University of Groningen

Can we improve children's thinking?

Jacobse, Annemieke Elisabeth

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2012

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Jacobse, A. E. (2012). *Can we improve children's thinking? A metacognitive approach to word problem solving*. [S.n.].

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 21-06-2022

Can we Improve Children's Thinking?

A Metacognitive Approach to Problem Solving in Mathematics

Annemieke Jacobse

rijksuniversiteit groningen
faculteit gedrags- en maatschappijwetenschappen

ico

Interuniversity Center for Educational Research

Cover design: Lowieke van der Heide
Layout: Robbert de Vries, Ipskamp Drukkers
Printing: Ipskamp Drukkers

RIJKSUNIVERSITEIT GRONINGEN

Can we Improve Children's Thinking?
A Metacognitive Approach to Word Problem Solving

Proefschrift

ter verkrijging van het doctoraat in de
Gedrags- en Maatschappijwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
maandag 11 juni 2012
om 16.15 uur

door

Annemieke Elisabeth Jacobse

geboren op 12 augustus 1982

te Groningen

Promotores:	Prof. dr. R.J. Bosker Prof. dr. E.G. Harskamp
Beoordelingscommissie:	Prof. dr. A. Desoete Prof. dr. M.P.C. van der Werf Prof. dr. C. Visscher

ISBN printed version: 978-90-367-5543-6

ISBN electronic version: 978-90-367-5542-9

Preface

As an elementary school teacher I noticed that students often have difficulties structuring their thoughts. More than once I heard them say things like ‘I don’t know where to begin’ and ‘I don’t think my answer is right but I don’t know what else to do’. When I was looking for a subject for my master thesis in educational science I thought back to such moments. Luckily, I was invited to work with Michelle Helms on a project about metacognition which gave me insight in the processes underlying such procedural difficulties. After writing my master thesis, Roel Bosker gave me the opportunity to write a PhD proposal to extend the line of research on metacognition. I am very grateful that he gave me this opportunity. After the proposal was approved, Egbert Harskamp guided me through the PhD process as my daily supervisor. Thanks to his vast amount of knowledge about mathematics, my interest for metacognitive processes was concretized in research on word problem solving. It has been very enjoyable to study how students can be supported in such a complex learning domain. I would like to thank Michelle for enthusing me for the topic and Egbert for sharing his knowledge and stimulating me to develop my research ideas as well as my own career as a researcher. Egbert, it has been inspiring to learn from you and to share our ideas about metacognition and about educational science in a broader sense. And Roel, thank you for all your helpful feedback on the manuscripts. Besides thanking my supervisors, I would also like to thank the research assistants who helped with the data collection. And I would like to express my gratitude to all teachers and students who participated in the studies. Thank you for helping me keep focused on the students who educational research should be all about.

In addition to learning from the findings of the studies, writing a dissertation has been a great learning process. From designing the interventions, to contacting schools, to collaborating with colleagues in the field…. throughout the ups and the downs I greatly value all the experiences. I thank my colleagues from the GION and colleagues from other universities who I’ve met along the way for adding to the process with the occasional breaks, laughs and conference outings which made it all so much more fun to do. And Sonja and Vera, thank you for helping me with all kinds of practical tasks. Furthermore, I would like to specifically thank my dear colleagues Mechteld and Coby who will stand beside me during the defense. I am glad to have some “vrollega’s” by my side on such an exciting day. And last but not least I would like to thank my family and friends. The feeling of their support has been encouraging. And of course they always provided ample opportunities for after-work-relaxation. Mom and dad, and my brothers Rob and Peter, thank you for believing in me and always stimulating me to aim for the top in anything I do. And my husband Peter, thank you for being there for me. Your love and trust always keeps me going.

CONTENTS

CHAPTER 1

General Introduction

“I cannot teach anybody anything; I can only make them think.”

- Socrates (470 - 399 BC)

1.1 Introduction

The quote by the classical philosopher Socrates is not far from the stance taken by most educators today; education is not just about teaching students facts, but rather it should focus on “making them think”. Students should be taught to reflect upon their own learning and become active learners who keep developing throughout their lives. This is a constructivist view of learning where students actively create, interpret and organize their knowledge. It requires a form of instruction in which learning processes are guided and not taught in a direct way (Windschitl, 1999). However, in practice, teachers have often misinterpreted constructivism by the idea that they should avoid any form of instruction whatsoever (Davis & Sumara, 2002). This is a worrisome observation. Review studies have shown that unguided learning is ineffective, and probably even negatively influences student outcomes (Kirschner, Sweller, & Clark, 2006; Mayer, 2004). So, before throwing students in at the deep end, we need to hand them enough skills to stay afloat. But what kind of skills do students need to perform well? Firstly, students need basic academic knowledge. But in order to become active learners, students also need to learn how to regulate and monitor their learning. Several reviews show that this is a factor largely influencing academic performance (Dignath & Buettner, 2008; Swanson, 2001; Wang, Haertel, & Walberg, 1990). However, in practice teachers spend little time explicitly guiding students to use adequate procedures to structure their learning process (Inspectie van het onderwijs [Inspectorate of Education], 2008; Kistner et al., 2010). More research is needed on how students can be guided to actively regulate their learning towards enhanced performance.

One important domain which especially requires students to carefully structure their learning processes is the domain of mathematical word problem solving. Word problems are verbal descriptions of problem situations in which mathematical information (usually numerical data) is embedded in text. In elementary education, word problems are frequently used to embed mathematical procedures in a realistic context (Cognition and Technology Group at Vanderbilt, 1997; Gravemeijer, 1997; Stanic & Kilpatrick, 1989). However, embedding mathematical operations in a problem description structurally changes the task. For instance, instead of practicing traditional tasks in grade three like 12 : 3 and 12 - 4, the same mathematical content can be embedded in a word problem such as: “Mother has bought 12 cookies. She told Martin, Sheila and Jasmine to share the cookies equally. Sheila immediately ate all her cookies. Martin and Jasmine saved theirs till later. How many cookies are there still left?” It is clear that when a task is presented in such a way, the solution procedure is more complex than directly applying an algorithm as can be done in the case of the traditional task format. Tasks are defined as problems when students do not have direct access to straightforward ways of finding an answer (Schoenfeld, 1985).

While students could probably solve the tasks 12 : 3 and 12 - 4 independently, solving a complex word problem requires different skills. Researchers stress that the procedural skills required to carefully structure the problem solving process are key determinants affecting performance (Desoete & Veenman, 2006; Desoete, 2009; Fuchs et al., 2010; Harskamp & Suhre, 2007; Schoenfeld, 1992). There are specific skills which come into play in word problem solving. What are these skills and how can we study if students have developed them or not? And can we develop instructional procedures to stimulate problem solving skills?

These questions are at the heart of the studies in this dissertation. The underlying issues behind these questions will be further elaborated below.

1.2 Word Problem Solving and Metacognition

In text books and standardized tests in elementary education a large amount of tasks are presented as word problems (KNAW, 2009; Mullis et al., 2005; OECD, 2003). However, national and international evaluations report that a lot of elementary school students experience difficulties solving word problems. For instance in the most recent TIMSS study of grades four and eight (Trends in International Mathematics and Science Study; Mullis, Martin, & Foy, 2008), it was found that on average across countries, only a median of five percent of the grade four students reached the advanced international benchmark score comprising problem solving abilities. The advanced benchmark is characterized as follows. “Students can apply their understanding and knowledge in a variety of relatively complex situations and explain their reasoning” (p. 68). In grade eight, only a median of two percent reached the advanced benchmark level where they could “organize and draw conclusions from information, make generalizations, and solve non-routine problems” (p. 69). In grade four twenty-six percent of the students reached the high benchmark in which they could “solve multi-step word problems involving operations with whole numbers (p. 68)”. And in grade eight the high level is reached by only fifteen percent of the students. In the Netherlands, the results of a survey in grade six show that in reference to Dutch standards for solving multistep word problems only sixteen percent of the students reached the average performance benchmark (Janssen, Van der Schoot, & Hemker, 2005). Even specifically high achieving students in the Netherlands were reported to have large difficulties solving complex word problems. High performing elementary school students showed a lack of motivation for solving word problems and insufficiently used aids such as note-taking to reach a solution (Van den Heuvel-Panhuizen & Bodin-Baarends, 2004).

Hence, word problem solving is an important, but problematic part of the current mathematics curriculum. Which skills do students need to solve a complex word problem? Students need to know how to apply mathematical knowledge, i.e. the knowledge of mathematical procedures and understanding of concepts, to solve the computation embedded in a text. However, this is not the core of the matter. In fact, although mathematics knowledge is needed, it explains only part of the ability to solve word problems (Fuchs et al., 2008; Mayer & Hegarty, 1996; Swanson & Beebe-Frankenberger, 2004). The main difficulty with word problems is that students have to construct a problem model by making inferences from text (Fuchs et al., 2008; Kintsch & Greeno, 1985). To do so, a student needs to perform a variety of activities to fully grasp the problem. Successful problem solvers generally spend relatively much time using problem solving strategies to analyze a new problem, construct a problem model and choose an appropriate solution plan. Meanwhile they monitor their performance. Conversely, novices were found to jump to calculations before carefully analyzing a problem (Schoenfeld, 1992; Verschaffel et al., 1999a). This shows that one of the most characteristic features of successful problem solving consists of two processes: Regulating and monitoring the use of problem solving strategies. These processes are called metacognition (Flavell, 1979).

In the literature, two categories of applied metacognition are distinguished: (a) metacognitive regulation and (b) metacognitive monitoring (Efklides, 2006; Nelson, 1996). Metacognitive regulation refers to regulating oneself to apply appropriate cognitive strategies to solve a task (Brown & DeLoache, 1978; Veenman, Elshout, & Meijer, 1997). And metacognitive monitoring refers to students' ongoing control over the learning process (Desoete, 2008). Successful problem solvers use a combination of metacognition (metacognitive monitoring and regulation) and cognitive strategies to structure their problem solving. For example, when starting with a complex word problem a student may judge that the problem is difficult to solve in reference to his/her prior knowledge and experiences. This is an example of monitoring behavior. Based on this observation, the student may choose to carefully analyze the problem and select relevant information from the text. The choice to do so is metacognitive regulation, while reading the text is the cognitive result of this choice. Then, when executing a solution plan, the student might notice that something is going wrong and consequently go back to selecting information from the text. The fact that the student notices a mistake is another example of monitoring behavior, which in turn again influences the use of metacognitive regulation of cognitive strategies to solve the problem. As the example explicates, metacognitive processes are conceptually different from cognitive processes, but they always work together in constant interplay. In problem solving, if a child does not have the skill to read a problem, understand the basic mathematical concepts or cannot do the mathematics operations involved in a word problem (these are cognitive strategies one can learn), then metacognition will not be productive. Metacognition can only thrive on a basis of cognitive strategies necessary to solve a problem.

The interplay of metacognitive and cognitive activities is also included in theories of self-regulated learning (SRL). Though theoretically related, SRL theories have a broader focus than metacognitive theory. SRL includes metacognitive and cognitive activities in learning, but also motivational regulation of for instance goal orientation, task value and regulation of emotion (Dinsmore, Alexander, & Loughlin, 2008). Depending on the focus of articles on self-regulated learning, some findings were used to inform theory about metacognition and learning. However, the focal point in this dissertation is specifically on the relation between metacognition and the use of cognitive strategies for problem solving, and not on self-regulation in a broader sense.

1.3 Issues Driving our Research

Referring back to the questions posed in the introduction about the processes involved in the use of metacognition and about how to develop instructional procedures to stimulate problem solving, we can state that there is already a vast knowledge base showing that metacognition is an important determinant of performance. However, the interest for research on students' metacognitive processes in word problem solving is relatively new and the field still is largely 'under construction'. To answer questions about the theoretical underpinnings and measurement of metacognition, and to study if students' metacognition can be used to improve word problem solving, more research is needed. The importance of these issues in reference to the current state of the art in the field will be further elaborated below.

First, a theoretical issue: Although the number of articles on metacognition has been booming - for instance in the 2011 JURE and EARLI conference over ninety presentations dealt with regulation of learning (Wegerif, Myhill, Vickers, Goodall, & Allan, 2011) - the definition of metacognition still remains somewhat 'fuzzy' (Dinsmore et al, 2008; Hacker, 1998; Veenman, Van Hout-Wolters, & Afflerbach, 2006). In the literature, metacognition has previously been related to intelligence (Alexander, Carr, & Schwanenflugel, 1995; Borkowski, Carr, & Pressley, 1987; Campione & Brown, 1978; Sternberg, 1984). This may lead to the question whether metacognition is a variable uniquely influencing performance or if it is merely an expression of students' intellectual abilities causing them to regulate themselves in a more or less sophisticated manner. When aiming to observe or even train metacognition, it is important to study how it is related to performance before treating it as a separate variable. Looking at the conceptual foundations, most intelligence tests used in education primarily measure the general ability to apply previously learned information (Minnaert & Janssen, 1999). Metacognitive measures on the other hand more strongly focus on the procedural activities needed to regulate one’s learning in a particular domain (Greene & Azevedo, 2010). In this sense, metacognition does seem to add something to intelligence. Indeed, in secondary education, most studies found some relatedness, but also some added value of metacognition on top of intelligence as a predictor of performance (Van der Stel & Veenman, 2008). However, little is known about these relationships in elementary education. Furthermore, it is unclear if these relationships are susceptible to influences of students’ cultural background. Students of a migrant background are mostly found to lag behind on various academic measures (Dagevos, Gijsberts, & van Praag, 2003; OECD, 2006). This could among other things be caused by inferior academic experiences or cultural bias in test items. In the literature, it is suggested that when addressing complex items students need to shift from relying on their intellectual abilities towards using metacognitive regulation to structure a task (Prins, Veenman, & Elshout, 2006). Up to date, it is unclear if migrant students use their metacognition in this way and if metacognition has comparable or even higher impact on their academic performance than for native students. If metacognition has comparable predictive validity for native and migrant students, they can communally be included in further studies studying the effect of metacognitive training. However, if migrant students’ metacognition is differentially related to performance, this group should be studied separately. Chapter 2 presents an observational study intended to clarify this issue. To do so, it is studied how native and migrant students’ scores for metacognition are related to scores on different intelligence measures and a standardized test of academic performance.

Secondly, the finding that less skilled problem solvers generally reside to ‘hit and run’ approaches instead of thoughtful regulation of problem solving episodes, raises the issue as to how these students can be trained to structure their word problem solving. Questions about the trainability of metacognition are crucial for the development of instructional theory and practice (Schunk, 2008). This issue was already formulated some decades ago by one of the founding fathers of metacognition research, Flavell (1976, p. 233). He asked: “Is there anything that could be taught that would improve [students’] ability to assemble effective problem solving procedures?” From there on, some studies have already found that training students in a metacognitive approach can benefit their word problem solving. However, the studies which have successfully been executed in elementary education are mostly time-intensive and comprehensive direct instructional approaches (Cognition and Technology Group at Vanderbilt, 1997; Fuchs et al., 2010; Verschaffel et al., 1999a). Little is known about the possibility to teach elementary school students a metacognitive approach by guided training. Guided training is characterized by a combination of an active role of the student and just-in-time instruction the student can choose, for instance in the form of hints or prompts (Mayer, 2004). The fact that there are few studies on guided training in elementary school word problem solving is a major caveat since it has important advantages over training with worked out solution steps determined by the teacher. Firstly, students have a more active role in their own learning processes. Because of this, they learn to make their own instructional decisions (Azevedo, 2005). Secondly, guided training facilitates better transfer of the learned processes to different types of problems (Jonassen, 2003; Moreno, 2006; 2009). And thirdly, giving students a certain level of control over their own learning processes can help them become more motivated to solve problems (Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004). More research is needed about how guided metacognitive training can be used as a tool to efficiently support elementary school students' word problem solving.

A final pressing issue which becomes apparent from the literature is the lack of consensus among researchers about how metacognition should be measured (Veenman, Van Hout-Wolters & Afflerbach, 2006; Winne, 2010; Winters, Greene, & Costich, 2008). To measure students' metacognition, self-report questionnaires are often used. However, self-report instruments have been strongly criticized to be inadequate indicators of students' actual metacognitive behavior and poor predictors of achievement (Desoete, 2007; Greene & Azevedo, 2010; Veenman, 2005; Winters et al., 2008). Students' verbalizations of thought processes on the other hand do seem to have high predictive validity for student achievement. But, using think-aloud measures is complicated and largely time-consuming (Azevedo, Moos, Johnson, & Chauncey, 2010). Researchers and practitioners lack an adequate but yet practical instrument which they can use to measure metacognition in word problem solving. Such an instrument should be able to map the relationship between students' metacognition and problem solving performance. And it should show concurrence with the highly valued think-aloud method (Veenman, 2011a). For the research field, as well as to empower practitioners wanting to document students' metacognitive processes, more research is needed on the measurement of metacognition to come to terms on valid and efficient instruments.

1.4 Overview of the Dissertation

As reported above, issues on the theory, measurement, and training of metacognition in word problem solving are the main focal points of the dissertation. The topics are structured over the different chapters as follows.

