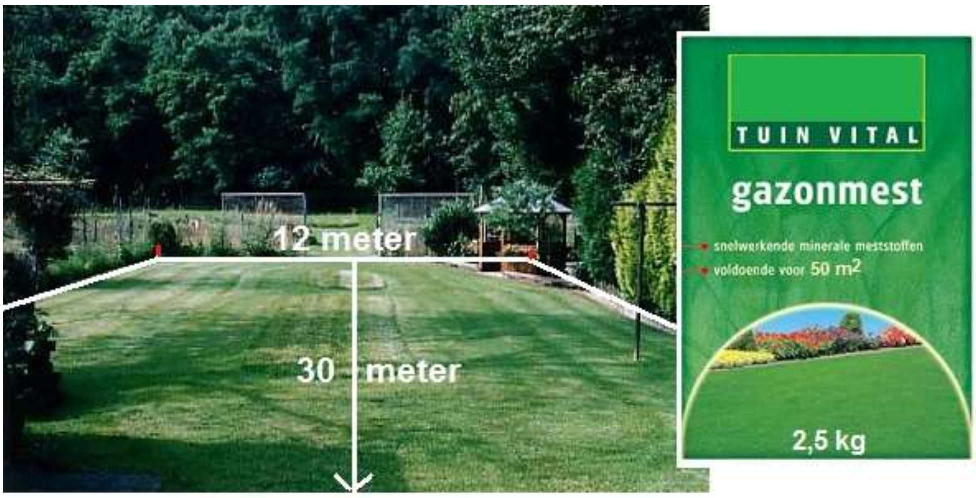
Hoeveel kost het om in dit raam dubbelglas te laten zetten?
€ [ ]</td><td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
<tr><td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td><td></td></tr>
<tr><td>

Hoeveel kost het om in dit raam dubbelglas te laten zetten?
€ [ ]</td><td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
</table>

<table>
<tr><td>Voor een vakantieklus wordt je gevraagd een gazon te bemesten.
Dit gazon is 30 meter lang en 12 meter breed.
Op het pak gazonmest van 2,5 kg dat je hebt gekocht staat dat er snelwerkende minerale meststoffen inzitten en dat het pak voldoende is voor 50 m^2.

Hoeveel kg kunstmest heb je nodig voor dit gazon?
[ ] kg</td><td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
<tr><td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

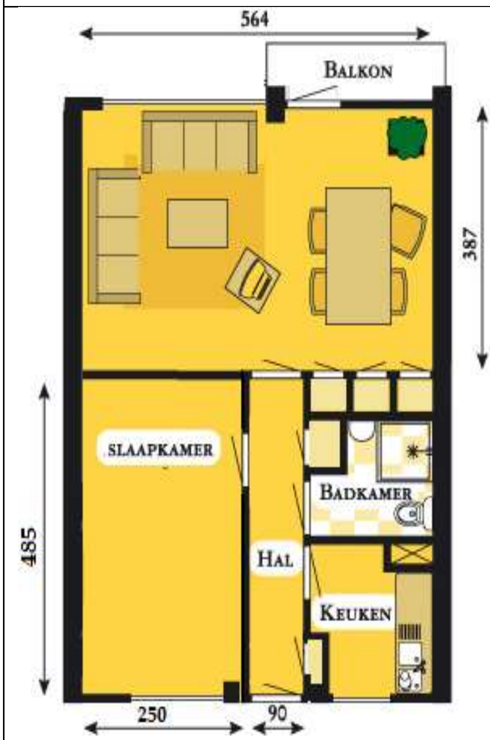
Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td><td></td></tr>
<tr><td>

Hoeveel kg kunstmest heb je nodig voor dit gazon?
[ ] kg</td><td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
</table>

Je gaat de wand van een badkamer betegelen. De afmetingen van de wand zijn 2,60 m bij 5,20 m. Per vierkante meter gebruik je 25 tegels. In een volle doos zitten 50 van deze tegels. Hoeveel hele dozen tegels zijn nodig voor deze wand? [ ] dozen	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel hele dozen tegels zijn nodig voor deze wand? [ ] dozen	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Op je bankrekening heb je een bedrag staan van € 255,72. Je koopt een laptop die is afgeprijsd van € 349,45 tot € 279,-. Wat is het saldo van je rekening na aankoop van deze laptop? € [ ]	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
  Wat is het saldo van je rekening na aankoop van deze laptop? € [ ]	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Een volle watertank op een watertappunt heeft een inhoud van 20 liter. Leerlingen vullen hieruit flesjes van 25 cl. Hoeveel flesjes water kun je uit een volle watertank halen? [ ] flesjes	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
  Hoeveel flesjes water kun je uit een volle watertank halen? [ ] flesjes	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

In een huis heeft de slaapkamer de volgende afmetingen: 485 bij 250. Je gaat de slaapkamer betegelen Tegels van het merk Tosca. Dat zijn 100% polyamide blauwe tegels van 50 bij 50 cm Hoeveel hele vloertegels heb je nodig voor de slaapkamer? [ ] vloertegels	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
  Hoeveel hele vloertegels heb je nodig voor de slaapkamer [ ] vloertegels	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr><td>Je hebt een fles met 0,70 L sterk gif voor onkruidbestrijding.
Voor elke sproeibeurt gebruik je daarvan 15 ml.

Hoeveel sproeibeurten kun je uit de fles halen?
[ ] sproeibeurten</td><td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
<tr><td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td><td></td></tr>
<tr><td>

Voor elke sproeibeurt tegen onkruid
gebruik je 15 ml uit deze fles.

Hoeveel sproeibeurten kun je uit de fles halen?
[ ] sproeibeurten</td><td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td></tr>
</table>

Hoeveel betaal je voor 1650 gram andijvie? € [ ] Op de markt kost de andijvie €1,25 per kilo	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel betaal je voor 1650 gram andijvie? € [ ]	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr>
<td>Je koopt een pak rundergehakt.
Op het etiket staat:
Rundergehakt
kiloprijs € 6,49
498 gr
€ 3,23
Voor 1 hamburger heb je 80 gram gehakt nodig.

Hoeveel hamburgers kun je maken?
[ ] hamburgers</td>
<td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
<tr>
<td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td>
<td></td>
</tr>
<tr>
<td>

Hoeveel hamburgers kun je maken?
[ ] hamburgers</td>
<td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
</table>

Je gebruikt een middel tegen hoestklachten. De hoestdrank zit in flacons van 100 ml. Volwassenen en kinderen vanaf 12 jaar mogen 5 ml per keer, maximaal 5 keer per dag gebruiken. Je gebruikt de maximale dosering. Hoeveel dagen doe je met een flesje? [ ] dagen	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel dagen doe je met een flesje? [ ] dagen	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr>
<td>In de kop van een krantenartikel over de staatsschuld staat dat de staatsschuld in Nederland stijgt met 1.105 euro per seconde.

Hoeveel miljard euro loopt de staatsschuld op in een jaar?
[ ] miljard euro</td>
<td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
<tr>
<td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td>
<td></td>
</tr>
<tr>
<td>
ELSEVIER
Abonneren Nieuwsbrief
NIEUWS OPINIE DOSSIERS STIJL WEEKBLAD SERVICE BIO'S
Rubrieken Laatste 24 uur Partners Vergelijk rente 288,32
Home > Nieuws > Economie > Staatsschuld Nederland stijgt 1.105 euro per seconde
ECONOMIE
Staatsschuld Nederland stijgt 1.105 euro per seconde
vrijdag 24 september 2010 10:50
Waardeer artikel

Hoeveel miljard euro loopt de staatsschuld op in een jaar?
[ ] miljard euro</td>
<td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
</table>

Je doet de catering voor een bruiloft. Je verwacht 150 personen. Bij de thee serveer je bonbons. Je schat dat er gemiddeld 2 bonbons per gast worden gegeten. In een doos bonbons zitten twee lagen van 20 bonbons. Hoeveel dozen moeten er minstens besteld worden om genoeg te hebben voor deze bruiloft? [ ] dozen	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 2 lagen bonbons  Hoeveel dozen moeten er minstens besteld worden om genoeg te hebben voor deze bruiloft? [ ] dozen	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Je rijdt in een Volkswagen Polo met een verbruik van 6,5 liter benzine per 100 km. Oftewel ongeveer 1 liter op 15,4 km. Je rijdt van Groningen naar Luxemburg. Dat is een afstand van 482 kilometer. Bij een benzinestation staat een bord waarop staat: Een liter benzine kost €1,62 Een liter diesel kost € 1,41 Wat gaat deze reis aan benzine ongeveer kosten? € [ ]	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Volkswagen Polo 55 kw Benzine 6,5 liter / 100 km = 1 liter op 15,4 km   Wat gaat deze reis aan benzine ongeveer kosten? € [ ]	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Bij de supermarkt is een aanbieding van 500 gram witlof voor € 0,89. Hoeveel betaal je voor 2,5 kg witlof? € [ ]	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel betaal je voor 2,5 kg witlof? € [ ]	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

A

Voor een project over wereldsteden vindt je de volgende gegevens op internet:

Amsterdam

Oppervlakte 219,33 km^2

Inwoners 783.364 (gemeente)

Sjanghai

Oppervlakte 6340,5 km^2 (31^e)

Inwoners (2010) 23 019 148 (3 2761 km^2)

Hoeveel keer meer inwoners heeft Shanghai dan Amsterdam?

[ ] keer meer

Level
- 0 2F--
- 0 2F –
- 0 2F
- 0 2F+
- 0 2F ++

Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?

O Ja

O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?

O Ja

O Nee, want

B

Hoeveel keer meer inwoners heeft Shanghai dan Amsterdam?

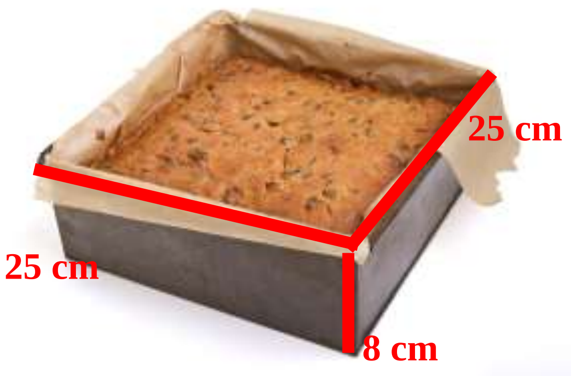
[ ] keer meer

Level
- 0 2F--
- 0 2F –
- 0 2F
- 0 2F+
- 0 2F ++

<table>
<tr>
<td>Een bakblik heeft als afmetingen 25 centimeter lang en 25 centimeter breed. De hoogte is 8 centimeter.

Hoeveel liter beslag heb je nodig om dit blik voor de helft te vullen?
[ ] liter</td>
<td>A

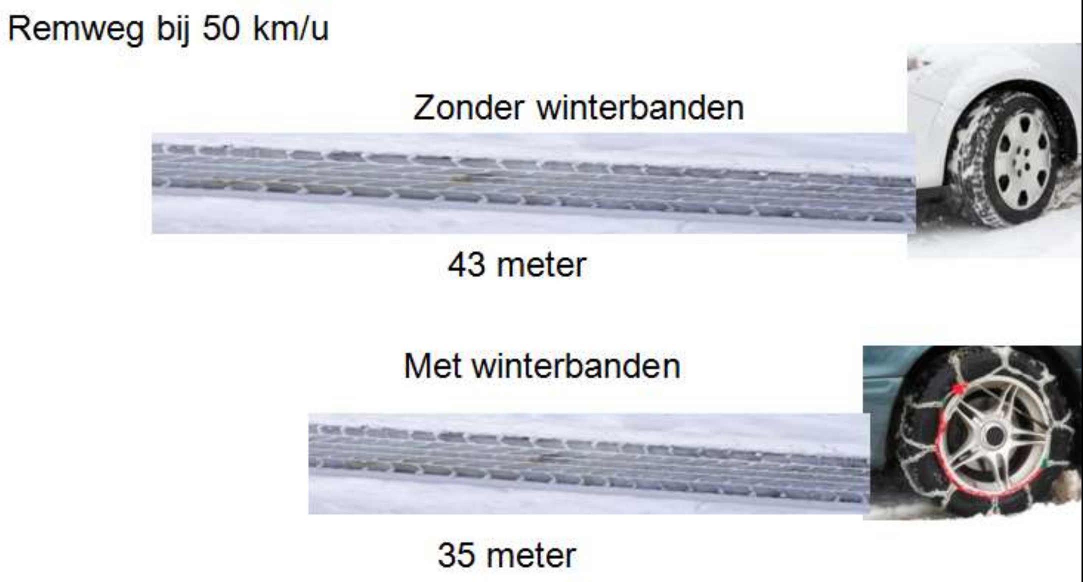
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
<tr>
<td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td>
<td></td>
</tr>
<tr>
<td>

Hoeveel liter beslag heb je nodig om dit blik voor de helft te vullen?
[ ] liter</td>
<td>B

Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
</table>

Je ziet een advertentie voor een polyester zwembad dat je in de tuin kan ingraven. Da afmetingen die genoemd staan in de advertentie zijn: 650 x 300 x 150 De prijs is € 6900,- Hoeveel kubieke meter grond moet je ongeveer uitgraven voor zo'n zwembad? [ ] m^3	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel kubieke meter grond moet je ongeveer uitgraven voor zo'n zwembad? [ ] m^3	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

Met winterbanden onder de auto is de remweg in sneeuw veel korter dan met gewone banden. Bij 50 km/ u is de remweg zonder winterbanden 43 meter en met winterbanden 35 meter. Hoeveel procent korter is de remweg met winterbanden? [ ] %	**A** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++
Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid? O Ja O Nee, want Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau? O Ja O Nee, want	
 Hoeveel procent korter is de remweg met winterbanden? [ ] %	**B** Level 0 2F-- 0 2F – 0 2F 0 2F+ 0 2F ++

<table>
<tr>
<td>Afra ontwerpt verpakkingsmateriaal. Ze heeft een doos voor luxe chocolaatjes gemaakt. De bodem is een vierkant met zijden van 10 cm. De hoogte is 4 cm.