In chapter 2, an observational study is reported about the relationship between metacognition, intelligence and performance in elementary education to explore the empirical basis for use of the different constructs. In this study, metacognition was found to have unique predictive validity for academic performance of both native and migrant elementary school students. Therefore we concluded it to be relevant to study if the impact of metacognition on performance can be reinforced by metacognitive training. Consequently, we turn to application of the theory about metacognition and learning by experimentally studying the effects of metacognitive training in word problem solving. A computer program with metacognitive hints and cognitive prompts was developed to support upper elementary school students' problem solving. Chapter 3 reports on the development of the training materials. Also, results of a first study on the effects of the metacognitive training on word problem solving performance are reported. Since highly positive results of metacognitive training were found in reference to a business as usual condition, it was decided to move on with the training program in a next study. Chapter 4 reports on the second study on metacognitive training in word problem solving. In this study, results of a metacognitive training group were compared to results of a control group working with a computer program with the same word problems, but without metacognitive hints. Using this more stringent experimental design, again positive results of metacognitive training on metacognition as well as word problem solving performance were found. When executing the studies, we noticed that metacognitive outcomes are not easy to measure. Therefore, in chapter 5 an explorative study is presented in which a newly developed measurement instrument designed to measure metacognition in word problem solving was compared to a think-aloud method and a self-report metacognition questionnaire. Finally, the findings and conclusions from the different chapters are brought together in a general discussion in chapter 6.

CHAPTER 2

The Relationship between Intelligence, Metacognition and Performance; Does Metacognition add to the Equation?

ABSTRACT

Metacognition has been identified as an important predictor of academic performance. However, the question rises if metacognition has unique significance for performance or if it is merely an expression of students' intellectual abilities. In secondary education, metacognition was already shown to predict students' performance beyond the influence of intelligence. But little is known about the impact of spontaneous use of metacognition in elementary education. Also, little is known about the influence of cultural background in students' use of metacognition. To study these issues an observational study is performed in grade six (*n* = 103). Our findings suggest that the construct metacognition is comparable for native and migrant students. Secondly, analyses of covariance show significant main effects of crystallized and fluid intelligence, but also of metacognition as a predictor of performance on a standardized test with reading comprehension tasks and word problems. A higher level of metacognition was found to benefit most students, additional to the effect of intelligence. We conclude metacognition to be important for native as well as migrant students. This finding can be of significant interest for intervention studies.

This chapter is based on the published book chapter:

Helms-Lorenz, M. & Jacobse, A. E. (2008). Metacognitive skills of the gifted from a coss-cultural perspective. In M. F. Shaughnessy, M. V. J. Veenman, & C. Kleyn-Kennedy (Eds.), *Meta-cognition: A recent review of research, theory, and perspectives* (pp. 3-43). Hauppage: Nova Science Publishers.

2.1 Introduction

In the past decades, research on metacognition has been steadily on the rise. Metacognition refers to students' ability to regulate and monitor their own cognitive activities while learning (Flavell, 1979). For instance, when performing a mathematical word task, students may regulate themselves to read the task well, analyze the problem, make a calculation plan, execute the plan and verify their answer (Schoenfeld, 1992). In such behavior, the activity in itself (i.e. writing down a calculation) is cognitive in nature, but the choice to do so is metacognitive. Metacognition is a particularly relevant variable for education since it has been found to largely influence academic performance (Wang et al., 1990). The importance of metacognition for academic performance has been reported in various domains such as mathematics and comprehensive reading (Mayer, 1998; Pressley et al., 2001; Roeschl-Heils, Schneider, & van Kraayenoord, 2003; Schoenfeld, 1992; Verschaffel et al., 1999a). However, more work is needed to position the construct between other variables influencing performance. Up to now, researchers have argued the concept of metacognition to be quite 'fuzzy' or vague (Dinsmore et al., 2008; Hacker, 1998). Defining the boundaries between metacognition and other variables can shed more light on its unique importance for educational research and practice.

To do so, an important step is to distinguish metacognition from another key variable influencing academic performance in education: Intelligence. Intelligence as measured by intelligence tests generally has moderate to high predictive validity for academic performance, depending on the IQ measure used in the study and the sample (Sternberg, Grigorenko, & Bundy, 2001). Some researchers have suggested there to be a relation between students' intelligence and their metacognition (Alexander, Carr, & Schwanenflugel, 1995; Borkowski, Carr, & Pressley, 1987; Campione & Brown, 1978; Sternberg, 1984). Findings from a review study for instance showed that highly intelligent students had more knowledge about metacognitive processes and started to use more sophisticated metacognitive regulation than students with lower intelligence around secondary school ages (Alexander et al., 1995). The question rises if metacognition is a variable uniquely influencing performance or if it is merely an expression of students' intellectual abilities.

When looking at the conceptual foundations, most intelligence tests used in education primarily measure the general ability to apply previously learned information (Minnaert & Janssen, 1999). Metacognitive measures on the other hand more strongly focus on the procedural activities needed to regulate one's learning in a particular domain (Veenman, 2005). In this sense, metacognition does seem to add something to intelligence. Indeed, there have been studies which found that metacognition has added value beyond intelligence as a predictor of performance (Veenman et al., 2006). However, most of these studies were executed in secondary and higher education. There are few observational studies assessing the relationship between metacognition, intelligence and school performance in the main academic domains of reading and mathematics in elementary education. Main aim of the present study is to analyze this relationship and discern if the statement that metacognition has its own importance as a variable affecting performance proves to be true for an elementary school sample.

The second aim is to analyze if measures for metacognition, intelligence and performance are structured differentially for native and migrant students. First and second generation migrant students in most countries generally perform under the level of their native counterparts (Dagevos et al., 2003; OECD, 2006). This is a very pressing issue since it affects a lot of students. In the year 2007/2008, about fourteen percent of elementary school students in the Netherlands were of non-western origin. Moreover, it is estimated that the proportion of non-Western migrants in elementary and secondary education will increase to about twenty percent by 2020 (Centraal Bureau voor Statistiek [Central Statistics Agency], 2003). Metacognition could be one of the sources which could help migrant students overcome performance gaps. Evidently, students need cognitive skills to reach a certain level of performance. But when other factors like comprehension of task situations or language complicate migrant students' learning, metacognition may help them to better structure complex tasks in order to support their performance (Prins et al., 2006). It would be interesting to observe if migrants' use of metacognition is related to their academic performance and if it has comparable or even higher impact on their performance than for native students. If a substantial relationship between metacognition and performance beyond intellectual abilities can be established, this could be an important pointer in the direction of supporting migrant students' learning processes.

The relationship between metacognition, intelligence and academic performance in elementary education is assessed in an observational study in grade six. First, the theory about these constructs and their relations in school settings is discussed in paragraph 2.2.

2.2 Theoretical Framework

2.2.1 Metacognition

The knowledge and regulation of one's thought processes is referred to as metacognition. In the pioneering work of Flavell (1979, p.906) the concept was characterized as "knowledge and cognition about cognitive phenomena" which has also been described to in terms such as 'thinking about thinking' or 'higher order cognition of cognition' (e.g. Alexander et al., 1995; Hacker, 1998; Veenman et al., 2006). Flavell (1979) defined metacognition as consisting of components of knowledge, monitoring of experiences and goals, and regulation. For instance, when solving a mathematics problem, a student will have to *know* that he/she can make a sketch to get an overview of the problem, the student will then *monitor* whether he/she needs to perform this action and then he/she has to *regulate* the actual execution of this action.

Although knowledge about metacognitive processes is important, metacognitive monitoring and regulation have the most direct influence on behavior and are thus major variables affecting performance (Veenman, Kok, & Blöte, 2005; Wang et al., 1990). For instance, in mathematics, metacognitive regulation and monitoring activities were found to be key determinants for successfully solving word problems (Mayer, 1998; Schoenfeld, 1992; Verschaffel, De Corte & Vierstraete, 1999b). Also, in other domains such as comprehensive reading, the importance of metacognition for performance has been shown in various studies (Pressley et al., 2001; Roeschl-Heils et al., 2003). There might be some metacognitive processes which can be used across domains. But, since metacognition is about the regulation of cognitive activities, metacognition primarily has a domain-specific character informing the processes students use when performing specific tasks.

2.2.2 Intelligence

The close connection between metacognition and cognitive skills in a particular domain differs from the more general abilities measured in intelligence tests. Intelligence tests attempt to measure some common ability characteristic which has been labeled the ‘g factor’ (Spearman, 1927). Spearman discovered that results of different ability tests tend to positively correlate with one another. He called this the ‘positive manifold’ phenomenon. From repeated findings on this phenomenon, Spearman concluded there to be a general intelligence factor (*g*), which represents what all valid intelligence tests have in common. Later on, researchers as Burt (1949), Horn and Cattell (1966) and Vernon (1950) discovered that besides this general factor, there are some more specific factors which come in to play in different intelligence tests. In the distinction of characteristics of different intelligence tests, researchers mostly differentiate between fluid and crystallized intelligence. Measures of fluid intelligence are new and unknown to the learner and measure one’s ability to reason and to solve problems. Measures of crystallized intelligence on the other hand, measure the ability to apply previously learned information. In a recent model of intelligence, the ‘Cattell–Horn–Carroll theory of cognitive abilities’ (McGrew, 2009) the distinction between crystallized and fluid intelligence has even been further extended with other types of cognitive abilities. For instance, within crystallized abilities researchers have discerned reading and writing abilities and mathematical knowledge (Flanagan, Ortiz, Alfonso, & Mascolo, 2002). However, the broad categories of crystallized and fluid intelligence stay intact.

An example of a test categorized as a more fluid measure is the Raven Standard Progressive Matrices test, which assesses non-verbal logical reasoning abilities in completing series of figures. Comparing the scores on fluid measures to measures of academic performance, mostly low to moderate correlations are reported (Raven, Court, & Raven, 1996). Measures of crystallized intelligence on the other hand, measure the ability to apply previously learned information (Minnaert & Janssen, 1999). Such measures may be more susceptible to background influences such as culture and learning experiences than fluid intelligence measures (Helms-Lorenz, Van de Vijver, & Poortinga, 2003). Crystallized intelligence tests are typically strongly related to achievement measures used in education. For instance, the NIO test (Van Dijk & Tellegen, 2004) was shown to correlate largely to a standardized achievement test (CITO) in the Netherlands.

2.2.3 Relationship between metacognition, intelligence and performance

To answer the question whether metacognition adds to performance on top of intelligence, a first step is to assess the underlying theoretical models of the relationship between the three variables. Veenman and colleagues (Veenman et al., 1997; Veenman & Spaans, 2005; Veenman et al., 2005; Veenman, Wilhelm, & Beishuizen, 2004) and Minnaert and Janssen (1999) have described the relationship between intelligence, metacognition and performance using three hypothetical models. The first model, referred to as the intelligence model, views metacognition as an integrated part of intellectual ability. Metacognition according to this view cannot predict learning independent of intellectual abilities because the two are completely interwoven. In a second, contrasting model, the independency model, intelligence and metacognition are seen as completely independent predictors of learning. Finally, the mixed model suggests that metacognition and intellectual abilities are related to some extent, but metacognition also has surplus predictive validity for learning outcomes over and above this relationship. In Figures 2.1 to 2.3 schematic representations of the three models are presented.

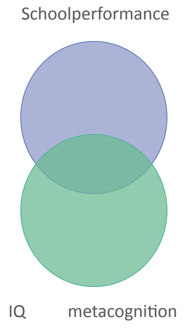


Figure 2.1 Intelligence model

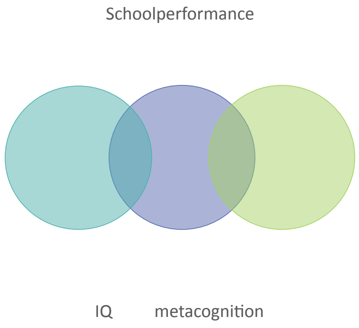


Figure 2.2 Independency model

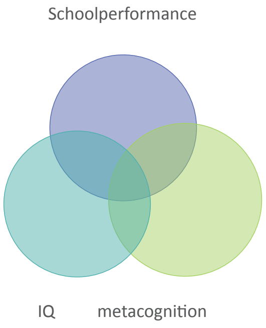


Figure 2.3 Mixed model

Initial support for the independency model was reported by Swanson (1990). In this study, students with self-reported high metacognition outscored students with low metacognition on two Piagetian problem solving tests, irrespective of their ability level. However, follow-up studies showed that metacognition was only partially independent of intelligence which is more in line with the mixed model (Swanson, 1992). Most other studies in secondary education have found support for the mixed model in which metacognition and intelligence have some overlap, but also some unique predictive validity for performance. Studies of Veenman, Elshout and Meijer

(1997) and Minnaert and Janssen (1999) showed that metacognition measured by think-alouds and a questionnaire respectively, attributed to performance over and above intellectual abilities. In the study of Veenman et al. a crystallized measure was used while Minnaert and Janssen found comparable patterns for both a fluid and a crystallized intelligence measure. For mathematical problem solving performance, a comparable finding was found for secondary education students. Scores on a crystallized intelligence test overlapped about twenty percent with metacognitive regulation measured by logfiles and systematical observation of think-aloud protocols. But, metacognition also uniquely explained about fourteen percent of the variance in problem solving scores, and intelligence uniquely explained about nine percent (Veenman & Spaans, 2005). In another study, the impact of metacognition respectively intelligence on biology and geography tasks in elementary and secondary education was assessed (Veenman et al., 2004). The results show a relationship between intelligence measured by a crystallized measure and performance. Moreover, metacognition uniquely affected performance in nearly all grades. Summarizing findings from different studies, on average intellectual ability was found to account for approximately ten percent of variance in learning, the proportion of overlap between intellectual ability and metacognition together predicted about twenty percent of the variance, and metacognition generally had a surplus value in predicting variance in learning performance of around seventeen percent (Veenman et al., 2006).

However, in elementary education, there are few studies explicitly addressing the relationship between intelligence, metacognition and performance. The aforementioned study of Veenman et al. (2004) showed that in grades four and six metacognition had unique value for performance in biology and geography. But on the other hand, Desoete (2008) found in her study that in grade three intelligence and metacognition measured by think-aloud protocols mainly overlap as predictors of math performance in line with the intelligence model. More research is needed to clarify if students as young as of an upper elementary school age can already use their metacognition independent of intelligence to benefit their performance in important academic domains such as math and reading.

2.2.4 Measurement issues

When studying the relationship between metacognition, intelligence and performance, it is likely that findings will be influenced by the measures used. As argued above, the type of intelligence measure can influence the degree to which intelligence is related to the academic performance. It is expected that crystallized intelligence measures are more strongly related to academic performance than fluid performance measures since they measure abilities that are taught in school. However, crystallized and fluid intelligence measures do not seem to show substantial differences in their relation to metacognition (Minnaert & Janssen, 1999). Secondly, the performance measure which is used can influence the findings. More complex measures are assumed to trigger metacognitive behavior more than other measures (Prins et al., 2006). Performance measures for which students use atomized, routine-like behaviors are not suitable for the use of metacognition and will as such probably show low correlations with metacognition. Lastly, the way metacognition is being measured could influence the findings. In studies on the relationship between metacognition, intelligence and performance mostly think-aloud measures are used. In think-alouds, students are asked to verbalize their thoughts while performing a learning task. Collecting think-alouds is best undertaken with learning tasks that are comparable to the tasks in the performance measure (Pajares & Miller, 1995).

2.2.5 Cultural background

Yet another variable which can influence students' performance on different measures is their cultural background. Cultural background may influence performance in various ways. Firstly, performance in learning is determined not only by genetic factors, but by an interplay of cognitive, affective and social factors. Factors from students' out-of-school environment such as supporting interactions, stimulation of learning, a variety of experiences, freedom and materials to play and experiment and emotional support can influence performance (Freeman, 1993). Secondly, native and migrant students may differ in mastery of the language used in school and cultural knowledge about their country of residence. Because of language and cultural loading, test instructions and item phrasing can unintentionally cause bias (Van de Vijver & Poortinga, 1997). For instance, in the most frequently used standardized test in grade six in the Netherlands, some items were found to be culturally biased, disadvantaging migrant students (Uiterwijk & Vallen, 2003). It is thus likely to expect migrant students to perform somewhat lower than natives on tests of crystallized intelligence and academic performance.

Concerning the question how migrant students specifically perform on metacognitive measures, little information is available. One study in secondary education in the Netherlands showed that migrant students claim to use more metacognition than native students (Blom, Hoek, & Ten Dam, 2007). However, it has not been observed if this is actually the case. Moreover, no information is available about how metacognitive-, intelligence- and performance measures are related for migrant students. As stated in the introduction, it could be that migrant students have learnt to rely more on their metacognition in academic tests than native students, to compensate for the problems they experience due to inferior understanding of test items. If this is true, metacognition may have a somewhat larger effect on migrant students' performance than for natives. This could explain the statements of migrant students in the study of Blom, Hoek and ten Dam. However, this question is yet to be answered.

2.3 Research Questions

Summarizing, most evidence up to date seems to point in the direction of metacognition having a moderate relation with intelligence but also some unique predictive validity for performance in line with the mixed model. We expect a comparable picture in our elementary school sample. Comparing both a crystallized and a fluid intelligence test, we hypothesize the crystallized measure to be most strongly related to academic performance. Due to influences of students' background, migrant students are expected to score lower than native students on intelligence and performance tests. But, it is suggested that metacognition may be more strongly related to performance for migrant students because they might use monitoring and regulation processes to compensate for difficulties they experience in test situations.

These hypotheses will be assessed by means of three research-questions:

1. Do different intelligence measures, a metacognitive measure and a performance measure show comparable results for native and migrant students?
2. How are metacognition, intelligence and academic performance related in the elementary school sample?
3. Does metacognition have additive predictive value for academic performance on top of intelligence?

2.4 Method

2.4.1 Sample

The sample of the study consists of 103 students, 57 boys and 46 girls. All subjects were in their final elementary school year. Students had an average age of 12.41 years old (*SD* = .53). The students were from schools with a large proportion of migrant students, mostly in or near large urban areas in the Netherlands. In the sample, the majority of 69 students were migrants, but also 34 students of native origin which were in these classrooms were taken up in the analyses. Students were classified as 'migrants' if at least one of their parents was born in a non-western country. Of the migrant students, 11 were born abroad (2 in an African country, 3 in Suriname, 2 in an Asian country, 2 in Afghanistan, 2 in Iraq) and of 3 subjects the birth country was unknown. The largest group of migrants, 43 students, had parents who were both of non-western origin. And 11 migrant students had one parent who came from a non-western country. Migrant students' parents were mostly from an African country (23% of the fathers and 22% of the mothers) or from Suriname (18% of the fathers and 17% of the mothers).