De inhoud van het tweede chocoladedoos is, in vergelijking met de eerste, [   ] zo groot.</td>
<td>A
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
<tr>
<td>Toetsen beide opgaven dezelfde rekenwiskundige kennis en/of vaardigheid?
O Ja
O Nee, want

Toetsen beide opgaven de rekenwiskundige kennis en/of vaardigheid op hetzelfde niveau?
O Ja
O Nee, want</td>
<td></td>
</tr>
<tr>
<td> 
20 bij 20 bij 8 cm
10 bij 10 bij 4 cm

De inhoud van het tweede chocoladedoos is, in vergelijking met de eerste, [   ] zo groot</td>
<td>B
Level
0 2F--
0 2F –
0 2F
0 2F+
0 2F ++</td>
</tr>
</table>

Appendix B: Instrument in Dutch

leren
inspireren

A

Appels van het ras Elstar worden verkocht in tassen van 2,5 kg.
Je weegt één appel en je vindt als gewicht 157 gram.

Hoeveel appels zitten er ongeveer in een tas?
[ ] appels

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel appels zitten er ongeveer in een tas?
[ ] appels

BEWAAR OVERSLAAN

leren
inspireren

A

Meneer Kremers woont in Groningen. Hij gaat met de auto naar Maastricht.
De afstand van Groningen naar Maastricht is 350 km.
De auto verbruikt 1 liter benzine per 16 km.

Hoeveel liter benzine gebruikt de auto voor de heen- en terugreis samen?
[ ] liter

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel liter benzine gebruikt de auto voor de heen- en terugreis samen?
[ ] liter

BEWAAR OVERSLAAN

leren
inspireren

A

Op de snelweg kom je een verkeersbord tegen waarop staat dat het nog 39 kilometer is naar Amsterdam.
Op deze snelweg mag je maximaal 100 kilometer per uur rijden.

Hoeveel minuten duurt de autorit naar Amsterdam met deze snelheid?
minuten

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel minuten duurt de autorit naar Amsterdam met deze snelheid?
minuten

BEWAAR OVERSLAAN

leren
inspireren

A

Een kleurentelevisie met gratis dvd-speler is afgeprijsd naar € 719,-
Je kunt de TV ook op afbetaling kopen. Dan kost het 39 keer € 20,- per maand.

Hoeveel duurder is het als je per maand betaalt?
€

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel duurder is het als je per maand betaalt?
€

BEWAAR OVERSLAAN


leren
inspireren
aps
A
Je koopt boodschappen voor € 21,30.
Je betaalt met een biljet van 50 euro en twee munten van een euro.
Hoeveel krijg je terug?
€
BEWAAR
OVERSLAAN


leren
inspireren
aps
B
Je moet betalen
Je betaalt met
SUPERMARKT
Daliastraat 4
5707 SJ Helmond
0492-527384
15 blik cola 330ml 0.90 13.50
13 chips flav. pnt. light 0.60 7.80
aantal art. 28 subtotaal 21,30
TOTAAL 21.30
50 EURO
Hoeveel krijg je terug?
€
BEWAAR
OVERSLAAN

leren
inspireren

A

Voor een picknick heb je een recept voor wraps gevonden.
In het recept staat dat je het volgende nodig hebt voor 5 personen:
2 pakken Honig wraps, 250 g verse roomkaas
1 zakje tuinkruiden, 300 gram fricandeau
groene krulsla, peper en zout naar smaak.

Hoeveel gram verse roomkaas heb je nodig voor 12 personen?
[ ] gram

BEWAAR | OVERSLAAN

leren
inspireren

B

HONIG

Picknick Wraps

5 personen

Ingrediënten

2 pakken Honig wraps
250 g verse roomkaas
1 zakje tuinkruiden
300 gram fricandeau
groene krulsla
peper zout naar smaak

Hoeveel gram verse roomkaas heb je nodig voor 12 personen?
[ ] gram

BEWAAR | OVERSLAAN

leren
inspireren

A

Harry ging in het voorjaar 2008 op vakantie naar IJsland. In IJsland gebruikt men de IJslandse Kroon (ISK). Tijdens de vakantie gold ongeveer: € 100 = ISK 13 400 en ISK 100 = € 0,74627. Een IJslands tijdschrift kostte ISK 670.

Hoeveel euro kost dit tijdschrift uit IJsland?
€

BEWAAR OVERSLAAN

leren
inspireren

B

In IJsland gebruikt men de IJslandse Kroon (ISK).

€ 100 = ISK 13 400
100 ISK = € 0,74627

Hoeveel euro kost dit tijdschrift uit IJsland?
€

BEWAAR OVERSLAAN

A

Een herenpak kost normaal € 399,-
Er is een speciale aanbieding waarbij geldt dat de AEX-index van vandaag de prijs is die je betaalt.
De AEX-index is vandaag 342,40.

Hoeveel procent korting krijg je als je vandaag dit pak koopt?
%

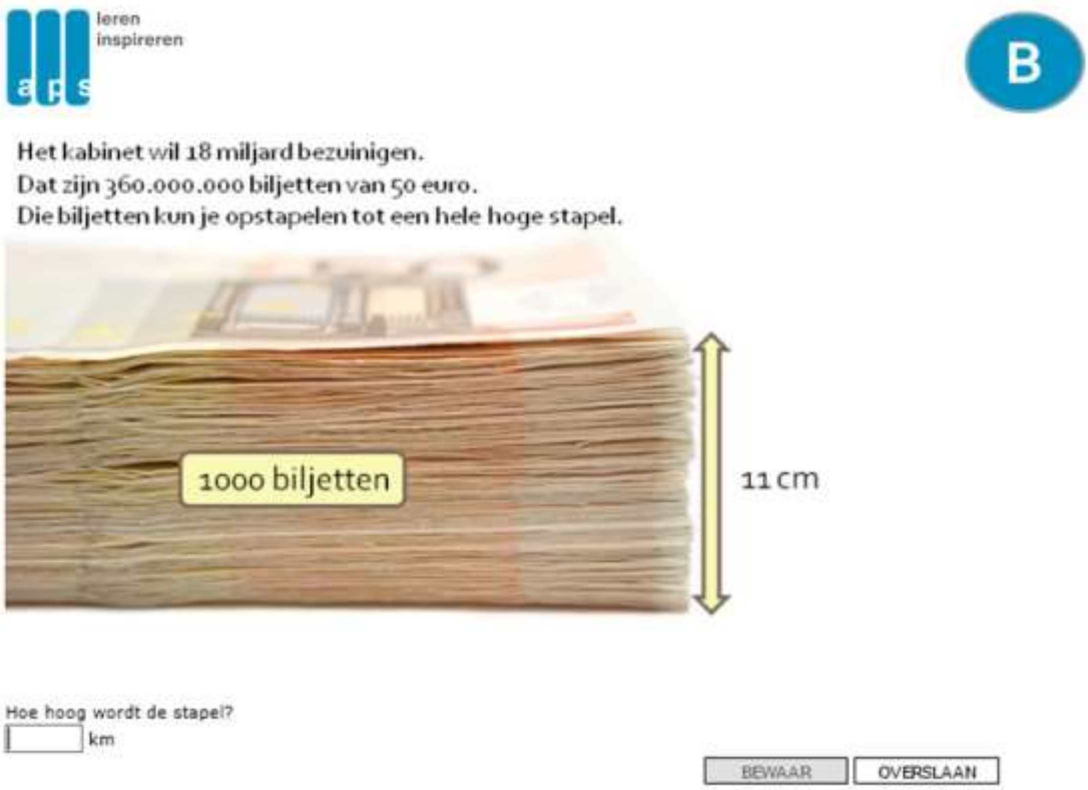
BEWAAR OVERSLAAN

B

Hoeveel procent korting krijg je als je vandaag dit pak koopt?
%

BEWAAR OVERSLAAN


leren
inspireren
aps
A
Het kabinet wil 18 miljard bezuinigen.
Dat zijn 360.000.000 biljetten van 50 euro.
Die biljetten kun je opstapelen tot een hele hoge stapel.
Een stapel van 1000 biljetten van 50 euro is ongeveer 11 cm dik.
Hoe hoog wordt de stapel?
km
BEWAAR
OVERSLAAN


leren
inspireren
aps
B
Het kabinet wil 18 miljard bezuinigen.
Dat zijn 360.000.000 biljetten van 50 euro.
Die biljetten kun je opstapelen tot een hele hoge stapel.
1000 biljetten
11 cm
Hoe hoog wordt de stapel?
km
BEWAAR
OVERSLAAN

leren
inspireren

A

Van een sportauto wordt een schaalmodel gemaakt.
De echte lengte van de auto is 3,60 meter.
De schaal die gebruikt wordt is 1:30.

Wat is de lengte van het schaalmodel?
[ ] cm

BEWAAR | OVERSLAAN

leren
inspireren

B

Wat is de lengte van het schaalmodel?
[ ] cm

BEWAAR | OVERSLAAN

leren
inspireren

A

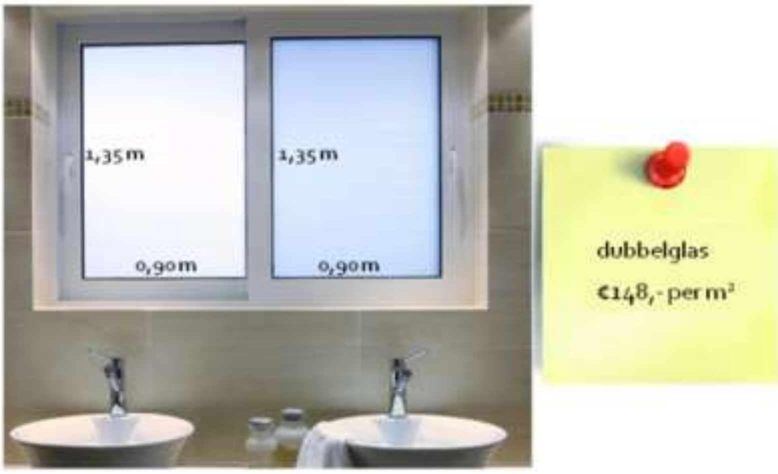
In de badkamer zitten twee ramen. Ze zijn 0,90 m breed en 1,35 m hoog.
Je wilt hier dubbelglas in laten zetten.
Dubbelglas kost € 148,- per m².

Hoeveel kost het om in deze ramen dubbelglas te laten zetten?
€ [ ]

BEWAAR | OVERSLAAN

leren
inspireren

B



Hoeveel kost het om in deze ramen dubbelglas te laten zetten?
€ [ ]

BEWAAR | OVERSLAAN

leren
inspireren

A

Je gaat de wand van een keuken betegelen.
De afmetingen van de wand zijn 2,60 m bij 5,20 m.
Per vierkante meter gebruik je 25 tegels.
Je kunt alleen hele dozen van 50 tegels kopen.

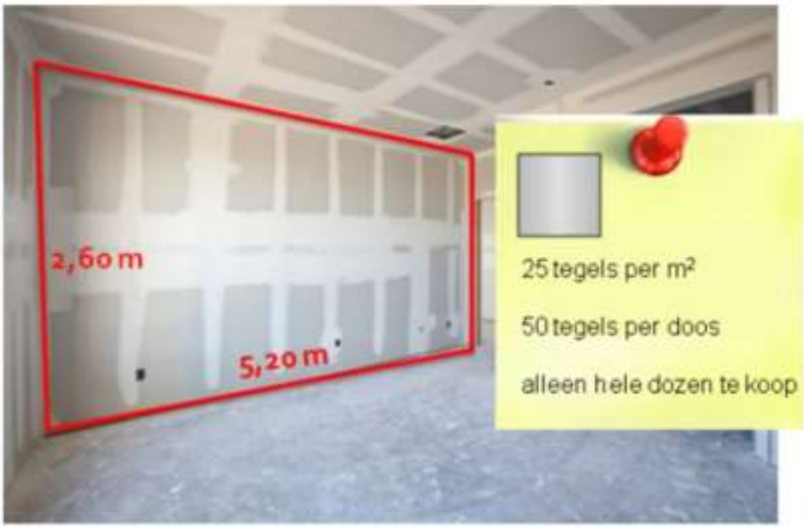
Hoeveel dozen moet je kopen voor deze wand?

[ ] dozen

BEWAAR | OVERSLAAN

leren
inspireren

B



Hoeveel dozen moet je kopen voor deze wand?

[ ] dozen

BEWAAR | OVERSLAAN

leren
inspireren

A

Een volle watertank op een waterkoeler heeft een inhoud van 20 liter.
Leerlingen vullen hieruit flesjes van 25 cl.

Hoeveel flesjes water kun je uit een volle watertank halen?

[ ] flesjes

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel flesjes water kun je uit een volle watertank halen?