2.4.2 Measurement instruments

Intellectual ability. Two intelligence tests were administered: the Raven Standard Progressive Matrices (Raven et al., 1996) and the NIO (Nederlandse Intelligentietest voor Onderwijsniveau [Dutch Intelligence test of Educational level], Van Dijk & Tellegen, 2004). These tests were administered in groups (school classes).

The NIO primarily focuses on learned abilities and can thus be categorized as a crystallized measure of intelligence (Van Dijk & Tellegen, 2004). The NIO presents multiple-choice tasks in the domains of language, math and spatial abilities. NIO results can be split up in sumscores for verbal and symbolical tasks which match with the subcategories of reading abilities and quantitative and visual abilities of the Cattell-Horn-Carroll Theory classification (Flanagan et al., 2002). The verbal tasks test students' knowledge of word meaning, the meaning of sentences and text comprehension. The symbolical tasks on the other hand test students' mathematical knowledge. Two translated sample items of the NIO are presented below. The left sample item is a verbal task, while the right item is representative for the symbolic part of the test.

Which word means the same as the word in bold print? *brave 1 light 2 cold 3 fear 4 courageous 5 new*	*Which sign x or : or + or – must be inserted to complete the calculation?* *8 … 2 = 10*

The Raven test on the other hand measures nonverbal reasoning abilities and is seen to be especially valuable in evaluating students who are limited in the dominant language skills (Raven, 2000). In the Raven test, students have to complete missing pieces from a series of 60 images. The images are of increasing complexity throughout the test. An example of an image from the Raven test is added in Figure 2.4. Although the test requires visual abilities, it primarily measures the ability of students to reason logically with new information and should thus be classified as a fluid measure of intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002).

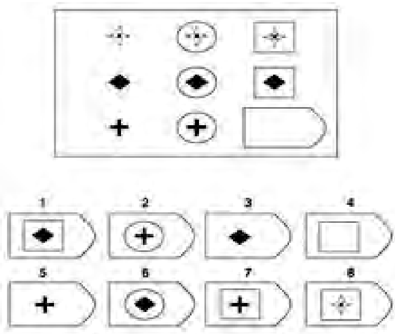


Figure 2.4 Sample item of the Raven Standard Progressive Matrices

In our sample, the NIO has a good internal consistency between the various subscales of $\alpha = .94$. The sixty items of the Raven test also have high internal consistency with an α of .92. Students scored a mean of 44.72 (*SD* = 6.49) on the Raven test, which corresponds to the 50th percentile of the standardized scores for the Netherlands (Raven et al., 1996). Students scored an average of 95.76 (*SD* = 17.00) points on the NIO test which is just a little lower than the standardized mean score of 100 in the population (Van Dijk & Tellegen, 2004).

Academic performance. As academic performance measure, scores on the standardized grade 6 Elementary School Leaving test of the National Institute for Educational Measurement (CITO) were used. This standardized test for school performance is used in over 80% of the Dutch school as an indicator of students' academic abilities at the end of elementary school (Citogroep Primair Onderwijs [Cito group elementary education], 2007). Most schools administer this test around January in the final elementary school year. The test has a language, mathematics, and an information section. A major part of 87 percent of the test requires reading comprehension (in both the language and the information section) or word problem solving (mathematics section). The tasks are novel but are in line with the domain-specific knowledge and abilities taught in school.

Scores on the test range from 501 to 550. Students in the sample scored a mean of 531.31 (*SD* = 11.35) which is a little lower than the general mean score in the Netherlands of 535 points (Citogroep Primair Onderwijs [Cito group elementary education], 2007).

Metacognition. To measure students' metacognition, a think-aloud measure is used. In this type of measurement, students are instructed to perform a task whilst verbalising their thought processes. Measuring metacognition with think-aloud protocols has the advantage that little information about metacognitive processes is lost since thoughts are directly recorded (Veenman, 2005). Using think-alouds is suitable for novice- as well as advanced learners as long as the task used is complex enough to prevent automatic execution (Prins et al., 2006). In our study, think-aloud protocols of students performing a mathematical word problem were recorded on video. Students performed the word problem individually. The only support the test leader was allowed to give was to encourage students to keep thinking aloud when they silenced (Ericsson & Simon, 1993).

The recorded think-aloud protocols of each student were assessed by means of a scoring rubric developed for systematical observation of think-aloud protocols in mathematical problem solving (Veenman, Kerseboom, & Imthorn, 2000). The scoring scheme in Table 2.1 distinguishes metacognitive regulation of the problem solving episodes of reading, analysing and exploring the task (orientation), planning and executing a solution (systematical orderliness) and verifying the answer (evaluation and reflection). Metacognitive monitoring is observed in items 8, 9 and 11 of the scoring scheme.

Students' activities were rated on a five-point scale ranging van '0' (activity not present), through '1' (small initiation of activity), '2' (activity partly present), and '3' (activity present, but not executed to the fullest) to '4' (activity fully present). In order to enhance reliability, two judges performed the assessment together, arguing until agreement was reached (c.f. Veenman et al., 2005; 2000; 2004). Sum scores were calculated for total instrument as well as for different problem solving episodes in the instrument. Students in the sample scored a mean of 12.09 points (*SD* = 5.86) on the instrument. The reliability of the instrument was $\alpha = .68$.

Table 2.1

Scoring scheme for systematical observation of think-aloud protocols in word problem solving

	Read, analyse and explore (orientation)					
1	Entirely reading the problem statement (as incomplete task analysis leads to trial-and-error behaviour);	0	1	2	3	4
2	Selection of relevant data (task analysis);	0	1	2	3	4
3	Paraphrasing of what was asked for (task analysis and goal setting);	0	1	2	3	4
4	Making a drawing related to the problem (task analysis);	0	1	2	3	4
5	Estimating a possible outcome (goal setting);	0	1	2	3	4
	Plan and implement (systematical orderliness)					
6	Designing an action plan before actually calculating (planning);	0	1	2	3	4
7	Systematically carrying out such plan (to avoid haphazard behaviour);	0	1	2	3	4
8	Calculation correctness (avoid sloppiness);	0	1	2	3	4
9	Avoiding negligent mistakes (such as inattentively switching numbers);	0	1	2	3	4
10	Orderly note-taking of problem solving steps (in order to keep an overview of problem solving steps and create an opportunity for checking outcomes);	0	1	2	3	4
	Verify (evaluation and reflection)					
11	Monitoring the on-going process;	0	1	2	3	4
12	Checking the answer;	0	1	2	3	4
13	Drawing a conclusion (recapitulating);	0	1	2	3	4
14	Reflecting on the answer (referring to the problem statement);	0	1	2	3	4
15	Relating to earlier problems solved (reflection with the aim to learn from one's experiences).	0	1	2	3	4

In our study, think-aloud protocols are collected of students performing a mathematical word problem. The academic performance measure we used mainly comprises of problem solving tasks as well as comprehensive reading tasks (see previous subparagraph). Since generally high correlations are found between word problem solving and comprehensive reading (Vilenius-Tuohimaa, Aunola, & Nurmi, 2008), we expect metacognition measured with a problem solving task to be related to this performance outcome measure. However, since the performance measure does cover a more broad range of tasks than used in think-aloud measurement, the relationship may be moderate.

Information on individual background. Information on background variables concerning age, gender and culture where gathered from the participating schools' administration.

2.5 Results

2.5.1 Descriptives of the instruments for native and migrant students

On the performance measure, migrant students in our sample were found to perform significantly lower (*M* = 528.38, *SD* = 10.20) than native students (*M* = 537.55, *SD* = 11.28; *t*(95) = 3.98, *p* = .00). The pattern of migrant students performing lower than native students was also found for both intelligence measures. For the crystallized measure (NIO), migrant students had an average score of 90.06 points (*SD* = 14.55) while native students on average scored 106.47 points (*SD* = 16.25). For the nonverbal fluid measure (Raven), migrant students surprisingly also have a lower average with a mean of 43.52 (*SD* = 6.02) versus native students' average score of 47.09 (*SD* = 6.82). Differences were significant for both the NIO (*t*(96) = 5.10, *p* = .00) as well as the Raven measure (*t*(99) = 2.69, *p* = .01). But, native and migrant students did not differ significantly in the amount of metacognition used in the think-aloud measurement. The mean score of natives is 13.12 (*SD* = 6.47) and of migrants 11.58 (*SD* = 5.52) (*t*(101) =1.26, *p* = .21).

Exploratory factor analysis was performed to determine the structure of the cognitive and metacognitive measures in our sample. One component was extracted for the native as well as the migrant sample. The component of the natives has a total eigenvalue of 4.499 explaining 75 percent of the variance, and for the migrant group a total eigenvalue of 3.506 explaining 58 percent of the variance was found. Assessment of factorial agreement by calculation of Tucker's phi (Tucker, 1951), shows that the factors of both the native and the migrant group are comparable with a coefficients per factor of .95 and .94.

Table 2.2

Principal component analysis (rotation varimax): Factor loadings for natives (left) and migrants (right)

	Component 1
Performance	.916
RAVEN IQ	.879
NIO IQ	.970
NIO IQ verbal	.859
NIO IQ symb	.896
Mc total	.637

	Component 1
Performance	.826
RAVEN IQ	.384
NIO IQ	.970
NIO IQ verbal	.913
NIO IQ symb	.872
Mc total	.376

Note. Performance = performance on CITO test; RAVEN IQ = total Raven score; NIO IQ = total NIO score; NIO IQ verbal = verbal subtest of NIO; NIO IQ symb = symbolic subtest of NIO; Mc total = sumscore metacognition

When forcing the extraction of two components, the rotated component matrices obtained by varimax rotation with Kaizer normalization, shows the following picture presented in Table 2.3. Here, the second factor has an eigenvalue of .684 explaining 11 percent of the variance for the natives, and .968 explaining 16 percent for migrants.

Table 2.3

Explanatory factor analysis: Component matrix for natives (left) and migrants (right)

	Component	
	1	2
Performance	.857	.330
RAVEN IQ	.796	.372
NIO IQ	.945	.272
NIO IQ verbal	.904	.09
NIO IQ symb	.807	.390
Mc total	.252	.952

	Component	
	1	2
Performance	.768	.305
RAVEN IQ	.188	.606
NIO IQ	.977	.151
NIO IQ verbal	.913	.159
NIO IQ symb	.886	.116
Mc total	.08	.863

Note. Performance = performance on CITO test; RAVEN IQ = total Raven score; NIO IQ = total NIO score; NIO IQ verbal = Verbal subtest of NIO; NIO IQ symb = Symbolic subtest of NIO; Mc total = sumscore metacognition

When a division over two components takes place, for the native sample the variables loading more strongly on the first factor are the crystallized (NIO) and the fluid (Raven) intelligence measure as well as the performance measure. The contribution of metacognition to the first factor is relatively small. In the second extracted component of the native group, metacognition has the greatest contribution. This is also true for migrant students in which the performance measure and the crystallized measure primarily load on the first component. The metacognitive measure and the fluid intelligence measure mainly load on the second factor. Referring to findings for both native and migrant students, we could label the first factor as a more cognitive academic factor while the second factor is more metacognitive in nature. This shows that the measures are structured in approximately the same way for native and migrant students. However, placing of the fluid intelligence measure Raven is ambiguous over the two groups.

2.5.2 Relationship between metacognition, intelligence and performance

In Table 2.4 bivariate correlations of the native and the migrant sample between the performance, ability, and metacognitive measures are presented. As expected, the crystallized measure of intelligence (NIO) which largely consists of language and mathematical items was strongly related to academic performance on the standardized test. This was true for both native and migrant students with correlations ranging from $r(59) = .59$ to $r(29) = .86$. However, the fluid measure (Raven) was less strongly related to performance for migrant than for native students. Referring to the content of the test it would be expected that the fluid measure correlates moderately or low with the performance measure. Since it is a non-verbal measure it is unclear why this finding varies this much between the subsamples.

For both native and migrant students, metacognition was significantly related to performance (*r*(29) =.52 respectively *r*(64) =.36). This was true for metacognition executed over different episodes of the problem solving process. Concerning the relation between metacognition and intelligence, it becomes clear that for natives there were significant positive relationships between crystallized and fluid intelligence measures on the one hand and metacognition on the other hand. Metacognition is related to both the verbal and the symbolic part of the NIO test, showing that metacognition is associated with language and mathematical abilities. For migrant students, correlations between the crystallized test and metacognition measure were close to significant. Splitting up metacognitive behaviour over the subscales shows that for native and migrant students higher intelligence scores are mainly related to metacognitive regulation in the planning episode of problem solving.

In order to assess the unique impact of metacognition and intelligence on performance, analyses of variance were performed with the performance measure as the dependent variable, native versus migrant background as an independent factor and metacognition and intelligence as continuous predictor variables. Analyses show that the crystallized NIO measure explained the largest part of the variance in performance. This was the case for both the total NIO test (*F*(1,88) = 98.53, *p* = .00) as well as for the verbal subtest (*F*(1,88) = 85.68, *p* = .00) and the symbolical subtest (*F*(1,88) = 50.88, *p* = .00). The fluid measure (Raven) also had a significant main effect on performance, although the effect was smaller (*F*(1,90) = 16.76, *p* = .00). Controlling for intelligence, a unique main effect of metacognition was found additional to all intelligence measures. Metacognition explained an unique part of the variance besides the total NIO test (*F*(1,88) = 5.55, *p* = .02), the verbal NIO subtest (*F*(1,88) = 9.66, *p* = .00), the symbolical NIO subtest (*F*(1,88) = 6.21, *p* = .02), and the fluid Raven test (*F*(1,90) = 9.87, *p* = .00). The main effect of students' cultural background on performance was not significant in any of the models except the model controlling for fluid intelligence (Raven).

Interactions between all independent variables were assessed. Firstly we wanted to determine if the predictive value of metacognition for performance is affected by students' background. No significant interactions between culture and metacognition were found controlled for the NIO total test (*F*(1,85) = .56, *p* = .46), the NIO verbal test (*F*(1,85) = 1.51, *p* = .22), the NIO symbolical test (*F*(1,85) = .32, *p* = .57), nor the Raven test (*F*(1,88) = .36, *p* = .55). This shows that metacognition explained a significant part of the variance of both native an migrant students' performance. For the NIO, there was an interaction between intelligence and metacognition (*F*(1,85) = 7.03, *p* = .01). But, besides this interaction, the main effects of intelligence and metacognition remain intact.

Table 2.4

Correlational matrix of bivariate correlations between performance, ability, and metacognitive measures for natives (below the diagonal) and migrants (above the diagonal)

	1	2	3	4	5	6	7	8	9
Performance (1)	--	.25*	71**	.70**	.59**	.36**	.24	.30*	.28*
p		.05	.00	.00	.00	.00	.05	.01	.02
N		64	61	61	61	66	66	66	66
RAVEN IQ (2)	.77**	--	.31*	.29*	.29*	.20	.01	.27*	.13
p	.00		.02	.02	.03	.10	.92	.03	.31
N	31		62	62	62	67	67	67	67
NIO IQ (3)	.86**	.79**	--	.94**	.91**	.24	.10	.25*	.18
p	.00	.00		.00	.00	.05	.42	.05	.17
N	31	34		64	64	64	64	64	64
NIO IQ verbal (4)	.78**	.68**	.92**	--	.71**	.21	.04	.24	.18
p	.00	.00	.00		.00	.10	.78	.06	.16
N	31	34	34		64	64	64	64	64
NIO IQ symb (5)	.78**	.77**	.91**	.67**	--	.24	.17	.22	.15
p	.00	.00	.00	.00		.06	.19	.09	.24
N	31	34	34	34		64	64	64	64

Mc total (6)	.52**	.52**	.54**	.44*	.55**	--	.71**	.83**	.74**
p	.00	.00	.00	.01	.00		.00	.00	.00
N	31	34	34	34	34		69	69	69
Mc Read & Explore (7)	.35	.39*	.40*	.33	.42*	.54**	--	.34**	.37**
p	.06	.02	.02	.06	.01	.00		.00	.00
N	31	34	34	34	34	34		69	69
Mc Plan & Implement (8)	.51**	.46**	.45**	.39*	.43*	.87**	.19	--	.41**
p	.00	.01	.01	.02	.01	.00	.28		.00
N	31	34	34	34	34	34	34		69
Mc Verify (9)	.26	.33	.35*	.23	.40*	.79**	.30	.51**	--
p	.15	.06	.04	.19	.02	.00	.09	.00	
N	31	34	34	34	34	34	34	34	

Note. Performance = performance on CITO test; RAVEN IQ = total Raven score; NIO IQ = total NIO score; NIO IQ verbal = verbal subtest of NIO; NIO IQ symb = symbolic subtest of NIO; Mc total = sumscore metacognition; Mc Read & Explore = metacognition in the reading and analysing episode; Mc Plan & Implement = metacognition in the planning and implementation episode; Mc Verify = metacognition in the verification episode.