[ ] flesjes

BEWAAR OVERSLAAN

leren
inspireren

A

In de plattegrond van je nieuwe huis staan bij de slaapkamer
de volgende afmetingen: 485 cm bij 250 cm.
Je gaat de slaapkamer betegelen met tapijttegels van het merk Tosca.
Dat zijn tegels van 50 bij 50 cm.

Hoeveel hele vloertegels heb je nodig voor de slaapkamer?
[ ] vloertegels

BEWAAR OVERSLAAN

leren
inspireren

B

Tegel Tosca

100% polyamide
blauw
50x50 cm
Art. nr: 20200111

Hoeveel hele vloertegels heb je nodig voor de slaapkamer?
[ ] vloertegels

BEWAAR OVERSLAAN

leren
inspireren

A

Op de markt kost de andijvie €1,25 per kilo.

Hoeveel betaal je voor 1650 gram andijvie?

€

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel betaal je voor 1650 gram andijvie?

€

BEWAAR OVERSLAAN

leren
inspireren

A

Je koopt een pak rundergehakt van 498 gram.
Daarvoor betaal je € 3,23.
Voor 1 hamburger heb je 80 gram gehakt nodig.

Hoeveel hamburgers kun je maken?
[ ] hamburgers

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel hamburgers kun je maken?
[ ] hamburgers

BEWAAR OVERSLAAN

leren
inspireren

A

Je gebruikt een middel tegen hoestklachten. De hoestdrank zit in flacons van 100 ml.
Volwassenen en kinderen vanaf 12 jaar mogen 5 ml per keer, maximaal 5 keer per dag gebruiken. Je gebruikt de maximale dosering.

Hoeveel dagen doe je met een flesje?
[ ] dagen

BEWAAR OVERSLAAN

leren
inspireren

B

Hoeveel dagen doe je met een flesje?
[ ] dagen

BEWAAR OVERSLAAN

leren
inspireren

A

In de kop van een krantenartikel staat dat de staatsschuld in Nederland stijgt met 1.105 euro per seconde.

Hoeveel miljard euro stijgt de staatsschuld in een jaar?

[ ] miljard euro

BEWAAR | OVERSLAAN

leren
inspireren

B

ELSEVIER

Abonneren Nieuwsbrief

NIEUWS | OPINIE | DOSSIERS | STIJL | WEEKBLAD | SERVICE | BIO'S

Rubrieken | Laatste 24 uur | Partners | Vergelijk rente | 288,32

Home > Nieuws > Economie > Staatsschuld Nederland stijgt 1.105 euro per seconde

ECONOMIE

Staatsschuld Nederland stijgt 1.105 euro per seconde

vrijdag 24 september 2010 10:50

Waardeer artikel

De staatsschuld in Nederland blijft maar stijgen en bedraagt eind 2010 ruim 382 miljard euro. Om dat in perspectief te plaatsen: iedere seconde loopt de schuld op met 1.105 euro.

Hoeveel miljard euro stijgt de staatsschuld in een jaar?

[ ] miljard euro

BEWAAR | OVERSLAAN

leren
inspireren

A

Een bakblik heeft als afmetingen 25 centimeter lang en 25 centimeter breed.
De hoogte is 8 centimeter.

Hoeveel liter beslag heb je nodig om dit blik voor de helft te vullen?
[ ] liter

BEWAAR | OVERSLAAN

leren
inspireren

B

Hoeveel liter beslag heb je nodig om dit blik voor de helft te vullen?
[ ] liter

BEWAAR | OVERSLAAN

leren
inspireren
aps

A

De remweg van een auto met winterbanden bij remmen in de sneeuw met een snelheid van 50 km/uur is 35 meter. Terwijl een auto met zomerbanden in die situatie maar liefst 43 meter nodig heeft om tot stilstand te komen.

Hoeveel procent korter is de remweg met winterbanden?
[   ] %

BEWAAR OVERSLAAN

leren
inspireren
aps

B

Hoeveel procent korter is de remweg met winterbanden?
[   ] %

BEWAAR OVERSLAAN

leren
inspireren

A

Afra ontwerpt verpakkingsmateriaal. Ze heeft een doos voor luxe chocolaatjes gemaakt.
De bodem is een vierkant met zijden van 10 cm. De hoogte is 4 cm.

De fabrikant vraagt haar net zo'n doos te maken, maar nu worden de zijden van het vierkant 20 cm en de hoogte is 8 cm.

De inhoud van de tweede chooladedoos is, in vergelijking met de eerste,
[ ] zo groot.

BEWAAR OVERSLAAN

leren
inspireren

B

De inhoud van de tweede chooladedoos is, in vergelijking met de eerste,
[ ] zo groot.

BEWAAR OVERSLAAN

leren
inspireren

A = B

Persoon	Kosten
Kind 1-3 jaar	€ 2,60
Kind 4-8 jaar	€ 3,78
Kind 9-13 jaar	€ 4,90
Man 14-65 jaar	€ 6,35
Vrouw 14-65 jaar	€ 5,81
Man en vrouw 65+	€ 5,63

Tabel: dagelijkse kosten voor voeding per persoon

Een huishouden bestaat uit:
- moeder van 31 jaar
- vader van 33 jaar
- kind van 12 jaar

Wat zijn de kosten voor voeding per dag voor dit huishouden?
€

BEWAAR OVERSLAAN

leren
inspireren

A = B

Wat is de waterstand in Texel om 8 uur 's avonds ongeveer?
cm

BEWAAR OVERSLAAN

leren
inspireren

Afstandentabel China

	Beijing	Chengdu	Dali	Guangzhou	Guilin	Hangzhou	Hong Kong	Kunming
Beijing		1697	2522	1967	1887	1200	2394	2266
Chengdu	1697		885	1390	975	1699	1907	711
Dali	2522	885		1632	1134	2600	1710	380
Guangzhou	1967	1390	1632		452	1099	135	1357
Guilin	1887	975	1134	452		1230	566	859
Hangzhou	1200	1699	2600	1099	1230		1092	2089
Hong Kong	2394	1907	1710	135	566	1092		1204
Kunming	2266	711	380	1357	859	2089	1204	

Afstanden zijn in kilometers en bij benadering

Wat is de afstand tussen Guilin en Hongkong?

km

BEWAAR OVERSLAAN

Appendix C: Instrument translated into English

1A

Apples are sold in bags of 2.5 kilograms. You weigh one apple and find it weighs 157 grams.

About how many apples are there in the bag?

☐ **apples**

1B

About how many apples are there in the bag?

☐ **apples**

2A

Mr. Smith lives in the city of Groningen.
He drives by car to the city of Maastricht.
The distance between Groningen and Maastricht is 350 kilometres.
The usage of the car is 1 litre petrol per 16 kilometres.

How many litres of petrol does a roundtrip take?

☐ **litres**

2B

How many litres of petrol does a roundtrip take?

☐ **litres**

3A

On the highway you see a road sign which says it is still 39 kilometres to Amsterdam.
On this highway the maximum speed is 100 kilometres per hour.

How many minutes will it take to reach Amsterdam with this speed?

☐ **minutes**

3B

How many minutes will it take to reach Amsterdam with this speed?

☐ **minutes**

4A

A color TV with free DVD player is marked down to € 719.-
You can pay the TV in installments.
Then you have to pay 39 times € 20.- a month.

How much more do you have pay if you pay in monthly installments?
€ ____

4B

How much more do you have pay if you pay in monthly installments?
€ ____

5A

You buy groceries for a total of € 21.30
You pay with a 50 euro bill and two coins of 1 euro.

What is the change?
€ ____

5B

You have to pay:

SUPERMARKET
Dahliastreet 5
Helmond

15 cans cola 330 ml	0.90	13.50
13 patato chips light	0.60	7.80
Number art. 13	sub total	21.30
Total		**21.30**

You pay with:

What is the change?
€ ____

6A

For a picnic you found a recipe for wraps. The recipe gives the ingredients for 5 persons: 2 packs of wraps, 250 grams of cream cheese, 1 sachet of garden herbs, 300 grams of fricandeau, green lettuce, pepper and salt to taste

How much cream cheese do you need for 12 persons?
____ **grams**

6B

How much cream cheese do you need for 12 persons?
____ **grams**

7A

Harry went in the spring of 2008 on a holiday trip to Iceland. In Iceland one pays with Icelandic Krona (ISK).
During the holiday trip the exchange rate was: € 100 = ISK 13 400 and
ISK 100 = € 0,74627.
An Icelandic magazine costs 670 ISK.

What is the price for this magazine from Iceland?
€ ☐

7B

In Iceland one pays with Iceland Krona (ISK).

€ 100 = ISK 13 400
100 ISK = € 0,74627

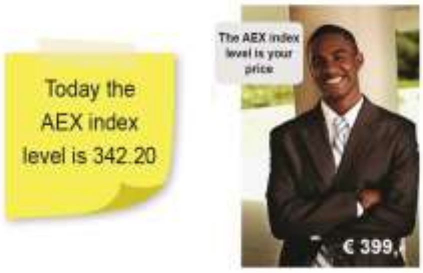
What is the price for this magazine from Iceland?
€ ☐

8A

The regular price for a men's suit is € 399.-
There is a special discount where you pay as the price the actual level of the Dutch stock market index (AEX).
The AEX index level today is 342.40.

What percentage is the discount if you buy the suit today?
☐ **%**

8B



What percentage is the discount if you buy the suit today?
☐ **%**

9A

The Dutch Government want to cut expenses by € 18 billion euro.
That are 360,000,000 bills of 50 euro.
Those bills can be piled up in a huge pile.
A pile of 1000 bills of 50 euro is about 11 cm in height.

What will be the height of the total pile?
☐ **km**

9B

The Dutch Government want to cut expenses by € 18 billion euro.
That are 360,000,000 bills of 50 euro.
Those bills can be piled up in a huge pile.

What will be the height of the total pile?
☐ **km**

10A

A sports car is made into a scale model.
The real length of the car is 3.60 metres.
The scale used is 1 : 30.

What is the length of the scale model?
☐ **cm**

10B

What is the length of the scale model?
☐ **cm**

11A

The bathroom has two windows.
They are both 0,90 m in width and
1,35 m in height.
You want to double glaze these windows.
Double glazing costs € 148,- per m^2

What is the cost of double glazing these windows?
€ ☐

11B

What is the cost of double glazing these windows?
€ ☐

12A

You have to tile a kitchen wall.
The wall measures 2.60 m by 5.20 m.
You use 25 tiles per square metre.
You can only buy boxes of 50 tiles.

How many boxes do you have to buy to tile this wall?
☐ **boxes**

12B

How many boxes do you have to buy to tile this wall?
☐ **boxes**

13A

A full water tank on a water dispenser contains 20 litres.
Students from the dispenser fill their 25 cl water bottles.

How many water bottles can be filled from a full water tank?
☐ **bottles**

13B

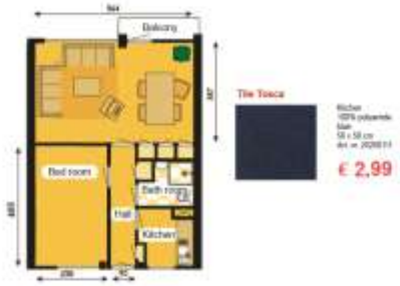
How many water bottles can be filled from a full water tank?
☐ **bottles**

14A

In the floorplan of your new house the size of the bedroom is 485 cm by 250 cm.
You are furnishing the bedroom with carpet tiles of the brand Tosca.
These tiles are 50 by 50 cm.

How many complete tiles do you need for the bedroom?
☐ **tiles**

14B



How many complete tiles do you need for the bedroom?
☐ **tiles**

15A

On the market the price of endive is € 1.25 per kilo.

What do you have to pay for 1650 grams of endive?
€ ☐

15B

What do you have to pay for 1650 grams of endive?
€ ☐

16A

You buy a pack of minced meat weighing 498 grams. The price is € 3.23.
For 1 hamburger you need 80 grams of minced meat.

How many hamburgers can you make?
☐ **hamburgers**

16B

How many hamburgers can you make?
☐ **hamburgers**

17A

You are using a medicine against coughing. The cough drops are sold in flasks of 100 ml. Adults and children of 12+ years old are allowed to take 5 ml as a dosage, with a maximum of 5 dosages per day. You use the maximum dosage.

How many days will one flask of cough drops suffice?
☐ **days**

17B

How many days will one flask of cough drops suffice?
☐ **days**

18A

A recent newspaper headline said that the public debt of the Netherlands increases with 1,105 euros per second.

How many billion euros the Dutch public debt increases in one year?
☐ **billion euros**

18B

How many billion euros the Dutch public debt increases in one year?
☐ **billion euros**

19A

A cake tin is 25 centimetres long and 25 centimetres wide. The height is 8 centimetres.

How many litres of cake mix do you need to fill this cake tin half full?

☐ **liter**

19B

How many litres of cake mix do you need to fill this cake tin half full?

☐ **liter**

20A

The braking distance of a car with snow tires in snowy conditions is about 35 metres from a speed of 50 km/h. A car with regular tires has in these conditions a braking distance of no less than 43 metres.

How much percent is the breaking distance smaller with snow tires?

☐ **%**

20B

Breaking distance from a speed of 50 km/h

With snow tires
35 metres

With regular tires
43 metres

How much percent is the breaking distance smaller with snow tires?

☐ **%**

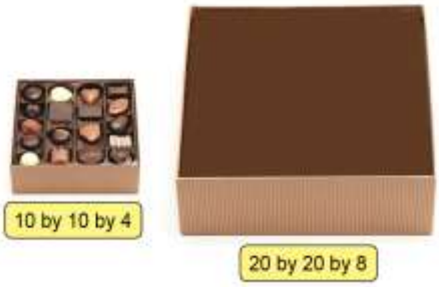
21A

Afra designs packaging materials.
She made a box for luxury chocolates.
The bottom is a square with sides of 10 cm.
The height is 4 cm.
The manufacturer asks her to design a similar box, but with as a bottom a square with sides of 20 cm and a height of 8 cm.

The volume of the second chocolate box is, in comparison to the first,

☐ **times as big.**

21B



The volume of the second chocolate box is, in comparison to the first,

☐ **times as big.**

22

What is the water level in Texel at 8 PM approximately? ☐ cm

23

Person	Costs
Child 1-3 years old	€ 2.60
Child 4-8 years old	€ 3.78
Child 9-13 years old	€ 4.90
Men 14- 65 years old	€ 6.35
Women 14- 65 years old	€ 5.81
Men & women 65+	€ 5.63

Daily cost of food per person

A household consists of:
mother 31 years old
father 33 years old
child 12 years old

What are the daily costs of food for this household? € ☐

24

Distance table China in km

What is the distance between Guilin and Hongkong? ☐ km

Summary

Images of Numeracy: Investigating the effects of visual representations of problem situations in contextual mathematical problem solving

Problem statement and research question

In this dissertation the following research question was addressed:

> In presenting contextual mathematical problems, what is the effect on student performance of changing a descriptive representation of the problem situation to a mainly depictive one?

Investigating this research question is likely to shed light on an important issue in mathematics education: why do students on so many occasions show a suspension of sense-making in answering contextual mathematical problems, and as result perform poorly, and develop inadequate problem-solving skills?

We argued that a mainly depictive representation of the problem situation, especially with photographs from real problem situations, might increase the likelihood that students address the problem with a problem-solving attitude, that is, taking into account all aspects and constraints that the problem situation demands. This could counteract or prevent the suspension of sense-making to some extent, with better performance as a result. And although there have been many studies on the issue of suspension of sense-making, we identified a knowledge gap in terms of the effect of specifically this kind of change in the representation of the problem situation on students' performance.

We found that our approach to investigating the effects on students' performance of a mainly depictive representation of the problem situation was under-researched in mathematics education. There is a very limited amount of literature on systematic comparisons of descriptive and depictive representations of problem situations in mathematical problem solving and the effect on students' performance.

The main steps on the road to answering the research question were the following. Firstly, in discussing the conceptual framework we laid out the chain of reasoning that led from the mentioned issue – the observed suspension of sense-making – to the research question. Secondly, we described how the question was addressed through a series of four studies; what the results were of the conducted studies; and how these results answered the research question. Thirdly, we argued how the results from our studies can be of relevance for the practice of mathematics education, and for the field of research in mathematics education.

Conceptual framework and chain of reasoning

Over the past few decades the focus in mathematics education has been shifting from a strong emphasis on mathematical procedures towards more emphasis on applying mathematics, modelling, problem solving, and fostering a problem-solving attitude in students. Current curricula and textbooks show more references to real-world phenomena and problems than 30 years ago (Blum et al., 2007). This is in particular true for the domains of mathematics education that deal with numeracy, mathematical literacy, and quantitative literacy. In these particular branches of mathematics education the central theme is how people can learn to use mathematics in order to cope with quantitative problems from daily life (Kilpatrick, 1996; Niss, 1996; OECD, 1999; Toner, 2011).

Contextual mathematical problems play an important role in mathematics education that focuses on modelling and problem solving. A typical contextual mathematical problem consists of a problem situation that is presented and a question that is raised. In most cases of contextual mathematical problems, the problem situation is a representation of a situation from real life, while the question aims at activating the student to answer the question by using mathematical reasoning, knowledge, and tools. The dominant form of representing a problem situation, in all countries and cultures, is to describe the problem situation in words. This genre of “word problems” (Lave, 1992; Verschaffel et al., 2000; Verschaffel et al., 2009) can be witnessed worldwide wherever mathematics education pursues problem-solving goals.

However, a serious issue in mathematics education is the way students deal with word problems. Probably as a consequence of the procedural emphasis in mathematics education, many students persistently see word problems as procedural exercises in disguise, and act accordingly; they try to extract as quickly as possible some numbers from the problem situation, upon which they perform an operation, without making sense of the problem and without taking realistic constraints into consideration. In the literature many such difficulties are reported in using contextual mathematical problems in general and word problems in particular (DeFranco & Curcio, 1997; Dewolf et al., 2014; Gellert & Jablonka, 2009; Inoue, 2009; Reusser & Stebler, 1997; Roth, 2009; Verschaffel et al., 1997; Verschaffel et al., 1994). The difficulties become poignantly visible in the many inadequate or literally nonsensical solutions that can be witnessed in students' work. This so-called suspension of sense-making by students (Schoenfeld, 1991; Verschaffel et al., 2000) is considered a serious issue in mathematics education. It is seen as a barrier to acquiring an adequate problem-solving mindset, as sense-making is considered crucial in an adequate problem solving process (Schoenfeld, 1992). An adequate problem-solving process starts with interpreting the problem situation, and students are expected to construct a model of the problem posed, in order to come to answers that make sense with respect to a solution to the problem (Schnotz, 2002).

As a countermeasure to the suspension of sense-making many mathematics educators suggest providing more authentic representations of the problems. This arguably improves the chances that students will stay in a problem-solving mindset, resulting in better performance. In our research we decided to pursue and investigate a specific countermeasure within this line of reasoning, that is, changing the problem situation from descriptive – as in word problems – to mainly depictive (Schnotz et al., 2010). For the depiction we used photographs from real situations (see Figure s.1). We argued that such a measure allows students to experience more authenticity in the problem, and hence increasing the chance of coming to an appropriate solution – in other words, decreasing the chance of suspension of sense-making.

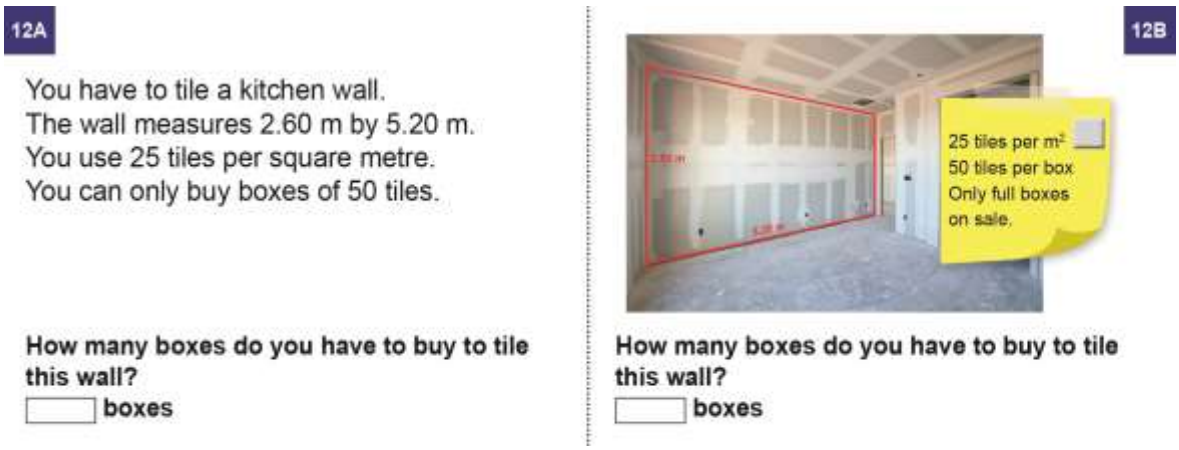


Fig. s.1 An example of an item in two versions: descriptive and depictive.

The theoretical framework for the research described in this dissertation combines several perspectives. Firstly, we made use of concepts like mathematical literacy (Jablonka, 2003, 2015; OECD, 2012), and comparable concepts like numeracy (Gal et al., 2003; Tout & Gal, 2015), quantitative literacy (Madison & Steen, 2008; Steen, 2001), matheracy (D'Ambrosio, 1997), and functional mathematics (Forman & Steen, 1999). Secondly, we used the extensive body of knowledge on problem solving, modelling, and word problems, and the difficulties word problems are said to trigger in mathematics education (Blum et al., 2007; Galbraith & Stillman, 2006; Verschaffel et al., 2000). Thirdly, we were informed by perspectives from cognitive psychology on mental representations and the use of visual elements for learning (Schnotz, 2005; Schnotz et al., 2010; Schnotz & Bannert, 2003; Schnotz & Kürschner, 2008). These perspectives are described in Chapters 1, 3, and 4.

Method and findings

This dissertation contains reports on four empirical studies that were conducted to find an answer to the research question: an exploratory study (Chapter 2); the design and validation of an instrument to measure students' performance on numeracy tasks (Chapter 3); a large-scale randomised controlled trial (Chapters 4 and 5); and a replication study (Chapter 6). These studies provided results that constituted the building blocks for answering the research question. The interview data from the exploratory study indicated what kinds of mathematical representations were making sense for students.

This knowledge was used to design image-rich, relevant, and appropriate numeracy problems as alternatives to word problems. The designed problems were subsequently used in the design and validation of an instrument that could measure the differences in students' performance when the representation of the problem situation of the numeracy tasks was changed from descriptive to mainly depictive. The instrument was used in a large-scale randomised controlled trial with students aged 11–20 years, and was replicated with adults as a target group.

We brought together the most important findings and conclusions of our studies to answer the research question on the effect of changing the representation of the problem situation. Moreover, we wanted to investigate whether our conjecture that students would perform better on contextual mathematical problems in which the problem situation is represented in a mainly depictive way, would hold true.

The first study was an exploratory study on the numerate world of vocational students (Chapter 2). After a series of semi-structured interviews with students, so-called "numeracy incidents" were identified in the transcription of the interviews. These numeracy incidents were categorised according to two major numeracy frameworks: the international PISA Mathematics Framework (OECD, 2013c); and the Dutch Literacy and Numeracy Framework (Ministerie van OCW, 2009). The findings showed that there was a gap between the numerate world as expressed by the interviewed students and the representation of the numerate reality in the most common numeracy assessment tasks of these students. Moreover, when interviewed on work related situations, in particular on artefacts the students produced themselves, we found they showed better understanding of mathematical concepts when the produced artefacts were at hand. This encouraged us to look for authentic visual representations of problem situations, for instance, by using photographs or real situations likely to arise in students' lives (see Appendix A for examples).

The second study entailed the design and validation of an instrument that could measure the differences in students' performance when changing the representation of the problem situation (Chapter 3). In the design of the instrument the results of the exploratory study were used to develop the representations of problem situations, which were closer to real-life situations (e.g., by using photographs in representing the problem situations), thereby increasing the perceived authenticity for the students.

For the validation of the instrument three steps were taken. First, alternative items for existing word problems were designed and referenced with the content domains of the LaNF. Second, two panels of experts from the field of mathematics and numeracy education were asked to comment on the problems used in the instrument, and to give their expert opinions on the degree of equivalence of the selected word problems as compared with the alternative image-rich numeracy problems. Third, a test run (N = 7,454) with a tentative instrument was held to see whether the condition of a controlled randomised trial could be met. In this way most threats to validity – content validity, construct validity, criterion validity, reliability, and feasibility – were countered. These activities around validations were carried out successfully, so that the final instrument could be used to measure validly the effect on students' performance of changing the representation of the problem situation. After the design of the instrument, it was programmed as an online test.

The actual trial with the instrument was held in a four-week period (see Chapter 4). The instrument was made available by internet for students aged 10-to 20-years, to be used as a numeracy test. In total, 31,842 students from 179 schools geographically spread across the Netherlands, participated in the test. To test the prediction that students scored higher on depictive representation, statistical tests were carried out (Chapter 4). Furthermore, to get more in-depth insight into the effects of the background variables on the results and their possible interaction effect with the manipulated variable, a probit model (a limited dependent variable model) was used. In the analyses the overall result was a 1.9 percentage points rise in students' performance when the representation was changed to a depictive problem situation. This difference was statistically significant ($p < .001$) with a very small effect size of Cohen's $d = 0.09$. So, although our prediction had been supported, this statement should be regarded with great caution given the small change and the small effect size. However, to put the findings in perspective, it is noteworthy that according to Slavin (2016) in educational studies using a randomised controlled trial, rarely effect sizes are found over 0.2, and that with research populations over 5000 participants valid inferences can be made with effect sizes as low as 0.1. Hence, following Slavin (2016) inferences seem justified in our case.

We further investigated the interdependent effect of selected background variables on the overall effect. The background variables, such as gender, ethnicity, educational track, grade and age, and maths grade, were taken as cross terms with the variable version in the statistical model. It appeared that most variables other than the version of the representation of the problem situation, had no significant effect on the difference in performance, with the exception of gender. This means that the measured effect was not dependent on the background variable, except for gender. For male participants the difference in scores on both versions of the problems was slightly higher than for female participants.

To get more insights in the relation between task characteristics and the overall results found in the trial, we conducted an in-depth statistical analysis (Chapter 5). Firstly, two variables were introduced to characterise the tasks: wordiness for the A-versions of the problems; and the number of pictorial elements for the B-version of the problems. From the statistical analysis we found a very tentative measure on the magnitude of the increase in students' performance for each word reduced, also on the magnitude of the decrease in students' performance for each depictive element added. The latter result indicated that adding more depictive elements had no positive effect on students’ performance.