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Table 2.5

Differences in mean scores on the performance measure of the below average and the above average metacognition group over different types of intelligence measures

Type and level of Intelligence		Performance low Mc group	Performance high Mc group	*t-value*
Crystallized:	Low IQ (*n*=21+9)	519.05 (7.48)	526.11 (8.04)	-2.32*
NIO IQ	Average IQ (*n*=16+15)	528.94 (6.06)	530.60 (6.45)	-0.74
	High IQ (*n*=13+23)	537.08 (10.87)	543.39 (6.34)	-1.92*
Crystallized:	Low IQ (*n*=21+10)	518.90 (6.49)	527.20 (8.90)	-2.95**
NIO IQ verbal	Average IQ (*n*=17+16)	531.24 (7.50)	531.25 (6.89)	-0.01
	High IQ (*n*=12+21)	534.75 (12.26)	543.81 (5.98)	-2.40*
Crystallized:	Low IQ (*n*=21+9)	519.38 (7.63)	527.33 (7.21)	-2.66**
NIO IQ symb	Average IQ (*n*=15+16)	530.07 (7.41)	532.94 (10.26)	-0.89
	High IQ (*n*=14+22)	534.79 (11.28)	541.77 (6.91)	-2.08*
Fluid:	Low IQ (*n*=24+11)	523.21 (9.86)	533.36 (9.87)	-2.83**
Raven IQ	Average IQ (*n*=16+12)	525.19 (7.94)	531.58 (9.32)	-1.96*
	High IQ (*n*=8+20)	538.50 (9.99)	539.42 (9.41)	-0.25

Note. Performance = performance on CITO test; RAVEN IQ = total Raven score; NIO IQ = total NIO score; NIO IQ verbal = verbal subtest of NIO; NIO IQ symb = symbolic subtest of NIO

**p<.01 *p<.05 (one-sided)

Table 2.5 illustrates how having below average versus above average metacognition affects performance scores of students of comparable intelligence. Three intelligence groups of about one third of the sample were computed for each intelligence measure to analyze differences in performance of students of comparable intelligence following the method of Minnaert and Janssen (1999). The results in Table 2.5 show that especially the low ability students, but also the high ability students who use more metacognition to regulate and monitor their learning activities achieve better mean scores on the academic test than comparably intelligent students with low metacognition. Average ability students benefit too, but this effect was only found for the fluid measure. Significant differences of the high versus the low metacognition group range from 13 to 21 percent better scores on the performance measure.

The fact that metacognition was found to be related to performance, but that metacognition also uniquely affects the academic performance of both native and migrant students of comparable intelligence, is in line with the mixed model.

2.6 Conclusion and Discussion

Researchers have suggested metacognition to be a key factor in academic performance. In order to determine the significance of studying this concept, it is important to establish how metacognition is related to another important predictor of performance: Intelligence. If metacognition is related to performance and does not show complete overlap with intelligence measures, it could be a variable of interest for research as well as for practical applications in education. In a theoretical sense, exploring the boundaries of the construct metacognition can help to clarify its definition. And on a more practical note, the question if metacognition has unique predictive validity for performance can give insight in whether it would be interesting to study if metacognition can be used to improve students' performance. One group of students which is consistently shown to lag behind on academic performance measures is the rising group of migrant students present in elementary school classes nowadays. The present study aimed to evaluate the relationship between intelligence, metacognition and performance for native as well as migrant elementary school students. To do so, outcomes were compared of a think-aloud measure as an indicator of metacognition, a crystallized and a fluid intelligence measure, and a performance test mainly including mathematical problem solving and reading comprehension tasks.

Firstly, the means and factorial structure of the different measures are compared for native and migrant students. Means show that, as expected, migrant students score significantly lower on the standardized performance measure as well as on intelligence measures. This can be caused by migrant students' inferior test-taking experiences or by poorer content-related and cultural knowledge needed to successfully answer test items. Surprisingly, although the fluid Raven measure is a nonverbal test, migrant students also perform lower on this test. But, migrant students in our sample do not have a lower level of metacognition than native students. Their think-aloud scores are comparable to the scores of native students. So the findings from secondary education where migrants reported to use more metacognitive regulation than natives (Blom et al., 2007) could not be repeated in our elementary school sample. Factorial analyses show that the different instruments are structured in approximately the same way for native and migrant students. When extracting a second factor, for both groups a metacognitive and a more cognitive factor could be determined. The fact that the division of the loadings over the two factors is comparable for native and migrants shows that the construct of metacognition is well comparable in the sample. However, this is not true for the Raven test of fluid intelligence. The Raven tests primarily loads on the academic factor for natives and on the metacognitive factor for migrant students. It seems that the Raven and the NIO tap the same latent variable in the native group, but not in the migrant group. Further research is needed to determine what causes this variation in outcomes on the Raven measure despite its supposed culture-free nature (Raven, 2000). In a follow-up study, more in depth analysis of the way students work with the Raven test would be needed to determine the source of these differences.

Concerning the relationship between intelligence and metacognition, it was predicted that these variables would be related to some degree, both for crystallized and fluid intelligence measures. This is indeed true for the native sample with correlations between metacognition and intelligence ranging from about .44 to .55. For both native and migrant students, the metacognitive subscale representing the degree to which students systematically made a calculation plan and executed the plan is most strongly related to intelligence. This shows that at least for some aspects of metacognition, students with higher intelligence will likely perform better. This is in line with findings from previous studies showing that around secondary school ages students with higher intelligence have a head start in regulatory activities (Alexander et al., 1995).

Secondly, as predicted, metacognition and performance were found to be moderately related with correlations ranging from .36 to .52. Overall, the measure of crystallized intelligence was found to explain most variance in academic performance of both native and migrant students. But, metacognition also has a significant effect on performance controlled for intelligence. The unique predictive value of metacognition for performance was found additional to both crystallized and fluid measures in line with the findings of Minnaert and Janssen (1999). There was no interaction between metacognition and students' cultural background as predictors of performance, showing that a higher level of metacognition can benefit both native and migrant students. The division of means of students of comparable intelligence varying in their level of metacognition again shows that a higher level of metacognition improves students' performance additive to the effect of intelligence. The unique effect of metacognition additive to the relation between metacognition and intelligence in our sample is in line with the mixed model. In this model metacognition and intelligence have some overlap, but also some unique predictive value for performance. The mixed model of the relation between intelligence, metacognition and performance was also found in most studies performed with adolescents (Veenman et al, 2006).

In comparison to other studies in secondary education, the relation between metacognition and performance in our sample is relatively moderate (c.f. Veenman & Spaans, 2005). As noted in the theoretical framework, this could have several reasons. Firstly, the intelligence measures were found to explain a major part of the variance in performance. Referring to the notion that for relatively easy tasks intelligence is a more determining factor for performance than metacognition (Prins et al., 2006), this might point out that the performance measure we used may not have had the level of complexity where a lot of metacognitive regulation is needed. This explanation seems to be supported by the spread of the performance score means reaching up to the maximum score in less than two standard deviations of the mean. Secondly, in our study, metacognition was measured using think-alouds for a mathematical problem solving task while the performance measure consists of both mathematical problem solving and reading comprehension tasks. Although word problem solving is related to comprehensive reading (Vilenius-Tuohimaa et al., 2008), the correspondence between metacognition measured in the think-alouds and metacognition used in the performance measure in this case is likely to be lower than with a more closely related performance measure. For future research we advise to measure metacognition with tasks that are closely aligned to the type of performance task for which metacognition is hypothesized to have predictive validity.

Regardless of these limitations, the study does permit us to draw some interesting conclusions about the way metacognition, intelligence and performance are related in elementary education. The results show that measures of crystallized intelligence have large predictive validity for academic performance. But, next to the unique impact of intelligence, metacognition was also found to affect performance. So, metacognition does add to the equation. Lastly, metacognition seems to be important for both native and migrant students. Students of both cultural backgrounds score comparably on the think-aloud measure and no interactions were found between cultural background and the effect of metacognition on performance. In this sense, stimulating metacognition may be used to support both native and migrant students.

The finding that metacognition influences performance in elementary education implies that metacognition can be used as a tool to stimulate young students' academic performance. The present study assessed the relation between spontaneous use of metacognition and performance. Possibly, metacognition can have more impact on performance when students are trained to apply metacognition to regulate and monitor their cognitive activities in academic domains. Since in our study metacognition used for mathematical problem solving was shown to influence performance, we suggest this to be an important domain in which metacognitive training has high potential. Some researchers have already shown the feasibility of executing metacognitive training in mathematics (i.e. Mevarech & Kramarski, 1997; Verschaffel et al., 1999a). However, as Schoenfeld (1992, p. 67) notes, "Developing self-regulatory skills in complex subject-matter domains is difficult. […] The task of creating the "right" instructional context, and providing the appropriate kinds of modeling and guidance, is challenging and subtle for the teacher". More research is needed which can give insight in the preconditions, the efficiency, and the effects of training for different age groups and on different outcome measures. Up to now, most intervention studies focused on teacher led training and cooperative learning as means for fostering metacognition. For future research we recommend to add to the knowledge base by studying computer supported instruction to aid teachers in fostering the metacognitive behavior of their students. As several reviews show (Kulik, 2003; Slavin & Lake, 2008), computers have the potential to be used as efficient and powerful instructional tools in contemporary education and we recommend to use this given to benefit metacognition research as well.

CHAPTER 3

Computer Supported Metacognitive Training for Solving Word Problems

ABSTRACT

Solving word problems plays an important role in elementary school mathematics education. However, many students have difficulties solving such tasks. In order to improve students' metacognition and problem solving skills, a computer program was developed consisting of word problems and metacognitive hints. The experimental group of grade 5 (*n* = 23) practiced with the computer program, in which students were free to choose metacognitive hints during problem solving. The control group (*n* = 26) did not work with the computer program. Results show that students using the metacognitive program outscore students in the control group on a problem solving posttest and improve their metacognition. Moreover, a relationship between mathematics performance and hint use was found. These results support the assumption that metacognition can be enhanced by students' free choice of metacognitive hints in a computerized learning environment, and that use of hints can increase students' performance in solving word problems.

This chapter is based on the published article:

Jacobse, A.E, & Harskamp, E.G. (2009). Student-controlled metacognitive training for solving word problems in primary school mathematics. *Educational Research and Evaluation, 15*(5), 447-463.

3.1 Introduction

In elementary school, word problems are offered to introduce new mathematical knowledge in more or less realistic situations. Word problems are especially used to let students apply their mathematical knowledge and make clear to them how mathematics can be used in different situations. But many students experience difficulties in solving word problems (Verschaffel et al., 1999a). Deficiencies may be due to the fact that students lack sufficient mastery of mathematical knowledge. However, mathematics textbooks and teachers pay much attention to remedy deficiencies in mathematical knowledge, and students still find it hard to solve word problems. Researchers therefore conclude that difficulties in solving word problems often originate from a lack of metacognition and not so much from a lack of mathematical knowledge (Hegarty, Mayer, & Monk, 1995; Schoenfeld, 1992; Verschaffel et al., 1999a). Metacognition concerns the procedural knowledge which enables learners to monitor and regulate their learning activities (Verschaffel et al., 1999a). This can help learners to analyze a task, to make a plan, or to evaluate and reflect upon their answers (Zimmerman, 2008). Using metacognition can strongly influence the mathematics performance of learners, even when the effect is controlled for the influence of intelligence (Helms-Lorenz & Jacobse, 2008; Van der Stel & Veenman, 2008; Veenman et al., 2005). This makes metacognitive support of great interest to educational research.

Although metacognition is very important in mathematics education in reference to its effect on performance and its role in self-regulated learning, metacognitive instruction seems underappreciated in practice. For instance, elementary school teachers in Dutch education spend a lot of time on instruction about cognitive mathematical procedures but give much less attention to discussing the metacognitive strategies which can be used to structure the execution of mathematics tasks (Inspectie van het onderwijs [Inspectorate of Education], 2008). Also, in mathematics text books and in pre-service teacher training little attention is given to metacognitive instruction. Apparently, teachers need tools which can help them incorporate metacognitive instruction into their daily teaching. These tools will have to integrate cognitive and metacognitive components in order to be effective. Especially for students of an elementary school age, metacognition seems to be quite domain specific and thus should be presented embedded in the cognitive content in which it can be applied (Veenman et al., 2006).

In the past, researchers have studied the effect of direct instruction about the use of metacognition to solve word problems (for an overview see Fuson, 2003). Students learned to solve problems step by step. In a first step, students used drawings or diagrams to build a schema of the problem situation. In the next step, they learned to recognize frequently occurring types of situations and ask themselves which mathematical operations are necessary to solve a problem. Then a solution plan followed, and in the last two steps students carried out the plan and checked their answer. Several studies, in different age groups, have demonstrated that direct instruction can be successful in teaching problem solving (e.g. Jitendra, 2002; Xin, Jitendra, & Deatline-Buchman, 2005). The results suggest that step-by-step instruction starting with representative diagrams helps students greatly in analyzing a problem situation and in choosing the right schema to solve a word problem. However, one drawback of step-by-step teaching is that it is effective only in cases aimed at solving similar problems in which the procedure for solving these problems is clear from the start (Jonassen, 2003; Moreno, 2006). With direct training, students will learn the steps to solve problems but are unable to adapt their procedures to new and diverse problem situations. This highlights the need for other types of learning environments in which students learn to take control over their own learning process and choose when to consult metacognitive support to help solve a problem. Instruction should be designed in a way that leaves room for the students to choose their approach to solve a problem. Students need to acquire a flexible way of problem solving that allows them to tackle different types of problems. In searching for a tool which can help teachers and students in this way, computer environments easily arise as a good possibility. Computer environments are used extensively both in and out of school and offer tools which teachers can use with relatively little effort. In the next paragraph, the possibilities of such environments for the support of metacognition are discussed, as well as some practical implications.

3.2 Computer Programs with Metacognitive Support

In the past years, computer environments have been used broadly to support learning in schools. However, some learners experience difficulties in using metacognition to regulate their learning in such environments, particularly in conceptually rich domains as mathematics (Azevedo, 2005). A new and promising research subject thus may be assessing the effects of computer environments which combine cognitive content with metacognitive support. Such programs can be designed in various ways, for instance by using intelligent tutoring systems, educational hypermedia systems, virtual or human agents, metacognitive hints, metacognitive question cards, and so on. Programs also differ in the phase in which the support is provided and the extent to which the learner has control over the metacognitive support. As discussed in the introduction, a good way of enhancing flexible metacognitive regulation is allowing a certain level of student control of the problem solving behavior. Schoenfeld (1992) is an important proponent of the approach to problem solving where students take the initiative. He investigated expert and novice problem solving behavior in mathematics education. On the basis of his research, he distinguished between five ‘episodes’ in the process of problem solving:

- survey the problem (read, analyze);
- activate prior knowledge (explore);
- make a plan (plan);
- carry out the plan (implement);
- check the answer (verify).

Schoenfeld found that experts and novices differ in their approach to solving problems. Novices almost immediately start to work out a poorly defined plan, whereas experts take time to analyze the problem and gather information before making and implementing a plan. Experts follow the episodes more systematically and take much time to analyze the problem. And they track back between episodes when, for instance, they want to check if a plan fits the problem situation. Based on these findings, Schoenfeld argued that novices need to learn to work through the different episodes more effectively and, by doing this, build up their metacognition. He proposed to teach students the use of the episodes through metacognitive questions and hints related to the episodes.

In the literature, there are already some examples of the effectiveness of using student controlled questions and hints in computer environments. For instance in the studies of Mathan and Koedinger (2005) and Harskamp and Suhre (2007). Mathan and Koedinger created a model of the 'intelligent novice'. Using this model, at first students worked out spreadsheet problems without help. If they wanted to move on before having solved a problem correctly, they would be advised to accept help in finishing the problem. Students using this model learned faster and performed better on a conceptual understanding test and on a transfer test than did a control group. Harskamp and Suhre tested the effectiveness of a training program based on Schoenfeld's problem solving episodes with hints to help secondary school students solve mathematical word problems. Students were free to use the hints in the manner they wanted. The students who used the program with the hints outscored students in the control group, who used the program without hints. There was a relationship between hint- use and the posttest problem solving achievement of students in the experimental group. Another study by Pol, Harskamp, Suhre, and Goedhart (2008) provided students with hints over the episodes analyze, explore, plan, implement, and verify in a web-based computer program in physics. The hints addressed metacognitive processes in problem solving as well as the cognitive content of the task. The study showed that students receiving problem solving hints increased their systematic use of the hints, whereas systematic hint use was linked to higher problem solving behavior. Thus, not only a practice effect of performing the tasks but also an effect of the problem solving hints provided in the program was found. A recent study of Azevedo, Greene and Moos (2007) made use of a human tutor giving external adaptive self-regulative support in a hypermedia environment in the domain of science. The study demonstrated that students gained more declarative knowledge and reached a higher mental model during the posttest when aided by external metacognitive hints by a human tutor. Aleven, McLaren, Roll, and Koedinger (2006) developed an intelligent tutoring system that guides students' metacognitive help seeking alongside cognitive hints. Research on this system shows much promise in understanding how help-seeking skills of students can be enhanced. Several studies coincide with the assumption found in aforementioned studies that metacognitive support in computer environments can have a positive effect on the performance of students of various age groups

(Bannert, 2006; Bannert, Hildebrand, & Mengelkamp, 2009; Clark & Mayer, 2008; Teong, 2003; Wood & Wood, 1999). However, not all metacognitive tools increase learning outcomes (Graesser et al., 2007). Besides providing a level of freedom for the students, another factor influencing success seems to be that metacognitive tools should not be too complex in order to avoid cognitive overload (Roll, Aleven, McLaren, & Koedinger, 2007; Schraw, 2007). Also, it should be clear to students why the metacognition is beneficial in order to motivate them for a change in their working method. Roll et al. (2007), for instance, added an instruction with video and a classroom discussion to enhance awareness among students about why to monitor help seeking.

Azevedo (2007) summarized requirements for effective metacognitive computer support:

(a) It requires students to make instructional decisions. Students should be able to make their own decisions regarding their working method. For instance, they should have some freedom in the use of the metacognitive support, as shown in the studies of Harskamp and Suhre (2007), Pol et al. (2008), and Wood and Wood (1999).