Secondly, we analysed the effect of task characteristics regarding the domain of the tasks on students’ performances. We expected from theory (Schnotz et al., 2010) that probably for the domain of measurement & geometry the effect of changing the representation of the problem situation could have a bigger effect than for tasks that belong to the domains of numbers and proportions, since the depictive elements in geometrical and measurement problems in many cases have a strong relation with the mathematical model of the problem. This expectation was corroborated by our analysis. We looked at the interdependent effects of the variables regarding the content domains and the variable regarding version, and we found a significant and negative effect of the cross terms. We found that the effect of changing the representation of the problem situation had the least effect in the domain of proportions, and the most effect in the domain of measurement & geometry. The effect in the domain of numbers was in between these effects, but closer to the effect of the domain of proportions. The effect on tasks in the domain of measurement & geometry was in line with our expectations.

The study described above focused on students aged 11 to 20 years. We wondered whether similar results could be observed with adults. Arguably, adults could be less inclined to the suspension of sense-making that is so often related to school mathematics. We conducted a replication study (Chapter 6) with the developed instrument with 420 adult participants in the Groot Nationaal Rekenonderzoek [Great National Numeracy Survey], a research initiative by the public broadcasting organisations VPRO and NTR, supported by the Netherlands Organisation for Scientific Research (NWO). The overall result in this study with adult participants was 1.1 percentage point higher scores on the problems with a depictive representation of the problem situation. This was in line with the overall results we found in the large school trial (1.9 percentage point higher scores on image-rich numeracy problems). We observed similarities and differences in scores when we compared the item scores in both trials. We found four tasks that in both trials showed a statistically significant better performance for the image-rich versions. Three of them are from the domain of measurement & geometry. This finding corroborates our earlier findings that with tasks from the domain of measurement & geometry the change in representation of the problem situation has the greatest positive influence on the performance of the participants.

To conclude, we found an overall positive result on students' performances of changing the representation of the problem situation to mainly depictive (Chapter 7). We were able to exclude most of the other variables as possible interdependent variables. Due to the large size of the sample inferences seem justified despite the small effect size. To establish whether we have a robust and general effect of changing the representation of the problem situation in contextual mathematics problems on students' performance, more replications or studies with larger item pools should be conducted. For that purpose, the instrument was translated into English and made available through open access which gives the opportunity for replication in English-speaking countries. Replications can be held with other and more items, for instance, to test specifically whether the effect holds true within the various content domains of the items.

After synthesising the conclusions, our answer to the research question is that changing the problem statement from descriptive to mainly depictive representations (by using photographs) improves the chance that students will correctly solve the stated problem. This is an indication that the chance of suspension of sense-making is likely to be reduced to some extent.

The changes in the aforementioned chances were small and depended to some extent on the content domain of the tasks. Limitations of the studies were identified, and hence the results of this study should be interpreted with care, that is, considering their limitations. Nevertheless, our findings add to the body of knowledge on contextual mathematical problem solving and sense-making. As far as we could ascertain, this is the first study that systematically compared different representations of the problem situation in contextual mathematical problem solving on this scale.

Relevance of the studies

General relevance

This dissertation contributes to the worldwide trend in mathematics education to emphasise problem solving. In investigating and answering the research question our aim was to shed more light on the effect of various kinds of representations of problem situations from real life on students' performances. More specifically, the reported studies addressed the intricacies of the genre of word problems and compared students' performance on word problems with their performance on tasks, in which the problem situation was represented with photographs of real problem situations. The results could be of interest for students, who could perform better and with less suspension of sense-making; for teachers and textbook authors, as it provokes reflections on how word problems – almost habitually – are used to present contextual mathematical problem; and for policy makers, since contextual mathematical problems are used increasingly for high stakes testing in many countries and in comparing countries in international comparative studies like PISA and PIAAC.

Relevance for future research

Future research in better understanding the actual behaviour of students in contextual mathematical problem solving, and the way they make sense of the problem situation, can use the findings from the reported studies.

The effects of habituation. Countries differ in the content of their mathematics curricula, on the extent in which numeracy tasks are used, and on the degree of authenticity in the contexts used. From the conjecture that students who are more used to depictive problem situations are less influenced by a change from descriptive to depictive representations, a follow-up research question could be: “Would the difference in students’ performance depend on the degree of realistic contexts being used in a particular curriculum? Or in the assessment practice?” These differences could be identified and taken into account in the statistical model to compare the scores on the instrument between students from different countries. For that purpose, the English translation of the instrument can be adapted for use in other countries or for comparable research designs.

Actual student behaviour in contextual mathematical problem solving. Future research could try to find answers to specific research questions, such as:

- Does it take it longer for students to solve image-rich numeracy problems?
- What is the typical reading behaviour of students in solving both versions of the tasks? This can be measured by eye-tracking and comprehension studies, (e.g., Van Silfhout, 2014) .
- What are the typical problem-solving considerations made by the students in both versions of representing the problem situation?

Incorporating Cognitive Load Theory in the theoretical framework. Cognitive Load Theory (CLT) gives insight into the way students process combinations of verbal and depictive information (Paas et al., 2010; Sweller, 2010; Sweller et al., 1998; Van Merriënboer & Sweller, 2005). We found a CLT-expert willing to make an expert estimation on the relative performance of the students on the two versions. The CLT-expert estimated students’ performance on image-rich numeracy problems systematically lower than the results from the trial indicated. The effect of using more depictive representation of the problem situation next to a verbally posed question was more positive than expected from the CLT perspective. It seems that the reduction in suspension of sense-making by using depictive representations of problem situations, could arguably have an opposite effect to what was predicted from the CLT perspective, and hence mitigate the prediction from the CLT perspective. This calls for a more systematic approach, for example with a randomised controlled design in which the perspectives are combined. The designed instrument can be adapted for that purpose.

Multimedia representation of the problem situation. Technological developments provide new possibilities. Instead of using photographs and other static information carriers, there are now more possibilities to represent problem situations, for example, with short video-clips and dynamic animations. In the near future, representations of problem situations could evolve to instances of virtual or augmented reality for educational purposes. Some educational uses have already been explored (Bower et al., 2014; Cuendet, Bonnard, Do-Lenh, & Dillenbourg, 2013; Sommerauer & Müller, 2014). This gives rise to inspiring new perspectives on how to bring problems from real life into classrooms to prepare students for mathematical problem solving in practical day-to-day real-life problem situations.

Relevance for the Dutch situation

The reported studies are particularly relevant for the Dutch context. In 2010 a “Referentiekader Taal en Rekenen” [Dutch Literacy and Numeracy Framework (LaNF)] passed into law. This law decreed a set of high stakes final examinations in numeracy, namely: Rekentoets vo [Numeracy examination secondary education] and Rekenexamen mbo [Numeracy examination vocational education]. Although there is a vast experience with contextual mathematics education in the Netherlands – in primary education since the 1970s, in upper secondary education since the 1980s, and in vocational education since the 1990s – the debate on the best ways to represent problem situations from real life in mathematics education in a fruitful and effective way is still very much alive. The use of contexts, and more specifically the use of word problems in, for instance, examinations is a recurring subject in debates in Dutch politics and (social) media. The theoretical framework and empirical results of the studies described in this dissertation can inform that discussion.

Samenvatting

Probleemstelling en onderzoeksvraag

In dit proefschrift wordt de volgende onderzoeksvraag beantwoord:

> Wat is bij wiskundige contextopgaven het effect op de score van leerlingen van een verandering in de representatie van de probleemsituatie van beschrijvend naar voornamelijk beeldend?

Het antwoord op deze vraag kan inzicht verschaffen in een belangrijke kwestie in reken- en wiskundeonderwijs, namelijk de vraag waarom leerlingen zo vaak hun gezond verstand lijken uit te schakelen bij het oplossen van wiskundige contextopgaven met slechte resultaten en een gebrekkige ontwikkeling van hun probleemoplosvaardigheden als gevolg.

We onderzoeken of een meer beeldende representatie van de probleemsituatie – met foto's van echte probleemsituaties – de kans vergroot dat leerlingen het probleem aanpakken vanuit een probleemoplossende houding, dat wil zeggen rekening houdend met alle aspecten en beperkingen die de probleemsituatie oproept. Wij verwachten dat deze benadering het "uitschakelen van het gezond verstand" tegengaat of zelfs voorkomt, met betere leerlingresultaten tot gevolg. Alhoewel er veel onderzoek is gedaan naar dit "uitschakelen van het gezonde verstand" [suspension of sense-making], ontdekten wij een kennishiaat als het gaat om het effect op leerlingresultaten van het veranderen van de representatie van de probleemsituatie naar meer beeldend. Er is hiernaar nog maar weinig onderzoek gedaan. Ook is er vrijwel geen literatuur over een systematische vergelijking van het gebruik van beschrijvende en beeldende representaties van probleemsituaties in wiskundige contextopgaven en het effect daarvan op leerlingresultaten.

De belangrijkste stappen in de gekozen onderzoeksopzet zijn de volgende. Ten eerste schetsen wij de redenering die de eerdergenoemde kwestie – het uitzetten van het gezond verstand – verbindt met de onderzoeksvraag. Ten tweede beschrijven wij hoe de onderzoeksvraag is aangepakt door middel van deelonderzoeken; welke resultaten de deelonderzoeken hebben opgeleverd; en hoe deze resultaten leiden tot een antwoord op de onderzoeksvraag. Ten derde betogen wij wat de relevantie kan zijn van de resultaten van het onderzoek voor de praktijk van het reken- en wiskundeonderwijs en voor het wetenschappelijk onderzoeksdomein reken- en wiskundeonderwijs.

Conceptueel kader en redeneerlijn

In de afgelopen decennia is de aandacht in het wiskundeonderwijs verschoven van een sterke nadruk op wiskundige procedures naar meer nadruk op toepassingen van wiskunde, modelleren, probleemoplossen en het creëren van een probleemoplossende houding bij leerlingen. In de huidige curricula en schoolboeken zijn veel meer verwijzingen naar (verschijnselen uit) de echte wereld zichtbaar dan dertig jaar geleden (Blum et al., 2007). Dit geldt met name voor de takken van het wiskundeonderwijs die zich richten op gecijferdheid, wiskundige geletterdheid en kwantitatieve geletterdheid, waarbij het centrale thema is hoe mensen kunnen leren wiskunde te gebruiken in het omgaan met kwantitatieve problemen uit het dagelijks leven (Kilpatrick, 1996; Niss, 1996; OECD, 1999; Toner, 2011).

Wiskundige contextopgaven spelen een belangrijke rol in wiskundeonderwijs dat de nadruk legt op modelleren en probleemoplossen. Een typische wiskundige contextopgave bestaat uit een probleemsituatie die wordt gepresenteerd en een vraag die daarover wordt gesteld. In de meeste gevallen is de probleemsituatie een weergave van een situatie uit het echte leven, terwijl de vraag gericht is op het activeren van de leerlingen om de vraag met wiskundig redeneren, wiskundige kennis en wiskundige instrumenten te beantwoorden. De overheersende vorm van weergeven van de probleemsituatie is deze te beschrijven met woorden. Dit genre van "talige contextopgaven" [word problems] (Lave, 1992; Verschaffel et al., 2000; Verschaffel et al., 2009) wordt wereldwijd aangetroffen in reken- en wiskundeonderwijs rond het onderwerp probleemoplossen.

De manier waarop de leerlingen omgaan met talige contextopgaven wordt in het wiskundeonderwijs gezien als problematisch. Waarschijnlijk als een uitvloeisel van de procedurele nadruk in het wiskundeonderwijs, benaderen veel leerlingen talige contextopgaven hardnekkig als procedurele oefeningen in vermomming, en handelen dienovereenkomstig: ze halen zo snel mogelijk een aantal getallen uit de probleemsituatie en voeren daar een bewerking op uit zonder het probleem te doorgronden en zonder rekening te houden met de beperkingen van de geschetste situatie.

De literatuur bevat veel verwijzingen naar deze problematische kant van talige contextopgaven (DeFranco & Curcio, 1997; Dewolf et al., 2014; Gellert & Jablonka, 2009; Inoue, 2009; Reusser & Stebler, 1997; Roth, 2009; Verschaffel et al., 1997; Verschaffel et al., 1994). Deze problematiek wordt pijnlijk zichtbaar in de ontoereikende of zelfs onzinnige oplossingen die in leerlingwerk veelvuldig worden aangetroffen. Het uitschakelen van het gezond verstand bij het oplossen van talige contextopgaven wordt in het reken- en wiskundeonderwijs beschouwd als een onwenselijke situatie (Schoenfeld, 1991; Verschaffel et al., 2000). Het wordt gezien als een belemmering om een probleemoplossende mindset te verwerven, aangezien actief betekenis geven aan het probleem wordt beschouwd als cruciaal voor een adequaat proces van probleemoplossen (Schoenfeld, 1992). Een adequaat proces van probleemoplossen begint met het interpreteren van de probleemsituatie en vervolgens een samenhangend model van het probleem construeren om te komen tot antwoorden die zinvol zijn als oplossing van het probleem (Schnotz, 2002).