(b) It is embedded in a particular learning context which allows decisions on contextual support or resources. For instance, a program about certain topics of science or mathematics in which contextual problems are posed with pedagogical agents or hints for the student to use as help to solve the problems (Clark & Mayer, 2008).

(c) It supports learners' self-regulatory processes. The tool should thus enhance cognitive, metacognitive, and/or motivational behavior by using hints, questions, models, or such. For instance, Teong (2003) used a computer program with Schoenfelds' episodes as a metacognitive framework for hints to help students solve mathematics word problems.

(d) It enhances domain- or activity-specific learning skills. The metacognitive support should thus be embedded in the cognitive content (Veenman et al., 2006). However, the metacognitive goals should not be ‘‘lost’’ within the cognitive content; students should pay attention to the metacognitive as well as the cognitive content (Roll et al., 2007).

(e) It comprises of external regulating agents (human or artificial) who support the learning process (c.f. Aleven et al., 2006; Azevedo et al., 2007).

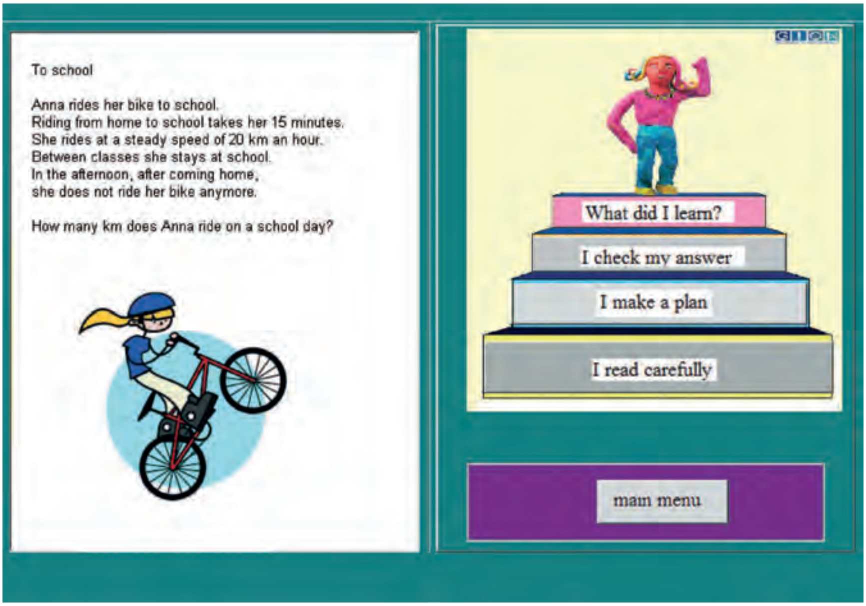
(f) It stimulates metacognitive processes prior to, during, and after performing a task. Thus, metacognitive processes during the whole problem solving sequence are important. Metacognitive support should be directly provided during problem solving, and there should be immediate feedback on errors (Roll et al., 2007).

Instructional support in computer programs which meets these requirements is suggested to help students effectively, if used well.

3.3 Development of a Metacognitive Computer Program

In order to assess the effect of metacognitive hints in a computer environment in this study, a computer program known from previous research (Harskamp & Suhre, 2007) is modified for use in the elementary school mathematics curriculum by replacing the original content with word problems suit for elementary students. Additionally – adhering to the implications mentioned above – metacognitive hints are added to the cognitive content of the word problems. The metacognitive hints are based on the so-called ''Task stairs procedure'' (see Figure 3.1) developed by Jacobse (2007). In this procedure, a picture of a staircase is offered with the steps 'I read carefully' (read/ analyze), 'I make a plan' (explore and plan), 'I check my answer' (implement), and 'What did I learn?' (verify). These steps have been used previously to teach metacognition in classroom instruction in elementary education.

The hints that are offered with each step of the Task stairs correspond to the metacognitive episodes proposed by Schoenfeld (1992) and Veenman, Kerseboom, and Imthorn (2000). The episodes are visualized in a systematic way suggesting a step-by-step procedure (see Figure 3.1). This is to stimulate students to use the hints systematically. However, students are free to choose the hints in any order they like. Thus, the staircase itself represents the metacognitive framework, and with each step content-embedded hints are offered to support the students during the episodes of problem solving. For instance, in the step 'I read carefully' (Figure 3.2), three orientation hints are offered pertaining to: The question ('What is the question?'), selection of relevant information ('Which numbers?'), and a graphical representation of the task ('Drawing the problem'). The graphical representation of the task is added to decrease cognitive load and help learners create a correct mental model of the task, as suggested by Jitendra (2002), Xin et al. (2005), and Harskamp and Suhre (2007).

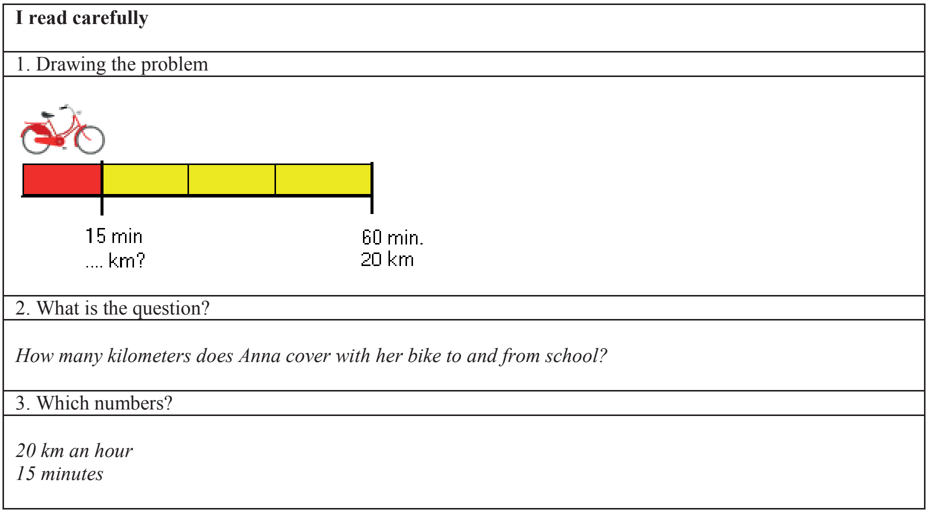


To school

Anna rides her bike to school.
Riding from home to school takes her 15 minutes.
She rides at a steady speed of 20 km an hour.
Between classes she stays at school.
In the afternoon, after coming home,
she does not ride her bike anymore.

How many km does Anna ride on a school day?

Figure 3.1 Interface of a mathematical word problem with Task stair hints.



I read carefully

1. Drawing the problem

15 min
.... km?

60 min.
20 km

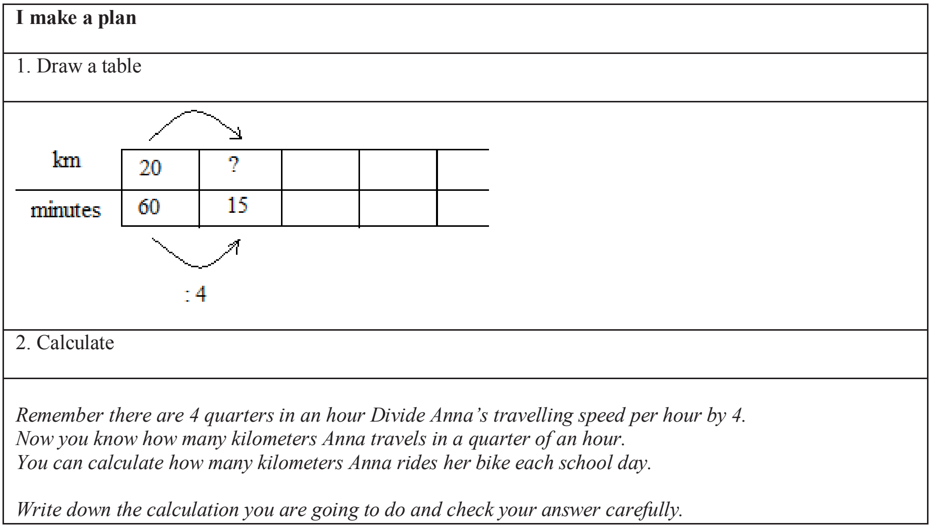
2. What is the question?

How many kilometers does Anna cover with her bike to and from school?

3. Which numbers?

20 km an hour
15 minutes

Figure 3.2 Example of hints in the Task stairs step ‘I read carefully’



I make a plan

1. Draw a table

km	20	?			
minutes	60	15			

: 4

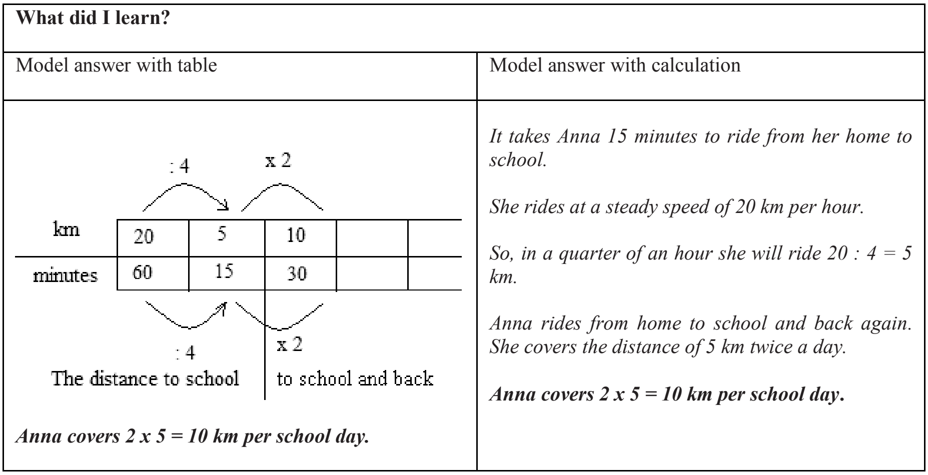
2. Calculate

Remember there are 4 quarters in an hour Divide Anna’s travelling speed per hour by 4.
Now you know how many kilometers Anna travels in a quarter of an hour.
You can calculate how many kilometers Anna rides her bike each school day.

Write down the calculation you are going to do and check your answer carefully.

Figure 3.3 Example of hints in the Task stairs step ‘I make a plan’

The metacognitive hint ‘I make a plan’ is divided into two hints with different solution methods for the task, so students can make their own decision about which plan to use (see Figure 3.3). After that, in the step ‘I check my answer’, students are prompted to check their calculations. When they think they are ready, they can fill in an answer in this step and get feedback if their answer is correct or not. If a student has filled in a wrong answer 3 times, the student is asked which metacognitive step would have needed more attention. Hereafter, students can choose to review the metacognitive hint ‘What did I learn?’ in which model answers for the two different solution paths are offered (see Figure 3.4).



What did I learn?	
Model answer with table	Model answer with calculation
: 4 x 2 km 20 5 10 minutes 60 15 30 : 4 x 2 The distance to school | to school and back ***Anna covers 2 x 5 = 10 km per school day.***	*It takes Anna 15 minutes to ride from her home to school.* *She rides at a steady speed of 20 km per hour.* *So, in a quarter of an hour she will ride 20 : 4 = 5 km.* *Anna rides from home to school and back again. She covers the distance of 5 km twice a day.* ***Anna covers 2 x 5 = 10 km per school day.***

Figure 3.4 Example of hints in the Task stairs step ‘What did I learn?’

The program consists of 40 novel word problems that need addition, subtraction, multiplication, or division to solve. There are 30 problems for all students and 10 extra problems for fast problem solvers. The word problems are developed for students in grade 5 of elementary school. In the problems, a combination of two mathematical operations is required to solve them (cf. Figure 3.1). Half of the problems also have an extra digit in the text not relevant for their solution. This type of problem is especially suitable to let students analyze the problem carefully, whereas the operations are not too hard to be performed by fifth-grade students. The problems are not included in mathematics textbooks.

Evaluating the use of Azevedo’s (2007) recommendations for effective metacognitive support leads to the following observations: (a) with the Task stairs program, students are able to make their own decisions about which hints to use; (b) the support is embedded in ‘Task stairs’, a learning context which allows decisions on contextual support (hints); (c) and (d) the hints consist of a combination of cognitive hints in a metacognitive framework (the staircase), and each hint is embedded in the context of this staircase. No use was made of human or virtual agents (e), but during the study a research assistant was present to support the students when needed; (f) the hints in the Task stairs program become visible when clicking on a word problem. In the program interface, the problem and the steps of the staircase are presented at the same time, and after reading the problem a student can decide to choose one of the steps for support. The support is thus provided during and after processing a word problem.

There is no instructional support before doing a task except for a preflight training with two sample problems that the researcher did with the students in the experimental group in order to show them how the program was best used. The assumption is that instructional support before solving a novel problem is not as effective as help during and after problem solving. This applies especially to nonstandard problems that do not resemble the usual text book problems (Moreno, 2006). When a wrong answer is given, immediate feedback is provided, and students proceed to the reflection question about their working method, directly after giving a right or three wrong answers. Also, during the preflight training a classroom discussion is guided by the research assistant about the usefulness of the metacognition in order to motivate the students.

3.4 Research Questions

This study assesses the effect of the computer program with metacognitive hints on metacognition and performance. The research questions are:

(1) Does practice by means of the computer program with metacognitive hints enhance the problem solving performance of elementary school students?

(2) Can the effect of the computer program be attributed to students' use of metacognitive hints?

(3) Is students' metacognition affected by use of the metacognitive hints?

Since the computer program with metacognitive hints adheres to most of the principles known to be effective from previous research, it is expected to have an effect on students' problem solving performance. This will in part be an effect of practicing the mathematical content, but an additive effect of the metacognitive hints is also expected (c.f. Pol et al., 2008). Furthermore, the metacognitive support offered in the hints is predicted to improve the metacognitive behavior of the students. By using the hints in the computer program, students are likely to adapt some of the metacognitive activities which are offered. The fact that students internalize some of these metacognitive activities is intended to lead to an increase in their own regulation, which in turn could lead to a decrease of hint-use in the program.

3.5 Method

3.5.1 Sample

The experimental study was executed in two fifth-grade classes of an elementary school in a city in the north of The Netherlands. Forty-nine students (mean age ≈ 11) participated, of which 23 were in the experimental condition and 26 were in the control condition. The sample consisted of 22 boys and 27 girls. The control group had 11 boys and 15 girls, and the experimental group had 11 boys and 12 girls. Students in both conditions were of comparable mean socioeconomic status, did not differ in mean mathematics performance scores on a norm-referenced test ($t(46) = .99$, $p = .33$), and did not differ on the word problem solving pretest ($t(44) = -.13$, $p = .90$).

3.5.2 Design

One of the two classes functioned as a control group, while students in the other class formed the treatment group. The two classes used the same mathematics textbook and covered content of the textbook at about the same pace. The experimental group received a preflight training of 30 minutes on how to use the Task stairs computer program. In the preflight training two sample problems were discussed. Then, during a fortnight, this group worked with the Task stairs computer program additional to the regular mathematics curriculum for 2 times a week of approximately 30 minutes. The control group worked with the mathematics curriculum in the usual fashion (without the Task stairs). For both groups, a word problem solving pre- and posttest were collected, and for the experimental group computer log files of the whole group and think-aloud protocols of 10 randomly chosen students were analyzed as well. Furthermore, video recordings of 14 students working with the computer program were observed.

3.5.3 Variables

Mathematics performance. As a mathematical performance measure, scores on two word-problem tests for mathematics were used (pretest/posttest). These word-problem tests consist of 15 items and were developed for this study. The items in the posttest were adaptations of items in the pretest. Of both tests, the items are considered comparable in style to the word problems in the computer program, but less complex. The pretest has a reliability of $\alpha = .78$.

Two sample problems are for instance:

Marianne went running. *She ran for 2 kilometers and then rested for 5 minutes.* *After this she ran three times as far.* *How many kilometers did Marianne run in total?*	*Jeroen has sold cookies.* *He sold them for €0,50 each.* *He got €5,- extra from his mother.* *He earned €35 in total.* *How many cookies did he sell?*

Log files. While working with the computer program, log files of each student were saved. These log files contain information about the number of word problems students completed and about which hints have been used. In the analyses, information from the log files of the first 20 word problems was used since this is the number of problems completed by most students in the experimental group (20 out of 22 students).

Metacognition. To analyze students' metacognition, two measures were used. Primarily, think-aloud protocols of 10 randomly chosen students from the experimental group were analyzed in the pre- and posttest. In order to do so, video recordings were made of students individually performing a word task while thinking aloud. Afterwards, the think-aloud protocols were evaluated by two judges using a scoring scheme which is an adaptation of the scheme introduced by Veenman et al. (2000). The number of items of the scheme has been reduced for this study because of the relatively small quantity of metacognitive regulation used in this age group in reference to the secondary school students in the study of Veenman et al. A copy of the adapted score scheme for assessing metacognition in elementary education is added in Table 3.1.

Table 3.1
Scoring scheme for systematical observation of think-aloud protocols in word problem solving

Step		Activity
I read carefully	1	Reading carefully
	2	Selection of information/ numbers
I make a plan	3	Making a calculation plan
	4	Taking systematic notes
I check my answer	5	Process monitoring
	6	Checking calculations and answers
	7	Drawing a conclusion
What did I learn?	8	Reflecting on the answer (i.e. in reference to a prior estimation)
	9	Reflection on what has been learned
Correctness	10	Correct answer?

Only items 1, 2, 3, and 4 of the list about students' metacognitive regulation in analyzing and planning were used for this study. The activities in the scheme were rated on a 3-point scale with scores ranging from 0 (activity not executed), through 1 (activity partly executed) to 2 (activity fully executed). Activity 10 (correct answer to the task) is not counted as a metacognitive activity. In this study, the total metacognition score thus ranges from 0 to 8 points. The scores of each individual student were given by two judges in the pretest as well as the posttest. After independently scoring each student, the two judges argued until agreement was reached and the scores were set, which is a method used frequently in assessment of think-aloud protocols (c.f. Veenman, 2005; 2000).