Als maatregel tegen het uitschakelen van het gezond verstand hebben veel reken- en wiskundedidactici de suggestie gedaan om de representatie van de problemen authentieker te maken en dichter bij het echte probleem te brengen. Dat verhoogt mogelijk de kans dat leerlingen in een probleemoplossende mindset blijven, hetgeen kan resulteren in betere prestaties. In ons onderzoek wordt daarom een specifieke maatregel tegen het uitschakelen van het gezond verstand onderzocht, namelijk het veranderen van de representatie van probleemsituatie van beschrijvend – zoals in talige contextopgaven – naar voornamelijk beeldend (Schnotz et al., 2010), waarbij voor de beeldende representatie foto's van echte situaties worden gebruikt. Wij betogen dat een dergelijke wijziging de leerlingen meer authenticiteit laat ervaren tijdens het werken aan het probleem, waardoor de kans om tot een adequate oplossing te komen toeneemt. Of anders geformuleerd: waardoor de kans op het uitschakelen van het gezond verstand afneemt.

Het theoretisch kader voor het onderzoek in dit proefschrift combineert verschillende gezichtspunten en invalshoeken. Ten eerste is gebruikgemaakt van concepten als wiskundige geletterdheid (Jablonka, 2003, 2015; OECD, 2012), en vergelijkbare begrippen als gecijferdheid (Gal et al., 2003; Tout & Gal, 2015), kwantitatieve geletterdheid (Madison & Steen, 2008; Steen, 2001), matheracy (D'Ambrosio, 1997), en functioneel rekenen of functionele wiskunde (Forman & Steen, 1999). Ten tweede is de uitgebreide kennisbasis benut die betrekking heeft op probleemoplossen, modelleren en het oplossen van talige contextopgaven, en de problemen die talige contextopgaven veroorzaken in het reken- en wiskundeonderwijs (Blum et al., 2007; Galbraith & Stillman, 2006; Verschaffel et al., 2000). Ten derde is gebruik gemaakt van gezichtspunten uit de cognitieve psychologie met betrekking tot mentale representaties en het gebruik van visuele elementen tijdens het leren (Schnotz, 2005; Schnotz et al., 2010; Schnotz & Bannert, 2003; Schnotz & Kürschner, 2008). Deze gezichtspunten en invalshoeken worden uitgebreid beschreven in de hoofdstukken 1, 3 en 4.

Methode en bevindingen

In dit proefschrift beschrijven we vier empirische deelonderzoeken die zijn uitgevoerd om een antwoord op de onderzoeksvraag te vinden: een verkennend onderzoek (hoofdstuk 2); het ontwerp en de validering van een instrument om de prestaties van leerlingen op verschillende rekentaken te kunnen meten en vergelijken (hoofdstuk 3); een grootschalig gerandomiseerd onderzoek in het primair en voortgezet onderwijs (hoofdstuk 4 en 5); en een replicatieonderzoek met volwassen deelnemers (hoofdstuk 6). De resultaten van deze deelonderzoeken leverden de bouwstenen voor de beantwoording van de onderzoeksvraag.

De interviewgegevens uit het verkennende onderzoek gaven een indicatie welk soort wiskundige representatie betekenisvol kan zijn voor de leerlingen. Deze kennis is gebruikt om beeldende en relevante rekenopgaven te ontwerpen als alternatief voor talige contextopgaven. De ontworpen taken zijn vervolgens verwerkt in het ontwerp en de validering van een instrument waarmee het verschil in de prestaties van leerlingen kon worden gemeten als de representatie van de probleemsituatie veranderd werd van talig naar voornamelijk beeldend. Het instrument is vervolgens ingezet in een grootschalige gerandomiseerde trial met 11- tot 20-jarigen, en is als replicatie herhaald met volwassenen als doelgroep.

De belangrijkste bevindingen en conclusies van de deelonderzoeken worden hier nu samengebracht om daarmee de onderzoeksvraag naar het effect op leerlingresultaten van het veranderen van de representatie van de probleemsituatie te kunnen beantwoorden. Daarnaast is onderzocht of het vermoeden kon worden bevestigd, dat leerlingen beter zouden presteren op wiskundige contextopgaven, wanneer de probleemsituatie voornamelijk beeldend werd gerepresenteerd,

Het eerste deelonderzoek was een verkennend onderzoek naar de gecijferde wereld van leerlingen uit het beroepsonderwijs (hoofdstuk 2). Een reeks semi-gestructureerde interviews met leerlingen leverde “gecijferdheidssituaties” op, dat wil zeggen situaties waarover de leerlingen vertelden en waarin ze een kwantitatieve redenering of handeling moesten uitvoeren. Deze gecijferdheidssituaties zijn vervolgens gecategoriseerd op basis van twee belangrijke raamwerken: het internationale PISA-raamwerk over rekenvaardigheid (OECD, 2013a) en het Referentiekader Taal en Rekenen (Ministerie van OCW, 2009). Deze categorisering bleek goed mogelijk, waarmee de relevantie van deze raamwerken om kwantitatieve situaties uit het dagelijkse leven te duiden, werd aangetoond. Tevens vonden wij dat leerlingen een veel beter begrip van wiskundige concepten toonden als zij geïnterviewd werden over aan het werk gerelateerde situaties, in het bijzonder als het ging over producten die zij zelf hadden gemaakt en bij de hand hadden. Leerlingen zetten veelvuldig gebaren en een eigen vocabulaire in bij het beantwoorden van de interviewvragen. Deze opbrengsten moedigden ons aan om te zoeken naar authentieke, visuele representaties van probleemsituaties – bijvoorbeeld door gebruik te maken van foto’s –, die zich voordoen in het leven en werk van mensen (zie Appendix A voor voorbeelden).

Het tweede deelonderzoek behelsde het ontwerp en de validering van een toetsinstrument waarmee het verschil in leerlingresultaten gemeten kon worden bij verschillende representaties van de probleemsituatie (hoofdstuk 3). In het ontwerp van het instrument werden de representaties van de probleemsituaties dichter bij echte situaties gebracht met behulp van foto's, waardoor de ervaren authenticiteit voor de leerlingen kon toenemen.

De validatie van het instrument bestond uit drie stappen. Ten eerste werden voor bestaande talige contextopgaven beeldende alternatieven ontwikkeld, zodat er koppels van opgaven ontstonden met verschillende representaties van eenzelfde probleemsituatie (zie Appendix A). Deze opgaven zijn vervolgens gekoppeld aan de domeinen uit het Referentiekader Taal en Rekenen. Ten tweede hebben twee panels van deskundigen op het gebied van reken- en wiskundeonderwijs commentaar geleverd op een eerste set opgaven voor het instrument en hun expert-mening gegeven over de mate van gelijkwaardigheid van de geselecteerde talige rekenopgaven en de alternatieve beeldende contextopgaven. Het gaat dan om gelijkwaardigheid op het gebied van de uit te voeren reken- of wiskundige handelingen om tot een oplossing van het probleem te komen. Ten derde is met een voorlopig instrument een test run (N = 7.454) uitgevoerd om vast te stellen of aan de vereisten van een gecontroleerde gerandomiseerde trial kon worden voldaan. Op deze manier zijn de belangrijkste bedreigingen van de validiteit van het instrument – inhoudsvaliditeit, begripsvaliditeit, criteriumvaliditeit, betrouwbaarheid en haalbaarheid – ondervangen. De drie genoemde stappen rond validatie zijn succesvol uitgevoerd, zodat het uiteindelijke instrument kon worden gebruikt om op valide wijze na te gaan wat het effect was op de leerlingresultaten als gevolg van het veranderen van de representatie van de probleemsituatie. Na het ontwerp is het instrument geprogrammeerd als een online toets.

De daadwerkelijke trial met het instrument is in een periode van vier weken gehouden (zie hoofdstuk 4). Het instrument werd via internet ter beschikking gesteld aan scholen om te gebruiken als een rekentoets. In totaal hebben 31.842 leerlingen van 179 scholen, geografisch verspreid over Nederland, deelgenomen aan de trial. Om te toetsen of leerlingen hoger zouden scoren op beeldende rekenopgaven, zijn diverse statistische tests uitgevoerd (hoofdstuk 4). Om dieper inzicht te verkrijgen in het effect van de achtergrondvariabelen en in mogelijke interactie-effecten is een probit model gebruikt.

Na analyse was het algehele resultaat een 1,9 procentpunt stijging in de prestaties van leerlingen bij een verandering van een beschrijvende in een beeldende representatie van de probleemsituatie. Dit verschil was statistisch significant (p <0,001) met een zeer kleine effectgrootte van Cohen's d = 0,09.

Dus, hoewel ons vermoeden werd bevestigd, moet gezien de kleine verandering en de kleine effectgrootte, deze uitspraak met de grootste voorzichtigheid worden gedaan. Om deze bevindingen in perspectief te plaatsen, verwijzen wij naar een recente publicatie van Slavin (2016), waarin hij laat zien dat in onderwijsonderzoek, dat gebruikmaakt van gerandomiseerde trials, zelden effecten te vinden zijn met een effectgrootte van meer dan 0,2 en dat met een onderzoekspopulatie van groter dan 5.000 valide gevolgtrekkingen gemaakt kunnen worden met effectgroottes van 0,1. Als we deze constatering van Slavin (2016) volgen, lijken geldige gevolgtrekkingen in dit onderzoek gerechtvaardigd.

Verder is nader onderzocht wat de onderlinge afhankelijke effecten van de achtergrondvariabelen zijn op het totale effect. In het statistische model zijn de achtergrondvariabelen, zoals geslacht, etniciteit, schooltype, leerjaar, leeftijd en wiskundeniveau als kruistermen aan de variabele versie gekoppeld. Het bleek dat geen van deze variabelen een significant effect hadden op het gemeten verschil in leerlingresultaat op de verschillende representaties van de probleemsituatie, behalve de variabele geslacht. Dit betekent dat het gemeten effect niet afhankelijk was van de achtergrondvariabelen, met uitzondering van geslacht. Voor mannelijke deelnemers was het verschil in scores op beide versies van de opgaven iets hoger dan voor vrouwelijke deelnemers.

Om meer inzicht in de relatie tussen de opgavekenmerken en de gevonden resultaten te krijgen, voerden we een uitgebreidere statistische analyse uit (hoofdstuk 5). Ten eerste werden twee variabelen geïntroduceerd om de opgaven te karakteriseren: het *aantal gebruikte woorden in de probleemsituatie* voor de A-versie van de opgaven en het *aantal gebruikte beeldelementen in de probleemsituatie* voor de B-versie van de opgaven. Uit de statistische analyse vonden we een zeer tentatieve indicatie voor de toename in leerlingresultaat per woord reductie en voor de grootte van de afname van het leerlingresultaat per toegevoegd beeldend element. Het laatste resultaat gaf aan dat het toevoegen van meer beeldende elementen geen positief effect heeft op de prestaties van leerlingen.

Ten tweede analyseerden we het effect op de leerlingresultaten van het domein waartoe de opgave behoorde. We verwachtten op basis van de theorie (Schnotz et al., 2010) dat in het domein Meten en Meetkunde het grootse effect zou optreden, meer dan bij opgaven in het domein Getallen of

Verhoudingen, aangezien de beeldende elementen in opgaven uit het domein Meten en Meetkunde vaak al een sterke relatie hebben met het wiskundige model dat nodig is om het probleem succesvol op te lossen. Deze verwachting werd bevestigd door onze analyse. We analyseerden de interactie-effecten van de variabelen domein en versie, en vonden een significant effect. We vonden het kleinste effect van het veranderen van de representatie van de probleemsituatie bij opgaven in het domein Verhoudingen en het grootste effect bij opgaven uit het domein Meten en Meetkunde. Het effect in het domein Getallen lag in tussen deze effecten, maar dichter bij het effect van opgaven uit het domein Verhoudingen. Het relatief grotere effect op de taken in het domein Meten en Meetkunde was in lijn met onze verwachtingen.

Het deelonderzoek dat hierboven beschreven is, richtte zich op leerlingen in de leeftijd 11 tot 20 jaar. Een vervolgvraag was of gelijkwaardige resultaten kon worden waargenomen bij andere doelgroepen, bijvoorbeeld volwassenen. Mogelijk zijn volwassenen minder gevoelig voor het uitschakelen van het gezonde verstand, aangezien dat verschijnsel vaak gekoppeld wordt aan schoolwiskunde en aan het maken van opgaven in de wiskundeles. We voerden een replicatieonderzoek (hoofdstuk 6) uit met 420 volwassen deelnemers aan het Groot Nationaal Rekenonderzoek, een onderzoeksinitiatief van de publieke omroepen VPRO en NTR, gesteund door Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). De algemene resultaten van de prestaties in dit onderzoek met volwassen deelnemers was 1,1 procentpunt hogere scores op de problemen met een beeldende voorstelling van de probleemsituatie. Dit was in lijn met de algemene resultaten die we vonden in de grote trial, namelijk 1,9 procentpunt hogere score op de beeldende contextopgaven. Bij het vergelijken van de item-scores in beide trials vonden we overeenkomsten en verschillen in scores. We vonden vier taken die in beide trials een significant betere prestaties voor de beeldende versie lieten zien. Drie van deze waren afkomstig uit het domein Meten en Meetkunde. Dit bevestigt onze eerdere bevindingen dat bij taken uit het domein Meten en Meetkunde de verandering van een beschrijvende representatie van de probleemsituatie naar een meer beeldende de grootste positieve invloed heeft op de score van de deelnemers.

Concluderend vonden we een positief hoofdeffect op de prestaties van de deelnemers door het veranderen van de beschrijvende representatie van de probleemsituatie naar een meer beeldende (hoofdstuk 7). Verder konden de meeste andere variabelen als mogelijke onderlinge afhankelijke variabelen worden uitgesloten. Vanwege de grote omvang van de steekproef lijken dergelijke conclusies gerechtvaardigd ondanks de kleine effectgrootte. Voor het vaststellen dat we te maken hebben met een robuust – in de zin van niet afhankelijk van andere variabelen – en algemeen effect veroorzaakt door het veranderen van de representatie van de probleemsituatie dienen meer replicaties of onderzoeken met grotere aantallen items te worden uitgevoerd. Daartoe is het instrument vertaald in het Engels en wordt het beschikbaar gesteld via open access, hetgeen de mogelijkheid biedt tot een replicatieonderzoek in Engelstalige landen. Replicaties kunnen ook gehouden worden met andere of meer items, bijvoorbeeld om nog specifieker te onderzoeken of en hoe het effect zichtbaar is in de verschillende domeinen.

Alle conclusies op een rijtje gezet hebbend, luidt het antwoord op de onderzoeksvraag als volgt: Het veranderen van de representatie van de probleemsituatie van beschrijvend naar voornamelijk beeldend door gebruik te maken van foto’s, vergroot de kans dat leerlingen tot een correcte oplossing van het probleem komen. We interpreteren dit als een indicatie dat het uitschakelen van het gezond verstand iets minder heeft plaatsgevonden. De veranderingen in de genoemde kans zijn klein en in zekere mate afhankelijk van het domein van de taak. We hebben de beperkingen van de diverse deelonderzoeken vastgesteld. Daarom dienen de resultaten van deze deelonderzoeken met zorg te worden verwoord, waarbij de beperkingen in de conclusies worden meegenomen. Niettemin kunnen onze bevindingen worden toegevoegd aan de kennisbasis met betrekking tot het oplossen van wiskundige contextopgaven en het stimuleren van het gebruik van gezond verstand bij het oplossen van wiskundige contextopgaven. Voor zover wij hebben kunnen nagaan, is dit het eerste onderzoek op deze schaal waarin een systematische vergelijking is gemaakt van leerlingresultaten bij het oplossen van wiskundige contextopgaven met verschillende representaties van de probleemsituatie.

Relevantie van het onderzoek

Algemene relevantie

Dit proefschrift draagt bij aan de wereldwijde trend in het wiskundeonderwijs om meer nadruk te leggen op het oplossen van wiskundige contextproblemen. Het onderzoeken en het beantwoorden van de onderzoeksvraag was erop gericht inzicht te verwerven in het effect op leerlingresultaten van verschillende soorten representaties van probleemsituaties. Ook brachten de gerapporteerde deelonderzoeken de complexiteit in beeld van het genre van talige contextopgaven en gaven zij zicht op de leerlingresultaten op talige contextopgaven in vergelijking met leerlingresultaten op meer beeldende contextopgaven.

De onderzoeksresultaten kunnen van belang zijn voor leerlingen die, door de verandering van de representatie van de probleemsituatie naar voornamelijk beeldend, mogelijk beter kunnen presteren met minder uitschakeling van hun gezond verstand. De onderzoeksresultaten kunnen ook van belang zijn voor leerkrachten en methodeschrijvers, aangezien het kan leiden tot een nadere doordenking van de vraag waarom wiskundige contextopgaven – bijna vanuit gewoonte – zo vaak als verhaaltjessommen worden vormgegeven. En de onderzoeksresultaten kunnen van belang zijn voor beleidsmakers, omdat wiskundige contextopgaven in grote mate voorkomen in summatieve toetsen met een groot civiel effect, zoals rekentoetsen, rekenexamens, en wiskunde-examens. Daarnaast kunnen de resultaten worden benut in onderzoeken die reken- en wiskundeprestaties van landen vergelijken zoals PISA en PIAAC.

Relevantie voor toekomstig onderzoek

Toekomstig onderzoek naar een beter begrip van het feitelijke gedrag dat leerlingen vertonen bij het oplossen van wiskundige contextopgaven kan voortbouwen op de resultaten van dit onderzoek.

Het effect van gewenning. Landen verschillen in de mate waarin aandacht in het reken- en wiskundecurriculum is voor gecijferdheid en de mate waarin de gebruikte contexten levensecht zijn. Vanuit het vermoeden dat leerlingen die meer gewend zijn aan beeldende contextopgaven minder gevoelig zijn voor een verandering van beschrijvende naar beeldende representaties, zou een follow-up onderzoeksvraag kunnen zijn: "Is het verschil in de prestaties van leerlingen afhankelijk van de mate waarin realistische contexten wordt gebruikt in het curriculum van een land?" Dergelijke verschillen kunnen worden vastgesteld en in het statistische model opgenomen om het effect ervan op de leerlingresultaten te analyseren. Daartoe kan de Engelse vertaling van het instrument worden aangepast voor gebruik in andere landen of voor vergelijkbare onderzoeken.

Daadwerkelijk leerlinggedrag bij het oplossen van wiskundige contextopgaven. Toekomstig onderzoek zou kunnen proberen antwoorden te vinden op specifieke onderzoeksvragen, zoals:

- Duurt het langer voor leerlingen om beeldende contextopgaven op te lossen? Gegevens hierover zijn reeds in de onderzoeksdatabase aanwezig.
- Wat is het typerende "leesgedrag" van leerlingen bij het oplossen van beide varianten van de opgaven. Dit kan worden gemeten door eye-tracking. Zie bijvoorbeeld het onderzoek van Van Silfhout (2014).
- Wat zijn typerende overwegingen die leerlingen maken bij het oplossen van beide versies van de opgaven?

Integratie van Cognitive Load Theory in het theoretisch kader. Cognitive Load Theory (CLT) geeft inzicht in de manier waarop leerlingen combinaties van talige en beeldende informatie verwerken (Paas et al., 2010; Sweller, 2010; Sweller et al., 1998; Van Merriënboer & Sweller, 2005). Een CLT-expert werd bereid gevonden om een inschatting te maken van de relatieve prestaties van de leerlingen op de twee versies van de opgaven. De CLT-expert schatte de leerlingresultaten op de beeldende opgaven systematisch lager dan de resultaten die uit de trial voortkwamen. Het effect van het gebruik van meer beeldende representatie van de probleemsituatie naast een talig gestelde vraag was minder negatief dan verwacht vanuit een CLT perspectief. Het lijkt dat het minder uitschakelen van het gezond verstand in de beeldende opgaven een tegengesteld effect is aan het effect dat verwacht wordt vanuit CLT. Dit vraagt om nader onderzoek met een meer systematische aanpak, bijvoorbeeld met een gerandomiseerd ontwerp waarin beide perspectieven worden gecombineerd. Het ontworpen instrument kan voor dat doel worden aangepast.

Multimediale representatie van de probleemsituatie. Technologische ontwikkelingen bieden nieuwe mogelijkheden. In plaats van foto’s en andere statische informatiedragers zijn er nu meer mogelijkheden om probleemsituaties te representeren, bijvoorbeeld korte videoclips en dynamische animaties. In de nabije toekomst kunnen representaties van probleemsituaties evolueren naar gevallen van “virtual reality” of “augmented reality” voor educatieve doeleinden. Enkele educatieve toepassingen van “augmented reality” zijn reeds verkend (Bower et al., 2014; Cuendet et al., 2013; Sommerauer & Müller, 2014). Gebruik van multimedia geeft aanleiding tot inspirerende nieuwe perspectieven over hoe problemen uit het echte leven in de klas gebracht kunnen worden om leerlingen voor te bereiden op het oplossen van problemen uit de dagelijkse praktijk.

Relevantie voor de Nederlandse situatie

De resultaten van de deelonderzoeken, met name de resultaten van de trials, zijn bijzonder relevant voor de Nederlandse context. De wet "Referentieniveaus Nederlandse taal en rekenen" is op 1 augustus 2010 in werking getreden. Als gevolg hiervan zijn twee verplichte summatieve toetsen in het onderwijs geïntroduceerd, namelijk de Rekentoets voor het voortgezet onderwijs en het Rekenexamen voor het mbo. En alhoewel er ruime ervaring in Nederland is met contextrijk reken- en wiskundeonderwijs – in het basisonderwijs sinds de jaren zeventig, in de bovenbouw van het voortgezet onderwijs sinds de jaren tachtig, en in het beroepsonderwijs sinds de jaren negentig – is het debat over de beste manieren om probleemsituaties uit het echte leven te (re)presenteren in het reken- en wiskundeonderwijs nog steeds springlevend. Het gebruik van contexten, en meer in het bijzonder het gebruik van talige contextopgaven in bijvoorbeeld de examens, is een terugkerend onderwerp in de debatten in de Nederlandse politiek (SLO, 2014) en de (social) media. Het gebruikte theoretisch kader en de empirische resultaten van de deelonderzoeken zoals ze beschreven zijn in dit proefschrift kunnen die discussie van een meer wetenschappelijke basis voorzien.

Dankwoord

Vele collega's hebben mij gesteund bij deze onderneming. Ingrid Verheggen (APS) gaf mij de gelegenheid om de eerste stappen op het promotiepad te zetten naast mijn professionaliseringswerk bij APS.

De volgende collega's hebben mij ondersteund en geïnspireerd bij het inhoudelijke werk aan deze dissertatie. Ik wil mijn beoogd promotoren Anne van Streun en Koeno Gravemeijer bedanken, die mij met hun grote en doorleefde kennis van reken- en wiskundeonderwijs hebben begeleid. Door de tijd die ik nam voor het afronden, overschreed ik echter hun *ius promovendi*. Steunpilaren gedurende het hele proces waren Arthur Bakker die met zijn voortdurende streven naar de hoogste wetenschappelijke kwaliteit mij veel heeft geleerd, en Jaap de Koning die met zijn enorme kennis van statistische technieken en modellen van grote waarde was bij het analyseren en interpreteren van de verzamelde gegevens.

I would like to say special thanks to Birgit Pepin, who took over as promotor. She was with her enormous working power, great knowledge of methodology and research in mathematics education, broad international orientation, ánd sound pragmatism, an invaluable guide for me – and just what I needed to finish the job.

Verder wil ik Mieke van Groenestijn en Martin van Reeuwijk bedanken als lid van de wetenschappelijke adviesraad bij het Onderwijsbewijs II project dat ten grondslag lag aan het hoofdonderzoek. Ook wil ik Ruben IJzerman en Teun de Lange bedanken die vanuit Intraquest de technische uitvoering van het hoofd- en replicatieonderzoek mogelijk maakten. Op het hoogtepunt deden gedurende een aantal dagen achtereen meer dan 3000 leerlingen tegelijk on-line de toets. Daarnaast wil ik Dédé de Haan bedanken die mij zeer regelmatig op het spoor zette van relevante literatuur. Ik ben Jan van den Akker (SLO) en Hans de Vries (SLO) zeer erkentelijk dat zij mij in de gelegenheid stelden het werk af te maken en de dissertatie te produceren.

En ten slotte wil ik mijn lief en levenspartner Monique bedanken voor de liefde, de steun bij alle ups en downs, en de thuisbasis. Zonder haar zou ik dit punt in mijn leven waarschijnlijk niet bereikt hebben.

List of publications

Publications marked with an asterisk (*) are partly or wholly included in the dissertation.

Peer reviewed journal articles and peer-reviewed papers

* Hoogland, K., Bakker, A., Koning, J. de, Pepin, B. & Gravemeijer, K. (2016). Representing contextual mathematical problems in descriptive or depictive form: Design of an instrument and validation of its uses. *Studies in Educational Evaluation, 50*, 22-32. doi: 10.1016/j.stueduc.2016.06.005

* Hoogland, K., Bakker, A., De Koning, J., Pepin, B., & Gravemeijer, K. (under review). Descriptive versus depictive representation of reality in contextual mathematical problems: the effect on students' performance.

* Hoogland, K., Pepin, B., De Koning, J., Bakker, A., & Gravemeijer, K. (under review). Word problems versus image-rich problems: An analysis of effects of task characteristics on students’ performance on contextual mathematics problems.

* Hoogland, K., & Pepin, B. (in press). The intricacies of assessing numeracy: Investigating alternatives to word problems. *Adults Learning Mathematics – An International Journal, 11*(2).

* Hoogland, K., & Pepin, B. (in press). The numeracy of vocational students: Exploring the nature of the mathematics used in daily life and work. *Proceedings of the 13th International Congress on Mathematical Education*. Hamburg, Germany.

* Hoogland, K., Bakker, A., De Koning, J., & Gravemeijer, K. (2012). Comparing students' results on word problems with their results on image-rich numeracy problems. In icme.org (Ed.), *Pre-proceedings of the 12th International Congress on Mathematical Education* (pp. 1722-1731). Seoul, Korea.

* Hoogland, K., & Stelwagen, R. (2011). A new Dutch numeracy framework. In T. Maguire, J. J. Keogh, & J. O’Donoghue (Eds.), *Mathematical eyes: a bridge between adults, the world and mathematics. Proceedings of the 18th international conference of adults learning mathematics – a research forum (ALM)* (pp. 193-202). Tallaght, Ireland: ITT.

* Hoogland, K. (2006). Mind and gesture: The numeracy of a vocational student. In M. Horne & B. Marr (Eds.), *Connecting voices in adult mathematics and numeracy: Practitioners, researchers and learners. Proceedings of the adults learning mathematics 12th annual*

international conference (ALM) (pp. 150-158). Melbourne, Australia: ACU.