Furthermore, fourteen randomly chosen students were videotaped while working with the computer program. These videotapes of their working method with the computer program were observed by means of a checklist particularly focusing on their hint use. For instance, observations were made referring to the points:

- Does the student make use of the hints when he/she cannot complete the task on his/her own?
- Are the hints used in a systematic order?
- Does looking at the hints lead to improvement of the problem solving behavior of the student?

3.6 Results

3.6.1 Performance and use of metacognitive hints

Table 3.2 shows the differences in mean mathematics score for the control group and the experimental group. The groups did not differ in the pretest, but in the posttest a significant difference was found ($t(45) = -1.73$, $p = .05$), which may be attributed to the intervention, leading to a medium effect size of $d = .51$.

Table 3.2

Mean mathematics scores (test of 15 items) of the control group and the experimental group on the posttest measure

	N	*M*	*SD*	*SE*
Control group	24	7.96	4.03	.82
Experimental group	23	9.74	2.93	.61

To assess what the effect of the intervention is independent of the mathematics proficiency of the students, an ANCOVA was performed controlling for the pretest in mathematics (see Table 3.3). The results confirm the finding that, apart from the large main effect of the pretest, the intervention had a significant main effect on mathematics performance ($F(1,41) = 7.23$, $p = .01$). Corrected for pretest differences, the metacognitive training program had a large effect on problem solving performance of $d = .81$.

Table 3.3

ANCOVA with dependent variable posttest mathematics, fixed factor computer program (intervention) and control variable pretest mathematics.

	Type III Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Corrected Model	332.06	2	166.03	32.18	.00**
Intercept	29.66	1	29.66	5.75	.01**
Computer program	37.29	1	37.29	7.23	.01**
Pretest mathematics	299.25	1	299.25	57.99	.00**
Error	211.57	41	5.16		
Total	4180.00	44			
Corrected Total	543.64	43			

Although this result seems promising, it does not yet make clear if this is merely a practice effect of working with the word problems in the computer program or if results can partly be attributed to the metacognitive hints. In order to assess this, the level of hint use and the relationships between the use of hints and mathematics performance were evaluated.

At first, descriptives of the log files were analyzed to see if students actually used the hints in the program (see Figure 3.5). On average, students used three hints per task. The ‘drawing’ hints with a visualization of the problem in the first step of the staircase ‘I read carefully’ was used most, this hint was used for 67% of the problems. The other hints within the first step of the staircase ‘What is the question?’ and ‘Which numbers?’ were used with 46% and 40% of the problems, respectively. The hints in the steps ‘I make a plan’ and ‘What did I learn?’ (reflection with a model answer) were also popular; they were used with 60% (plan) respectively 58% (model answer) of the problems. Consequently, the conclusion can be made that the metacognitive hints in the computer program were used to a relatively large extent.

When we look at the stability of hint use over the 20 tasks, we can see that the intensity of hint use generally decreases over time. The differences in hint use between the first 10 word problems and word problems 11 – 20 are portrayed in Figure 3.6.

The average number of hints used in the first 10 tasks and in the last 10 tasks was taken as a starting point. The average in total hint use is almost 30 for the first 10 tasks and about 25 for the last 10 tasks. The difference between these averages is significant ($t(19) = 1.75$, $p = .05$); thus, in solving the first 10 tasks, students used more hints than in the last 10 tasks. Additionally, the differences in the average number of times students used the different hints in Task stairs were analyzed. Students used the hint ‘Drawing the problem’ about 7 times in the first 10 problems and about 6 times in the second half of the program ($t(19) = 1.82$, $p = .04$). Also the use of the hint ‘Which numbers?’ decreased significantly from

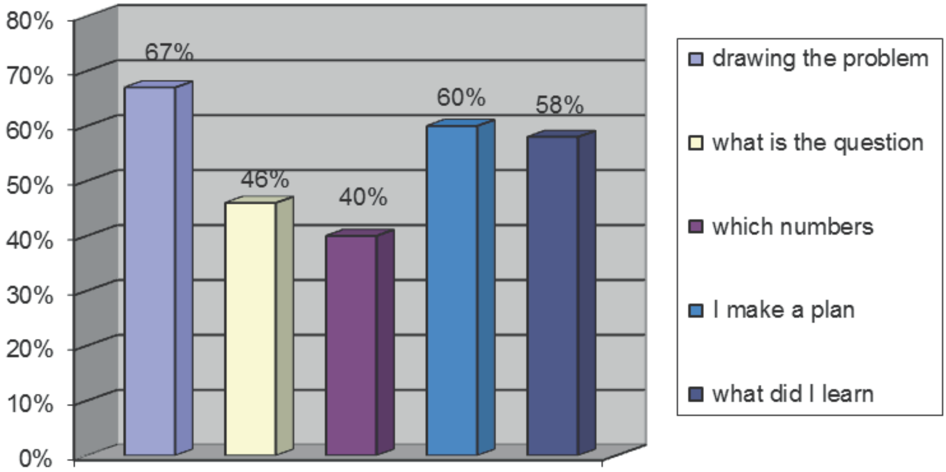


Figure 3.5 Percentage of metacognitive hints used over the different steps of the Task stairs in 20 word problems

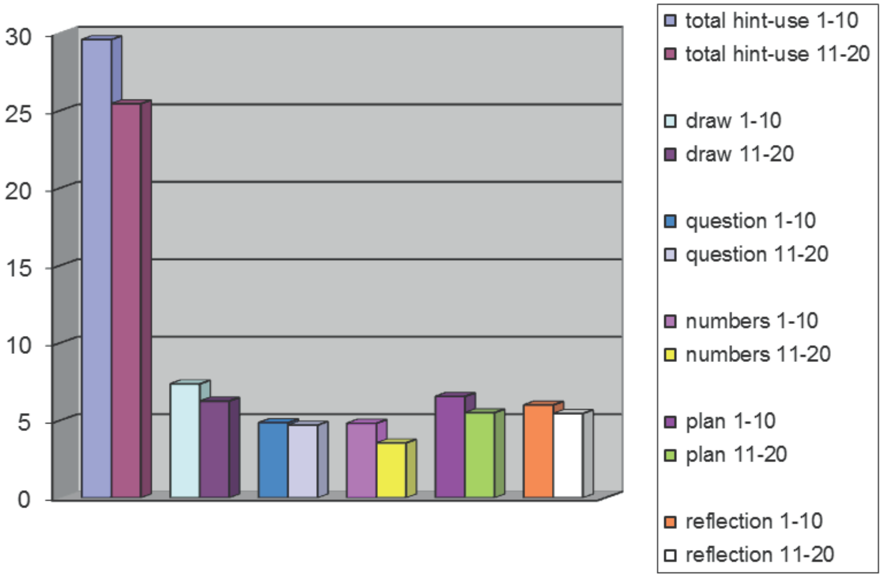


Figure 3.6 Average number of times hints were used in the first ten and the second ten problems in the program

over 4 to about 3 times (*t*(19) = 1.81, *p* = .04). Both hints belong to the step 'I read carefully', which supports analysis of the problem. There was no significant decline in use of the hints in the steps 'I make a plan' and 'What did I learn?'.

Correlational analysis makes clear that the use of these different hints was strongly interrelated. The hints within the step 'I read carefully' ('Drawing', 'Which numbers?', and 'What is the question?') and the steps 'I make a plan' and 'What did I learn?' were all significantly related. Correlations range from .42 to .74. This indicates that students were generally consistent in their use of hints. So, although hint-use slightly decreased over time, the hints have been used frequently and consistently throughout the program.

In order to assess if there is a relationship between the use of metacognitive hints in the 20 word problems and the word problem posttest, partial correlations were calculated controlling for scores on the word problem pretest. The partial correlations reveal that mathematics problem solving performance in the experimental group is significantly related to the use of hints (*r*(16) = .47, *p* = .03). Having a closer look at the different hints reveals that this was mainly due to the hints in the step 'I read carefully' (*r*(16) =.47, *p* = .02).

3.6.2 Metacognition

Think-aloud protocols of 10 randomly chosen students from the experimental group trying to solve a word problem were analyzed in the pre- and posttest. In order to do this, individual video recordings of students were made. The metacognitive activities in the scheme were rated on a 3-point scale with scores ranging from 0 to 8. The mean scores and standard deviations on pre- and posttest were 3.33 (*SD* = .87) and 4.56 (*SD* = 1.13), respectively. A paired samples *t*-test of think-aloud protocols reveals that the students progressed significantly from pre- to posttest in analyzing and planning activities (*t*(8) = -2.23, *p* = .03).

Besides the think-aloud protocols, students were also videotaped while working with the computer program. Qualitative analyses of the videotapes of 14 students working with the computer program show that use of the hints actually led to problem solving behavior. For instance, after clicking on the orientation hint 'Drawing the problem', students often took notes. The hint 'I make a plan' was also used frequently. Most students have a look at both hints within 'I make a plan' that show a way to solve the problem (e.g., a hint with a plan to do a calculation and a hint with a plan to use a table). Students generally expressed a preference for the calculation plan with a table or pictorial representation.

3.7 Conclusion and Discussion

The present study aimed at assessing the usefulness of working with a computer program with metacognitive hints for enhancing metacognition and problem solving in mathematics in elementary school grade 5. Previous research has shown that such interventions can be effective in different domains, age groups, and with different types of metacognitive tools (i.e. Azevedo et al., 2007; Bannert, 2006; Pol et al., 2008). However, more research is needed to explore the notion that metacognitive hints can be effective in student-controlled learning environments. The computer program developed for this study offers mathematical word problems supported by hints. The structure of the hints is metacognitive, whereas the information the hints carry is primarily content related aimed at helping students to solve the problem. The hints are presented following a systematic sequence (the Task Stairs procedure, see Jacobse, 2007). In a preflight training, students are informed about why to use the hints in the sequence of the staircase, and they are given some practice with the program. During the experiment, the students are free to choose which hints to use when solving the problems in the program.

Analysis of the log files reveals that students use the hints to a relatively large extent (40% – 67% of the specific hints were used). Especially the visual representation of the problem and the planning hint were used a lot. However, hint use decreases slightly over time. This was expected, because it is likely that students internalize skills of analyzing and planning over time, so they may need less external support for these metacognitive processes. After practicing with the computer program with metacognitive hints during ten lessons, students of the experimental group significantly outperform students of the control group on the problem solving posttest, controlling for pretest problem solving scores.

Although using the computer program has a positive effect on students' mathematics performance, this could merely be an effect of practicing the word problems in the program. In order to estimate if the metacognitive hints have an additive effect on the performance of the students, the relationship between hint-use and performance was assessed. The analyses reveal that there is a significant relationship between the two, especially for the hints that support the episode of reading and analyzing word problems with activities such as reading ('What is the question?'), selecting information ('Which numbers?'), and making a graphic representation ('Drawing the problem'). But this finding may be due to the nature of the word tasks in the program. The tasks contain relatively much text and often included an extra digit irrelevant for the solution of the problem. Students had to analyze the tasks carefully and could not jump easily to a solution plan, as novices use to do (Schoenfeld, 1992). Our first recommendation is to do further research with a more varied set of tasks to assess which hints might benefit students most.

Although a relationship was found between hint-use and problem solving, a causal relation could not yet be established. In order to get more information about the effect of the use of metacognitive hints, qualitative analyses and think-aloud protocols were collected. Both the observations of students working with the computer program and the think-aloud protocols of a test problem indicate that students in the experimental group enhance their problem solving behavior due to the use of the metacognitive hints. The second recommendation is to track the movement of students through the program with hints. It is of interest to study if students adapt their behavior in working with the computer according to the hints they use and if this affects their growth in metacognitive skillfulness over time.

Lastly, the results in this study support the idea that, apart from a practice effect, there is an additive effect of the use of metacognitive hints on mathematics performance. In the experimental group, there was a relationship between use of hints and problem solving achievement. However, the extent to which both factors (practice by doing tasks and use of hints improving metacognition) impact students' problem solving remains unclear due to the design of the study. The study did not control for a practice effect (Clark & Mayer, 2008). A final recommendation is to execute further research with an experimental group practicing with the computer program with metacognitive hints versus a control group receiving practice with a the same word problems in a computer program, but without hints. Such a research design will provide the opportunity to disentangle the effect of practice from the effect of metacognitive support on students' metacognition and problem solving performance.

CHAPTER 4

Using Metacognitive Training to Improve Problem Solving in Mathematics: A Guided Approach

ABSTRACT

Students in elementary school often experience difficulties in solving word problems. Several authors have suggested that training metacognitive monitoring and regulation of problem solving with cues, hints or supportive questions in computer based learning environments can help students to improve their performance. For problem solving, guided instruction is preferred over training in which students are taught fixed procedures to solve word problems. However, little is known about effects of guided training on problem solving performance of elementary school students. In a previous study, a computer environment with metacognitive hints building on the problem solving model of Schoenfeld was developed for grade 5. Using this program, a six week experimental study was executed (*n* = 73) to examine the effects of guided metacognitive training in comparison to a practice-only condition. Students working with the metacognitive hints were found to solve more problems in the computer program than the control group. Posttest results show that the students in the experimental condition learned to judge their performance more accurately than students in the practice-only condition. Moreover, metacognitive training had an effect on students' problem solving performance beyond the effect of mere practice. Limitations and possibilities for further research are discussed.

This chapter is based on the submitted article:

Jacobse, A.E., & Harskamp, E.G. (submitted). Using metacognitive training to improve problem solving in mathematics: A guided approach.

4.1 Metacognition and Mathematical Problem Solving

In upper elementary school, one of the main topics to learn in mathematics is word problem solving. It plays an important role in standards for mathematics education and testing (Schoenfeld, 2002). Word problems are based on a textual description of a realistic context that requires students to apply their mathematical knowledge. An example from grade 5:
Julie prepares snacks for a family reunion. There will be 54 people present.
She makes snacks with eggs. She prepares two eggs for each person.
There are twelve eggs in a box. How many boxes of eggs does she need?

This is a two-step problem. Students first have to multiply 54 x 2 and then divide by 12. They need a proficient level of mathematical knowledge to perform these calculations. Mathematical knowledge concerns the knowledge of mathematical procedures to and the understanding of concepts. However, although mathematical knowledge is needed, a high level of mathematical knowledge does not necessarily lead to successful word problem solving (Fuchs et al., 2008; Mayer & Hegarty, 1996; Swanson & Beebe-Frankenberger, 2004). The main difficulty with word problems is that students have to construct a problem model of a context by making inferences from the text (Fuchs et al., 2008; Kintsch & Greeno, 1985). Students have to understand which variables are involved (*people and eggs*) and which relationships to deal with (*'person and 2 eggs' and 'total of boxes and 12 eggs per box'*). International evaluations show that trying to solve word problems causes difficulties for a large proportion of elementary school students (Mullis et al, 2008). In word problem solving, most students were found to quickly select numbers from the word problem and make a calculation, irrespective of whether the calculation fits the problem. Also, students sometimes unjustly generalize solution procedures to different types of problems (Verschaffel, et al., 1999b). Although students generally understand the calculations embedded in the word problems, it seems that they mainly encounter difficulties because they rarely take the time to regulate and monitor the use of cognitive strategies. This causes them to skip or misinterpret information from the problem and choose inappropriate solutions (Verschaffel et al., 1999a).

To overcome such difficulties, students need metacognition to help them structure their problem solving process (Mayer, 1998; Verschaffel et al., 1999a). Metacognition refers to thought processes on a meta-level which can be used to control cognitive processes (Flavell, 1979). Metacognition can be applied to monitor the learning process and to regulate cognitive activities (Efklides, 2006; Nelson, 1996). Metacognitive monitoring refers to students' ongoing control over the learning process (Desoete, 2008). If a student thinks he or she cannot solve a problem readily, then it is time to seek further resources. Monitoring then influences the use of metacognitive regulation (Nelson, 1996). Metacognitive regulation refers to mental activities used to consciously apply cognitive strategies (Efklides, 2006). Both aspects of metacognition are very important to help avoid an unstructured approach to word problem solving.

To successfully solve word problems, both metacognitive and cognitive processes are involved. One cannot do without the other. For example, when starting with a complex word problem, a student may judge that the problem is difficult to solve, based on his/her prior knowledge and experiences. This is an example of monitoring behavior in which the student also thinks ahead about the solution plan. Consequently, the student may choose to take some more time to carefully analyze the problem and select relevant information from the text. The choice to do so is an example of metacognitive regulation, while reading the text is the cognitive result of this choice. Then, while executing a solution plan, the student might notice that something is going wrong and go back to selecting information from the text. The fact that the student notices a mistake is another example of monitoring behavior, which in turn again influences the use of metacognitive regulation to solve the problem. This example shows that metacognitive processes are conceptually different from cognitive processes, but they work together in constant interplay. Thus, to help students to become more able problem solvers, training has to entail both metacognitive directions of what procedure to follow, but also cognitive suggestions of what to do to solve a problem.

By observing expert problem solvers in college, Schoenfeld (1992) distinguished that metacognitive monitoring and regulation in mathematical problem solving manifest themselves in five distinct episodes: Read/ analyze, explore, plan, implement and verify. In our example of the eggs served at the family reunion, careful reading and analyzing the problem will lead the student to understand the problem situation. And conscious exploration will reveal the relationships and type of calculations in the problem (*‘2 eggs per person’, ‘12 eggs in a box’, this is a multiplication and division problem*). Then the numbers and relations in the task then have to be translated into an appropriate mathematical model for a solution plan (*in this case the calculation (54 x 2) / 12*) and this plan needs to be executed thoroughly, for instance by writing the calculation down. Lastly, when finding an answer, experts do not quit directly. They were found to first verify their answer by checking their calculation, referring back to the question and possibly by thinking back to their initial estimations. This makes students less vulnerable for formulating a wrong answer. Schoenfeld suggested that asking students metacognitive questions that go with the episodes (What is the problem about? What is the question? Which mathematical procedures are involved? etc.) can help them to bring relevant strategies to mind. He successfully trained college students to solve problems by teacher led instruction. However, he did not study how to train elementary students.