Other scientific publications

* Hoogland, K., De Koning, J., & Tanis, O. (2013). *Rekenen in beeld: eindrapport* [Images of numeracy: final report]. Rotterdam, the Netherlands SEOR.

* Hoogland, K., & De Koning, J. (2013). *Dataset: Rekenen in beeld* [Dataset: Images of numeracy]. http://dx.doi.org/10.17026/dans-za6-5q6c

* Hoogland, K. (2012). Images of numeracy. In T. Maguire, J. J. Keogh, & J. O'Donoghue (Eds.), *Mathematical eyes: a bridge between adults, the world and mathematics. Proceedings of the 18th international conference of adults learning mathematics – a research forum (ALM)* (pp. 2-14). Tallaght, Ireland: ITT.

Hoogland, K. (2010). Realistic Numeracy Problems. In H. Christensen, J. Diez-Palomar, J. Kantner, & C. M. Klinger (Eds.), *Maths at work – Mathematics in a changing world. Adults learning mathematics, 17th international conference (ALM)* (pp. 58). Oslo, Norway: VOX.

Hoogland, K. (2010). *Sophisticated numeracy learning materials.* Paper presented at the International Society for Design and Development in Education (ISDDE), Cairns, Australia.

Hoogland, K. (2009). Beyond word problems. In G. Griffiths & D. Kaye (Eds.), *Numeracy works for life: Proceedings of the 16th international conference of adults learning mathematics – A research forum (ALM)* (pp. 300-307). London, UK: London Southbank University.

Hoogland, K. (2009). Going beyond word problems. *Numeracy Briefing*, October 2009, 13-16.

Sembiring, R. K., Hoogland, K., & Van den Hoven, G. H. (2009). *Initiation, implementation and institutionalization of realistic mathematics education in Indonesia.* Paper presented at the international conference on school improvement and effectiveness (ICSEI), Vancouver, Canada.

Sembiring, R. K., & Hoogland, K. (2008). *PMRI: A North–South partnership for improving mathematics education in Indonesia.* Paper presented at the international conference on school improvement and effectiveness (ICSEI), Auckland, New Zealand.

Hoogland, K. (2007). Towards a multimedia tool for numeracy education. In T. Maquire, N. Collearan, O. Gill, & J. O'Donoghue (Eds.), *The changing face of adults mathematics education: Learning from the past, planning for the future. Proceedings of the adults learning mathematics (ALM) – 14th international conference 2007* (pp. 165-176). Limerick, Ireland.

Books and book chapters

Hoogland, K. (in preparation). Ensuring usability: Reflections on a Dutch mathematics reform project for students aged 12–16. In M. van den Heuvel-Panhuizen (Ed.), *Reflections from inside on the Netherlands didactic tradition in mathematics education*. Utrecht, The Netherlands: University Utrecht.

Sembiring, R., Hoogland K., & Dolk M. (Eds.) (2010). *A decade of PMRI in Indonesia.* Utrecht, The Netherlands: APS International.

Amin, S., Julie, H., Munk, F., & Hoogland, K. (2010). The development of learning materials for PMRI. In R. Sembiring, K. Hoogland, & M. Dolk (Eds.), *A decade of PMRI*. Utrecht, The Netherlands: APS International.

Hadi, S., Zulkardi, & Hoogland, K. (2010). Quality assurance in PMRI. Design of standards for PMRI. In R. Sembiring, K. Hoogland, & M. Dolk (Eds.), *A decade of PMRI*. Utrecht, The Netherlands: APS International.

Sembiring, R. K., Hadi, S., & Hoogland, K. (2010). The future of PMRI. In R. Sembiring, K. Hoogland, & M. Dolk (Eds.), *A decade of PMRI*. Utrecht, The Netherlands: APS International.

Hoogland K. (2005). Numeracy. In APS International (ed.). *Twenty two theories*. Utrecht, The Netherlands: APS International.

Books and brochures (in Dutch)

Stelwagen, R., & Hoogland, K. (2015). *Het belang van rekenen en gecijferdheid in een leven lang leren.* Ede, The Netherlands: Steunpunt Taal en Rekenen Volwasseneneducatie.

Wijers, M., Hoogland K., & Jonker, V. (2010). *Rekenwijzer AKA.* Ede, The Netherlands: Steunpunt Taal en Rekenen mbo.

Hoogland, K., Van Reeuwijk, M., Sjoers, S., Vliegenthart, M., Van Vugt, R., & Van Wijk, P. (2009). *Rekenen in het mbo. Waarom? Wat? Hoe?* Utrecht, The Netherlands: APS.

Hoogland, K., Van Reeuwijk, M., Sjoers, S., Vliegenthart, M., Van Vugt, R., & Van Wijk, P. (2009). *Rekenen in het voortgezet onderwijs. Waarom? Wat? Hoe?* Utrecht, The Netherlands: APS.

Hoogland, K., & Meeder, M. (2007). *Gecijferdheid in beeld.* Utrecht, The Netherlands: APS.

Hoogland, K., & Kok D. (1997). *Wiskunde: een vakvertaling bij Onderwijskunde*. Groningen, The Netherlands: Wolters Noordhoff

Wijers, M., & Hoogland K. (1994). *Vijf jaar wiskunde A–lympiade*. Utrecht, The Netherlands: Freudenthal Instituut

(Teacher) Journal publications (in Dutch)

Hoogland, K. (2016). Columns gecijferdheid. *Euclides* 91(4,5,6,7)
Hoogland, K. (2015). Columns gecijferdheid. *Euclides* 91(1,2,3)
Hoogland, K. (2012). Getallentheorie. in: *Euclides* 2012 special, 3-10.
Hoogland, K. (2011). Rekentoetsen in vo en mbo. *Panama Post reken-wiskundeonderwijs: Onderzoek, ontwikkeling, praktijk*, 30(4), 29-31.
Hoogland, K. (2009). Nostalgische terugblik op de staartdeling. *Nieuw Archief voor Wiskunde* 5/9, 279-281.
Hoogland, K. (2008). Een andere kijk op de kwantitatieve wereld. Gecijferdheid voor zwakke rekenaars. *APS Nieuws*, 61.
Hoogland, K. (2008). Groen gecijferd. *Vakblad Groen Onderwijs*, 50(12), 18-20.
Hoogland, K. (2008). Van sommen maken naar gecijferdheid. *Beroep Docent*, 2008(5), 26-27.
Hoogland, K. (2007). Rekenen is meer dan sommen maken. Begripsverwarring verduistert rekendiscussie. *VO Magazine*, 1(7), 23-26.
Hoogland, K. (2007). Weg met de angst voor rekenen. *APS Nieuws*, 58.
Hoogland, K. (2006). "Wel weten wat a is ….". *Nieuwe Wiskrant. Tijdschrift voor Nederlands Wiskundeonderwijs*, 26(1), 7.
Hoogland, K. (2006). Pleidooi voor gecijferdheid. *Van 12 tot 18. Praktijkblad voor het voortgezet onderwijs* 2006(5), 28–29.
Hoogland, K. (2005). Hoe gecijferd is een basisberoepsgerichte leerling in het VMBO? *Nieuwe Wiskrant. Tijdschrift voor Nederlands Wiskundeonderwijs*, 24(3), 36–39.
Hoogland, K. (2005). Gecijferd. *Euclides*, 80(4), 186–189.
Hoogland, K. & Jablonka E. (2003). Wiskundige geletterdheid en gecijferdheid. *Nieuwe Wiskrant. Tijdschrift voor Nederlands Wiskundeonderwijs*, 23(1), 31-37.

Curriculum vitae

Kees Hoogland was born on 13 March 1961 in Vlaardingen, the Netherlands. He finished pre-university education (Gymnasium β) in 1979 in Vlaardingen and studied pure mathematics at the University of Leiden, graduating in 1988 with specialisations in didactics of mathematics and educational sciences. From 1984 he worked in mathematics education, as a teacher, teacher educator, researcher, textbook author and editor of *Euclides*, the journal of the Nederlandse Vereniging van Wiskundeleraren [Dutch Association of Mathematics Teachers]. From 1998 he worked at APS, the national institute for school development, as in-service teacher educator, coach, advisor, international educational consultant on mathematics education, with a special interest in mathematical literacy and numeracy, and CEO. He was active as consultant in Belarus, Dutch Antilles, Greece, Indonesia, South Africa, and Surinam. From 2001 to 2011 he was project leader of the Pendidikan Matematika Realistik Indonesia (PMRI) project in Indonesia on empowering primary school teachers to improve mathematics education. He is a trustee of Adults Learning Mathematics – A research forum (ALM), and a fellow of the International Society for Design and Development of Education (ISDDE). He is a member of the Educational Committee of the Platform Wiskunde Nederland (PWN) [Dutch Organisation for Mathematics]. Additionally, he is the designer and co-developer of state-of-the-art blended learning materials in the domain of numeracy: ffLeren Rekenen. Since 2015 he has been working as a researcher and curriculum developer at Nationaal Expertisecentrum Leerplanontwikkeling (SLO) [Netherlands Institute for Curriculum Development] at Enschede, the Netherlands.

ESoE dissertation series

Sande, R. A. W. van de (2007). *Competentiegerichtheid en scheikunde leren: Over metacognitieve opvattingen, leerresultaten en leeractiviteiten*.

Hooreman, R. (2008). *Synchronous coaching of trainee teachers: an experimental approach.*

Rajuan, M. (2008). *Student teachers' perceptions of learning to teach as a basis for supervision of the mentoring relationship*.

Raessens, B. A. M. (2009). *De E-kubus: Een analysemodel voor curricula*.

Rohaan, E. J. (2009). Testing teacher knowledge for technology teaching in primary schools.

Oemar Said, E. (2009). *De Da Vinci Case: Een onderzoek naar de relaties tussen vernieuwende leeromgevingen en de motivatie en regulatievoorkeuren van leerlingen in het MBO*.

Koopman, M. (2010). *Students' goal orientations, information processing strategies and knowledge development in competence-based pre-vocational secondary education.*

Mittendorff, K. M. (2010). *Career conversations in senior secondary vocational education.*

Crasborn, F. J. A. J., & Hennissen, P. P. M. (2010). *The skilled mentor. Mentor teachers' use and acquisition of supervisory skills*.

Bragt, C. A. C. van (2010). *Students' educational careers in Higher Education: A search into key factors regarding study outcome*.

Bakker, G. de (2010). *Allocated only reciprocal peer support via instant messaging as a candidate for decreasing the tutoring load of teachers*.

Vos, M. A. J. (2010). *Interaction between teachers and teaching materials: On the implementation of context-based chemistry education.*

Bruin-Muurling, G. (2010). *The development of proficiency in the fraction domain.*

Cornelissen, L. J. F. (2011). *Knowledge processes in school-university research networks.*

Kraemer, J.-M. (2011). *Oplossingsmethoden voor aftrekken tot 100.*

Stiphout, I. M. van (2011). *The development of algebraic proficiency*.

Saeli, M. (2012). *Teaching programming for secondary school: A pedagogical content knowledge based approach*.

Putter-Smits, L. G. A. de (2012). *Science teachers designing context-based curriculum materials: Developing context-based teaching competence*.

Ketelaar, E. (2012). *Teachers and innovations: On the role of ownership, sense-making and agency*.

Dehing, F. (2012). *Preparing students for workplace learning in higher engineering education. 216*

Vrijnsen-de Corte, M. C. W. (2012). *Researching the teacher-researcher. Practice-based research in Dutch professional development schools*.

Doppenberg, J. J. (2012). *Collaborative teacher learning: Settings, foci and powerful moments*.

Linden, P. W. J. van der (2012). *A design-based approach to introducing student teachers in conducting and using research*.

Diggelen, M. R. van (2013). *Effects of a self-assessment procedure on VET teachers' competencies in coaching students' reflection skills.*

Gorissen, P. J. B. (2013). *Facilitating the use of recorded lectures: Analysing students' interactions to understand their navigational needs.*

Keuvelaar-van den Bergh, L. (2013). *Teacher feedback during active learning: The development and evaluation of a professional development programme.*

Pillen, M. T. (2013). *Professional identity tensions of beginning teachers.*

Maree, T. J. (2013). *Scripted collaborative enriched skeleton concept mapping to foster meaningful learning.*

Gomez Puente, S. M. (2014). *Design-based learning: Exploring an educational approach for engineering education.*

Coninx, N. S. (2014). *Measuring effectiveness of synchronous coaching using Bug-In-Ear device of pre-service teachers.*

Griethuijsen, R.A.L.F. van (2015). *Relationships between students' interest in science, views of science and science teaching in upper primary and lower secondary education.*

Want, A. C., van der (2015). *Teachers' interpersonal role identity.*

Bogert, N. J. van den (2016). *On teachers' visual perception and interpretation of classroom events using eye tracking and collaborative tagging methodologies.*

Beer, H. T. de (2016). *Exploring instantaneous speed in grade 5: A design research.*

Kock, Z. D. Q. P. (2016). *Towards education in agreement with the nature of science: Grade 9 electricity as a case.*

Slangen, L. A. M. P. (2016). *Teaching robotics in primary school.*

Leeferink, H. (2016). *Leren van aanstaande leraren op en van de werkplek.*

Bent, G.J.W. (2016). *Exploring geography teaching in primary education: Perspectives of teachers, teacher educators and pupils.*

Hoogland, K. (2016). *Images of numeracy. Investigating the effects of visual representations of problem situations in contextual mathematical problem solving.*

slo