The main aim of this study is to evaluate whether metacognitive support can help elementary school students to use problem solving episodes better and to subsequently improve their problem solving achievement. This issue is studied using an experimental design with an metacognitive training program in grade five.

4.2 Metacognitive Training in Word Problem Solving

4.2.1 Different approachs to train word problem solving

To train students in word problem solving, there are two different types of approaches. The first type is training in which students are taught specific step-by-step solution methods. In most of such metacognitive training programs teachers teach students how to recognize types of problems and subsequently to use a predetermined approach. Instruction is typically given by the teacher modeling solution steps or by using text or schema-based worked examples. It has already been shown that such an approach to metacognitive training can help students to improve their performance in the types of problems that are trained. This may be especially beneficial for low achieving students (Fuchs et al., 2009; Xin et al., 2005). However, this type of training also has some drawbacks. Because the solution processes are directed by the teacher, students do not learn to make their own instructional decisions about how to solve the problems. And when fixed procedures are used for specific problem types, effects rarely transfer to other types of problems (Jonassen, 2003; Moreno, 2006).

Another possible approach to train problem solving is guided training. Guided training is based on the idea that the learning environment provides instructional support, but that the student directs the learning process by choosing when and how to use the support (Mayer, 2004). By letting students take control, they learn to reflect on their approach instead of memorizing fixed solutions. For this type of training, mostly computer based learning environments are used which offer hints or questions which remind the students of metacognitive and cognitive processes (Azevedo, 2007). Sometimes the hints or questions contain suggestions about how to solve the problem, but the solution steps are not fixed; students should “connect the dots” themselves. Guided training has the major benefit over training with fully worked out step-by-step procedures that students learn to make their own decisions about how to solve problems and become flexible in solving various types of problems (Mayer, 2004). This connects well to the idea of problem solving as a non-standard, flexible process (Schoenfeld, 1985). Previous research has shown that training with a high level of student control requires students to have sufficient metacognitive abilities to navigate through the instructional program (Winters, Greene & Costich, 2008). Thus, when offering guided training to students, it is important to embed enough metacognitive and cognitive support about the problem solving process to help students to make the right instructional decisions.

As discussed above, direct and guided training both have their pros and cons. However, to teach students how to flexibly solve different types of mathematical problems, guided metacognitive training seems particularly suitable. Therefore, in this study we have chosen a training program which gives students suggestions about how to solve problems but also gives them some freedom to make their own instructional choices. The effectiveness of such an approach has been studied in secondary education, but little is known about the effectiveness of guided metacognitive training in elementary school. Some examples from the literature which were used in the development of the training materials are discussed below.

4.2.2. Computer supported metacognitive training with hints and prompts

In secondary education, there are already some examples of effective guided metacognitive training with computer based learning environments. The studies of Harskamp & Suhre (2007), Aleven & Koedinger (2002), Kramaski & Gutman (2006) and Kramarski and Mizrachi (2006) have shown that computer supported training with student-directed metacognitive prompts or questions can considerably benefit students' problem solving performance. The metacognitive features in the computer environments prompted students to structure their learning over various problem solving episodes. For instance, in the studies of Kramarski and colleagues, students received a worksheet with metacognitive questions which they could ask themselves such as: "What is the problem all about?" and "Which strategies are appropriate for solving the problem?" Additional to the general metacognitive questions, during and after the solution process specific questions were provided about the word problem such as "Why didn't the slope of the line change?" In the study of Aleven and Koedinger, students were prompted to explain their solution path a by asking them questions such as "Some rules dealing with parallel lines are highlighted in the Glossary. Which of these reasons is appropriate?" The hints got progressively more specific about the cognitive content when the student could not provide the right explanation. As these examples show, the studies connected guidance by student-directed metacognitive questions and prompts to the cognitive suggestions about how to solve the problems. Effects of the metacognitive prompting were reported on improvement of metacognition (Harskamp & Suhre, 2007; Kramarski & Gutman, 2006; Kramarski & Mizrachi, 2006), as well as on near- and far transfer measures containing word problems (Aleven & Koedinger, 2002; Harskamp & Suhre, 2007; Kramarski & Gutman, 2006; Kramarski & Mizrachi, 2006). Moreover, all of these studies had a control group working with the same mathematical content but without hints. Thus, effects were not merely caused by practice but also by the metacognitive features of the programs. This shows that guided metacognitive training is a promising instructional tool to improve secondary students' performance.

However, in elementary education, guided metacognitive training studies are scarce. The metacognitive training studies which use some form of student-control specifically targeted low achieving students. And since low achieving students need extra support, these studies included teacher directed training as well. For instance in the studies of Teong (2003) and Chang, Sung and Lin (2006) low achieving students from grade five were first instructed by the teacher and later students could work on their own using questions and prompts in a computer environment. In the guided instructional phase, metacognitive questions were used such as “Have I read and understood what I am supposed to find?” and “Am I on the right track?” to help students to reflect upon their problem solving process (Teong, 2003). Findings of these studies in elementary education show that for low achievers, metacognitive training with a combination of direct instruction and computer supported guided training can improve problem solving performance.

Nevertheless, the effects of guided metacognitive training in a more inclusive sample of elementary school students have not yet been studied. Thus, the question whether guided training with metacognitive hints can be applied as an instructional tool to improve metacognition and problem solving performance in elementary school classes remains unanswered. To study this question, we developed a computer supported metacognitive training program based on the design characteristics described below.

4.2.3 Design characteristics for computer supported metacognitive training

As shown in the studies discussed above, computer environments are well suit for guided metacognitive training. When students work with a computer environment, hints, questions or prompts can be used to support their individual needs. From the literature, we summarize six characteristics of effective metacognitive training programs which can be used to design computer supported metacognitive training.

(a) Students should be given some freedom to choose the instructional support they need (Azevedo, 2007). In order to use hints or questions at their own initiative, students will have to judge their performance in order to evaluate which help they need. Giving students’ control of their problem solving process makes them better able to transfer the learned skills to new problem situations (Kapa, 2007; Moreno, 2009). Accordingly, the studies of Aleven and Koedinger (2002), Kramarski and Gutman (2006) and Kramarski and Mizrachi (2006) show that student-controlled metacognitive training has substantial effects on transfer measures of problem solving.

(b) Metacognitive questions should be used to guide students’ learning. Several studies have shown that letting students practice with metacognitive questions can improve their metacognition and problem solving performance (Aleven & Koedinger, 2002; Kramarski & Gutman, 2006; Kramarski & Mizrachi, 2006; Teong, 2003). Metacognitive questions can be used in addition to cognitive and metacognitive suggestions as seen in the studies of Harskamp and Suhre (2007), Aleven & Koedinger (2002) and Chang et al. (2006).

(c) Metacognitive support should be embedded in cognitive content. Thus, metacognitive hints or questions should refer to the word problem at hand such as is the case in the studies of Harskamp and Suhre (2007), Aleven and Koedinger (2002) and Chang et al. (2006). The computer program should not only provide general metacognitive questions and hints but also give hands-on suggestions about how to solve a problem.

(d) Metacognitive hints and/or questions should be provided for all episodes of problem solving. All of the training studies discussed in the previous paragraph provided metacognitive hints or questions for a variety of episodes, supporting the whole problem solving process. This has been identified as a success factor of metacognitive training in computer environments (Azevedo, 2007; Kapa, 2007).

(e) In the studies of Harskamp and Suhre (2007) and Aleven and Koedinger (2002) students were presented with graphic representations of the word problems. Research on the multimedia principle has shown that supporting domain specific facts or concepts by means of representational graphs can support problem solving (Clark & Mayer, 2008). Graphic problem representations can be used as an extra aid to support students' problem solving.

(f) Concerning the length of the training, a comparison of the guided training studies shows that most programs were implemented for a period of four to six weeks. This coincides with the finding from a meta-analysis of metacognitive training programs where it was shown that duration of a month or more is positively related to effects on outcome measures (Hattie, Biggs, & Purdie, 1996).

Following these design characteristics we can conclude that, if designed well, guided metacognitive training with cognitive and metacognitive support can help students to develop their metacognition and to transfer their approach to improve performance on different types of problems. How the six features mentioned above are given shape in our program will be further elaborated in the method section.

4.3 Hypotheses

Based on the effectiveness of guided training reported in the literature and the design characteristics discussed above, we have formulated the following hypotheses:

(1) Training word problem solving with metacognitive hints and specific prompts will improve students' metacognition more than mere practice with word problems.

(2) Training word problem solving with metacognitive hints and specific prompts will improve students' word problem solving performance more than mere practice with word problems.

4.4 Method

4.4.1 Sample

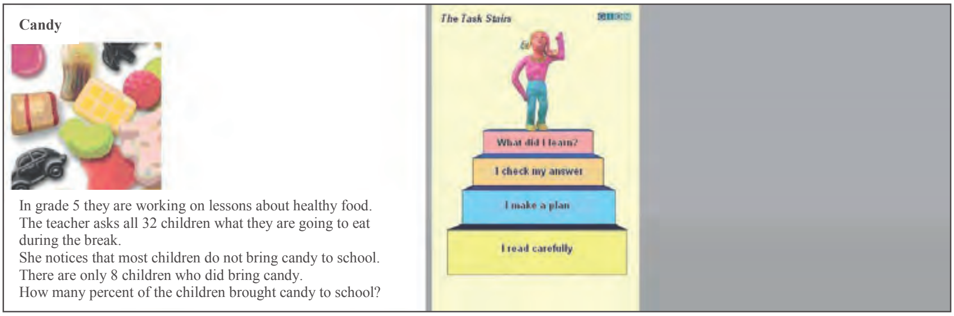
A total of 73 grade five students of three middle sized schools in the Netherlands participated in the experiment in the spring of 2010 (35 boys, 38 girls). Students with extremely low mathematical knowledge (less than 25% correct in a test with non-textual computation problems) were left out of the study. The mean age of the students was 10.9 years (*SD* = 0.38). Students within classes were randomly assigned to the experimental conditions. There were 36 students in the experimental group (18 boys) and 37 students in the control group (17 boys). The two experimental groups were part of a larger study on word problem instruction.

4.4.2 Training materials

For this study, an adaptation of a computer program with metacognitive hints applied in previous research is used (Jacobse & Harskamp, 2009). A first study has shown that this program is suitable for supporting elementary school students' word problem solving, but the effect of metacognitive hints versus mere practice has not yet been studied. The adaption which has been made in the computer program is that metacognitive hints are presented by spoken voice, since this may be more motivating for students (Mayer, Sobko, & Mautone, 2003). Furthermore, the substeps within the 'read/analyze' and the 'plan' hint were removed to reduce the number of actions students have to take to listen to the hints. In the version of the computer program used in this study (see Figure 4.1), metacognitive hints are presented in the form of the so called 'Task Stairs' with the problem solving episodes: Read/analyze ('I read carefully'), explore and plan ('I make a plan') implement ('I check my answer') and verify ('What did I learn?'). The episodes are presented in personalized language to motivate students.

In reference to design characteristic (a) students are given some freedom to choose which hints they use. This is expected to trigger monitoring behavior and to support transfer. In the program, the 'read/analyze' ('I read carefully') and the 'explore and plan' ('I make a plan') hint could be freely selected by the students. As suggested in characteristic (b) metacognitive support in the computer program is provided by a combination of metacognitive prompts and questions. Presenting the image of the Task-stairs with the different problem solving episodes in itself can be seen as a metacognitive prompt for students to structure their learning over the problem solving episodes (c.f. Harskamp & Suhre, 2007). Additionally, when students judge they need further support they can choose a hint. Students first have to select in which cognitive episode they need help by clicking one of the steps of the Task stairs (see Figure 4.1), and then get the metacognitive support that goes with the cognitive activities in the episode. Metacognitive questions and prompts are given such as: "What is the question you have to answer?" (read/ analyze) and "Don't forget to write down your calculation neatly" (plan). Although these metacognitive questions and prompts are quite general in nature, they are presented embedded in the episodes and connected to cognitive suggestions specifically aimed at the word problem abiding to design characteristic (c). Furthermore, hints are presented for episodes throughout the whole problem solving process (characteristic d). And as suggested in characteristic (e), students' problem representations are supported by providing graphic representations of the problems in the hints. Lastly, students will practice with the computer program over a period of six weeks (characteristic f).

In Figure 4.1 the interface of the computer program is shown. After logging in and choosing a task from the start frame, the word problem is presented on the left side of the frame. The computer program consists of a total of 72 novel (primarily multistep) word problems developed for grade five that need addition, subtraction, multiplication, or division to solve. The word problems are slightly more complex than word problems in standardized mathematics tests used by most schools in the Netherlands (Janssen & Engelen, 2002). Students work through the word problems in their own pace and word problems are of progressive difficulty. Students of higher ability might pass the first tasks more quickly, but eventually the word problems will have a level of complexity for which the use of metacognitive support is assumed to be suitable for all students (Prins, et al., 2006).



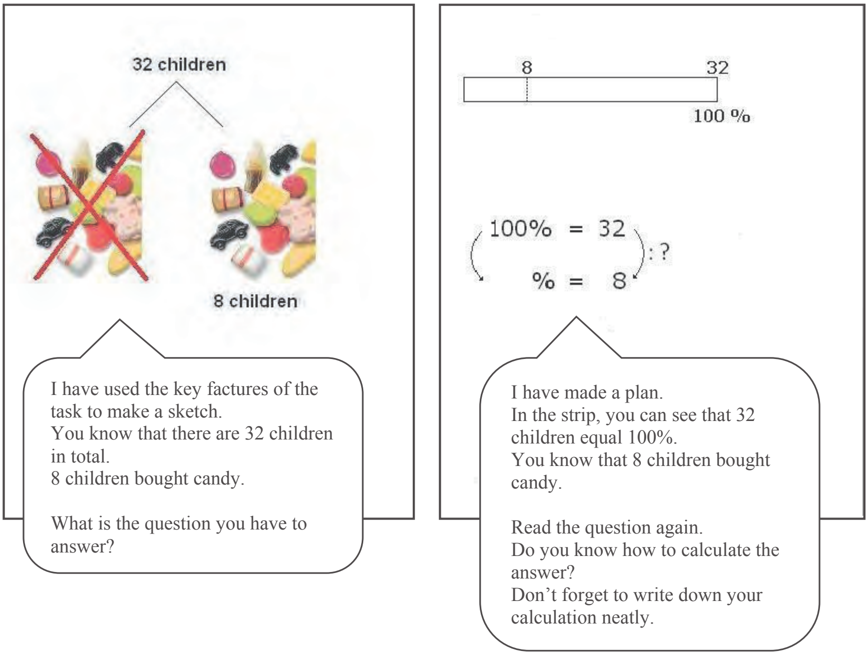
Candy

In grade 5 they are working on lessons about healthy food.
The teacher asks all 32 children what they are going to eat during the break.
She notices that most children do not bring candy to school.
There are only 8 children who did bring candy.
How many percent of the children brought candy to school?

Figure 4.1 Interface with word problem in Task Stairs computer program

When students choose a hint, the hint appears on the right side of the screen, the grey area in Figure 4.1. The first hint students can choose by clicking the bottom step of the Task stairs is the 'I read carefully' hint (see Figure 4.2). In this hint, a visual representation of the main features of the word problem is given, added by spoken text which starts when the student presses a button. This 'read/analyze' hint is intended to help students to create a situational model by analyzing the problem and drawing students' attention to key information. For instance, the metacognitive question "What is the question you have to answer?" prompts students to define the goal of the problem themselves.

In the ‘I make a plan’ hint (Figure 4.2), multiple mathematical models representing the word problem are provided to support students’ solution plans. The modeling type of instruction used in the description of the models is direct, but the use of the hint is student-directed. And within the hint students need to choose which approach they want to use and how to apply this to reach an answer. Students are prompted to monitor if they are able to perform the task and to use metacognitive regulation to analyze the task, to compare their plan to the actual question and to write down their calculation.



I have used the key factures of the task to make a sketch.
You know that there are 32 children in total.
8 children bought candy.

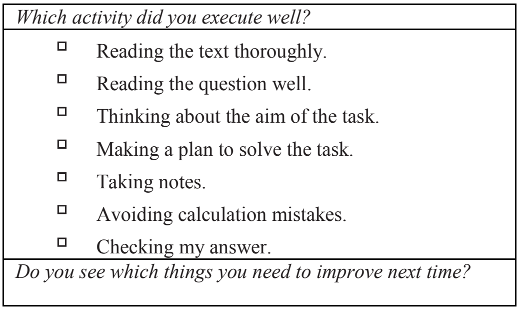
What is the question you have to answer?

I have made a plan.
In the strip, you can see that 32 children equal 100%.
You know that 8 children bought candy.

Read the question again.
Do you know how to calculate the answer?
Don’t forget to write down your calculation neatly.

Figure 4.2 Example of an ‘I read carefully’ (left) and an ‘I make a plan’ hint (right) with spoken text which starts when a student clicks the button.

After executing their calculation plan, students from the experimental group can click the ‘I check my answer’ step of the Task stairs. In this hint students are encouraged to check their calculation. When they think they have the right answer, they fill it in here. Students get feedback about whether their answer is correct or incorrect. They get three chances to fill in a correct answer. After that the program proceeds to the next episode: ‘What did I learn?’ In this episode a brief model answer of the problem is provided so students can compare this to their own problem solving. After viewing the model answer, a popup comes up asking students to reflect on their problem solving by checking boxes. The text presented in the popup is shown in Figure 4.3.



Which activity did you execute well?

- ☐ Reading the text thoroughly.
- ☐ Reading the question well.
- ☐ Thinking about the aim of the task.
- ☐ Making a plan to solve the task.
- ☐ Taking notes.
- ☐ Avoiding calculation mistakes.
- ☐ Checking my answer.

Do you see which things you need to improve next time?

Figure 4.3 Reflection popup in the hint 'What did I learn'.

4.4.3 Instruments

Implementation. In order to assess the implementation of the training, information was recorded about students' attendance of the training lessons. Additionally, log files were collected from the computer program, which give information about the number of problems students solved and the hints they used to solve the problems.

Performance judgments. As a metacognitive measure, we used concurrent performance judgments (Schraw, 2009). Students were given four word problems. They made judgments before solving each word problem about how well they thought they could solve the problem. This provides us with an indicator of metacognitive monitoring relevant to word problem solving (Boekaerts & Rozendaal, 2010). Students were asked to predict their performance in relation to four paper and pencil word problems before the start of the training (pretest) and four word problems after the training (posttest). For each word problem, they were asked to fill in a traffic light with 3 options: Red (I am sure I will execute this problem incorrectly), orange (I am not sure whether I will execute this problem correctly or incorrectly) and green (I am sure I will execute this problem correctly). Students read the word problem in 30 to 40 seconds time (depending on the length of the text) and directly afterwards judged their ability to solve it. Then they moved on to solve the problem at hand.

The word problems used for the performance accuracy measurement were different from those in the problem solving outcome measure. This is in line with the advice not to use the same but comparable items to avoid biasing the results (Marsh, Roche, Pajares, & Miller, 1997). An example of a word problem used for collecting performance judgments is added below.

A tourist wants to travel 120 kilometers. He walked the first 5 kilometers. Then he got a ride from a truck driver.
When the truck driver dropped him off, he still had half of his journey to go.
How many kilometers did the tourist travel in the truck?

By comparing students' performance judgments to their actual performance on each word problem, a bias index score is calculated for each student. The bias index, classified as an absolute accuracy measure, gives insight in students' average degree of over- or under confidence in predicting their own performance (Schraw, 2009). For a correct prediction a score of 0 is given while overconfidence is represented by a positive score and underconfidence is represented by a negative score. The scoring scheme used in this study is presented in Table 4.1.

Table 4.1.
Scoring scheme bias index

	Answer correct	Answer incorrect
Prediction correct (green)	0	+ 100
Prediction uncertain (orange)	- 50	+ 50
Prediction wrong (red)	- 100	0

The formula used for calculation of the bias index (Schraw, 2009) is:

$$1/n \sum_{i=1}^{n}\left(c_i - p_i\right)$$

c_i = confidence rating p_i = performance n = number of problems

Word problem solving. As a mathematical problem solving measure, scores on two word problem tests for mathematics were used (pretest/ posttest). The word problems in the tests were structurally comparable but more complex than word problems from Dutch standardized tests used in most elementary schools (Janssen & Engelen, 2002). Both word-problem tests consist of 20 novel word problems which were not used in the computer program. The items typically require good regulation and flexible use of strategies to give the right answer. For instance in the sample problem below, students who don't regulate their learning well might be inclined to give the answer "3/4th part is 45" while the answer should be "15" instead. The tests contained addition, subtraction, multiplication, and division problems. Word problems in the posttest had different content than items in the pretest but did have the same structure.

Jonathan has 60 stickers.
One day, he gives ¾ of the stickers to Tobias.
How many stickers does Jonathan have left?

Students got 0 points for an incorrect answer and 1 point for a correct answer. Students on average solved 6.86 word problems correctly on the pretest (*SD* = 4.78). The reliability of the pretest as indicated by Cronbach's α is .87. The posttest has a reliability of α = .84.

4.5 Procedure

4.5.1 Procedure of the experiment.

The study is an experimental design with students within classes being randomly assigned to the conditions. Students in the experimental condition practiced with the computer program with word problems and metacognitive hints. Students in the control condition practiced with a computer program with the same word problems but without metacognitive hints. In the control condition students could fill in an answer and they received feedback about if their answer was correct or incorrect, but no additional support was provided. Before working with the computer program, students from the experimental group received a short preflight training informing them about how to use the computer program and about the rationale for using the hints. Control group students were only shown how the program works. Hereafter, students in both groups worked on the computer individually for two times a week (approximately 20 minutes each time) during six weeks. During this time, a researcher was present to keep order but students were encouraged to work on their problem solving tasks independently. The problem solving practice in the computer program was layered on top of the regular mathematics lessons of the teacher in which some word problems were used but no explicit instruction was given about problem solving processes.

Pretests were collected about one week before the start of the intervention. In the pre-measurement students individually completed the word problem test and four word problems with prediction judgments in the classroom. Posttests were collected about a week after the experiment took place. The procedure in the post-measurement was comparable to the pre-measurement.

4.5.2 Data-analysis.

Before addressing the hypotheses, preliminary analysis are performed to check for initial differences between the groups and to check for implementation of the training. Then, to assess the hypotheses, the following analyses are planned:

Hypothesis 1) To analyze the effect of the computer supported metacognitive training on metacognition, experimental and control students' performance judgments are compared. This is done by comparing students' accuracy of performance judgments controlled for pretest performance judgment scores in an analysis of covariance (ANCOVA). The data is also checked for an interaction effect between pretest and condition to assess if the training equally affects students with varying prior word problem solving abilities.

Hypothesis 2) To analyze the hypothesized effect of metacognitive training on word problem solving performance, irrespective of prior performance, analyses of covariance (ANCOVA's) are performed with two dependent variables: Performance in the computer program and performance on the word problem solving posttest. Effects on the first dependent variable show if students' word problem solving during practice improves by using the metacognitive hints. The second analysis shows whether this effect transfers to the posttest where students need to structure their own word problem solving without guidance. The data is also checked for an interaction effect between pretest and condition to assess if the training affects students with varying prior word problem solving abilities equally.

4.6 Results

4.6.1 Preliminary analyses

Comparability of the groups. As a check of the randomization process, mean scores on the pretests for word problem solving and performance judgments were compared over the two conditions. In the mathematical problem solving pretest the experimental group solved an average of 7.55 problems (*SD* = 5.04) and the control group solved 6.18 problems correct (*SD* = 4.47). For performance judgments in the pretest, the experimental group had a mean bias of 29.93 (*SD* = 33.41) while the control group had a mean bias of 41.21 (*SD* = 25.68). So, students in both groups were somewhat overconfident in their performance judgments. The differences in the pretest are not significant for problem solving (*F*(1,71) =1.50, *p*=.23) nor performance judgments (*F*(1,71) = 2.62, *p* = .11). This shows that the randomization processes went well. However, we will also correct for the pretest by using the it as a covariate in analyses of the posttest results.

Implementation. Students in both conditions were intended to practice with the computer program for a total of 12 lessons. This was indeed the case in both groups (experimental group *M* = 11.7 lessons, *SD* = 0.6; control group *M* = 11.7 lessons, *SD* = 0.7). Students on average completed 63 word problems (out of 72) in the computer program (*SD* = 8.1). Students in the experimental condition completed less problems (*M* = 59.7, *SD* = 9.2) than students in the control group (*M* = 66.6, *SD* = 5.2). This is to be expected since using the hints takes a bit more time than working without hints.

On average, the metacognitive hints in the experimental condition were used for about 56 percent of the word-tasks. Students on average consulted 34 percent (*SD* = 16) of the 'I read carefully' hints and 34 percent (*SD =* 16) of the 'I make a plan' hints. Most hints were well used. However, the last hint 'What did I learn', was not always used in the intended manner. The majority of students directly turned away from this hint without reviewing the model answer.

4.6.2 *Training Effects*

Effect of metacognitive training on performance judgments. In the pretest, students of both groups tended to overestimate their performance. But, analysis of covariance shows that in the posttest, there was a significant main effect of the intervention on the accuracy of performance judgments (*F*(1,70) = 8.06, *p* = .01). In the posttest students in the experimental group on average judged their performance more accurately and overestimated less than students in the control group (see Table 4.2). The experimental metacognitive training condition had a medium to large effect on performance judgments (*d* = .67). This effect was found for students of all levels of mathematical knowledge on the pretest, no interaction effect of the intervention with pretest scores was found (*F*(1,69) = .20, *p* = .66).

Table 4.2.
Corrected means and standard errors of the overestimation of performance judgments on the posttest for the control- (n=37) and the experimental group (n=36).

	*M * (SD)*	*SE*
Control group	43.27 (23.66)	3.89
Experimental group	27.40 (23.66)	3.94

Note. * corrected for a pretest value of 35.65

Performance judgments were related to problem solving performance in the word problem post-test. The negative correlation shows that the less students overestimated their performance, the higher scores they reached in the word problem solving test (*r(72)* = -.39, *p* = .00).

Effect of metacognitive training on mathematical problem solving. Analyses of students' word problem solving in the computer program show that students in the experimental group outperformed students in the control group on number of correctly answered problems in the computer program (*F*(1,71) = 10.91, *p* = .00). Corrected for the word problem pretest, we see that the control group solved a mean of 51 percent (*SD* = 19) of the total number of problems while the experimental group solved 62 percent (*SD* = 20) of the word problems correctly (*F*(1,70) = 12.01, *p* = .00). The effect of the experimental condition on the percentage of correctly solved problems in the computer program was large (*d* = .82). Hint-use in the program was negatively related to students' success in the problem solving pretest (*r*(72) = -.64, *p* = .00). Below average students used hints for 78% of the word problems, while above average performers used the hints for 42% of the word problems. This indicates that students actually monitored if they needed the student-directed hints. Moreover, the percentage of correctly solved problems in the computer program was strongly and significantly related to scores on the independent word problem solving posttest (*r*(72) = .78, *p* = .00).

On the independent word problem solving test, paired samples t-tests show that both the control group (*t*(36) =7.31, *p* = .00) and the experimental group (*t*(35) = 7.67, *p* = .00) on average increased their performance from pre- to posttest. Moreover, in line with hypothesis 2, the experimental group outperformed the control group (*F*(1,70) = 5.21, *p* = .03). Students working with the computer supported metacognitive training program solved more problems than students only practicing word problems (see Table 4.3). The metacognitive training program had a medium effect of *d* = .54 on problem solving performance. No interaction effect was found between the experimental conditions and the problem solving scores on the pretest (*F*(1,69) = .02, *p* = .89), indicating that the effect of metacognitive hints affects the problem solving of students of varying mathematical levels in the sample.

Table 4.3.
Corrected means and standard errors of the number of correct answers on the problem solving posttest for the control- (n=37) and the experimental group (n=36).

	M (SD)*	*SE*
Control group	8.53 (2.56)	.43
Experimental group	9.90 (2.56)	.42

Note. * corrected for a pretest value of 6.86

4.7 Conclusion and Discussion

In the literature on word problem solving, metacognition is considered to be a key factor affecting performance. In line with this assumption, we studied if word problem training with metacognitive questions and hints in a computer environment could enhance students' metacognition, which in turn would positively affect their problem solving performance. In secondary education, several studies have already shown that guided metacognitive training has strong benefits for students' word problem solving. However, in elementary school mathematics there is a paucity of guided metacognitive training studies. Up to date, it was unclear if computer supported metacognitive training could benefit word problem solving of elementary school students of varying abilities just as it did in secondary education. This issue was evaluated with an experimental design in a grade five sample. In the design, performance of an experimental group practicing with a computer program with metacognitive hints was compared to the performance of a control group practicing with a computer program with only word problems.

Based on the metacognitive training literature, we hypothesized that guided metacognitive training could positively affect students' metacognition (hypothesis 1). More specifically, we tested if training with a computer program with metacognitive hints could affect students' metacognitive monitoring by measuring the accuracy of their performance judgments. Some studies in secondary education have already reported effects of metacognitive training on monitoring (Kramarski & Gutman, 2006; Kramarski & Mizrachi, 2006). In line with our hypothesis, the logfiles show that students' hint-use is affected by their prior performance. This is a first indication for monitoring behavior. Additionally, in the posttest a main effect was found of the metacognitive computer training on the accuracy of students' performance judgments (*d* = .67). Practice with the computer environment with student-directed metacognitive questions and prompts helps students to better monitor their performance. This is true for students of varying abilities. Monitoring accuracy in turn is related to word problem solving performance.

Consequently, we expected that adding metacognitive hints to a computer environment with word problems would positively affect students' word problem solving performance (hypothesis 2). In a previous study (Jacobse & Harskamp, 2009), the computer program with metacognitive hints was already shown to positively affect students' word problem solving performance. However, it had not yet been experimentally studied if the metacognitive hints in the computer program affect problem solving performance beyond the effect of practice. Findings from the present study show that practice with word problems in the computer program benefits students of both the control and the training group. But besides this, working with the metacognitive hints improves problem solving additive to the effect of practice. Even when students in the experimental group practice with fewer problems in the computer program as was the case in our study, their problem solving significantly improves. This conclusion can be supported by the following arguments: Firstly, students in the experimental group largely outperform students in the control group on amount of correctly answered items in the computer program (*d* = .82). This shows that the metacognitive hints support students problem solving during the training. Secondly, to assess if students internalized the metacognitive procedures enough for transfer to novel word problem solving, results on an independent mathematics posttest were analyzed. Analyses of the word problem post-test show students in the experimental group to outperform the control group on the word problem test with an effect size of *d* = .54. No interaction between prior performance and the effect of the training was found. Since students of lower prior ability were found to use significantly more metacognitive hints, it seems that this helps them to compensate for their initial deficits.

Our findings on the effectiveness of guided word problem solving training in elementary mathematics education are thus promising in respect to increasing metacognition and problem solving performance. This is in line with studies in secondary education (Aleven & Koedinger, 2002; Harskamp & Suhre, 2007; Kramarski & Gutman, 2006; Kramarski & Mizrachi, 2006). However, there are some limitations which should be kept in mind. One limitation of the study is that we did not include a follow-up measure to assess long-term effects of the metacognitive training. Results of follow-up tests and standardized transfer measures should be included in further studies to assess the robustness of the training effects. Our results show that when students solve problems independently, the effect of the metacognitive training slightly decreases. This is not surprising since the metacognitive hints support students to solve problems. However, further studies should evaluate which training duration is best for students to fully internalize the procedures suggested in the training program. Furthermore, we recommend collecting more information on students' problem solving processes during the program. Although we found evidence that the metacognitive hints positively affect students' performance judgments and word problem solving skills, we do not have data for more fine-grained analyses of students' thought processes while using the hints. For instance, no information is gathered about the metacognitive thoughts students are expected to exert while using the hints, such as “I will click the planning hint because I am not sure how to get to the calculation”. In a follow-up study, recording think-aloud protocols or collecting trace-data during problem solving in the computer program could help fill this gap (Azevedo, Moos, Johnson, & Chauncey, 2010; Greene & Azevedo, 2010). Such extra measures could also be used to address a final limitation of our study concerning the metacognitive measure we used. In order to assess outcomes of the training on metacognition, we used students' performance judgments. These are primarily indicative for monitoring behavior and possibly also for already thinking ahead about an initial plan. But, predictions do not directly address all regulatory activities students can use to solve problems. Although monitoring and regulation of thought processes are closely linked (Nelson, 1996; Roebers, Schmid, & Roderer, 2009), the current design of the study does not yet allow fully defining the relationship between the improvement of students' metacognition and improved problem solving.

To summarize, our study shows that it is possible to enhance word problem solving performance of a broad variety of word problems for all students in upper elementary school by means of guided metacognitive training. In the training, a problem solving structure was presented (Task Stairs) with metacognitive hints. This improved students' metacognitive behavior. Students became more accurate in monitoring their own performance by using the hints. We suggest that metacognitive thinking for problem solving is best served through a guided training approach. But a mix of both a direct and guided training may be most profitable for low-achievers. We recommend further studying how metacognitive processes unfold during the problem solving process and how this affects students' cognitive activities.

CHAPTER 5

Towards Efficient Measurement of Metacognition in Mathematical Problem Solving

ABSTRACT

Metacognitive monitoring and regulation play an essential role in mathematical problem solving. Therefore, it is important for researchers and practitioners to assess students' metacognition. One proven valid, but time consuming, method to assess metacognition is by using think-aloud protocols. Although valuable, practical drawbacks of this method necessitate a search for more convenient measurement instruments. Less valid methods that are easy to use are self-report questionnaires on metacognitive activities. In an empirical study in grade five (n = 39), the accuracy of students' performance judgments and problem visualizations are combined into a new instrument for the assessment of metacognition in word problem solving. The instrument was administered to groups of students. The predictive validity of this instrument in problem solving is compared to a well-known think-aloud measure and a self-report questionnaire. The results first indicate that the questionnaire has no relationship with word problem solving performance, nor the other two instruments. Further analyses show that the new instrument does overlap with the think-aloud measure and both predict problem solving. But, both instruments also have their own unique contribution to predicting word problem solving. The results are discussed and recommendations are made to further complete the practical measurement instrument.

This chapter is based on the article:
Jacobse, A.E., & Harskamp, E.G. (in press). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning.*

5.1 Introduction

In the past years, metacognition has been recognized as one of the most relevant predictors of accomplishing complex learning tasks (Van der Stel & Veenman, 2010; Dignath, & Buttner, 2008). Metacognition refers to meta-level knowledge and mental actions used to steer cognitive processes. In our study, we adopt the view of applied metacognition as consisting of metacognitive monitoring and regulation (Efklides, 2006; Nelson, 1996). Metacognitive regulation refers to mental activities used to regulate cognitive strategies to solve a problem (Brown & DeLoache, 1978). For instance, when taking a note, the decision to do so is metacognitive, while the writing in itself is cognitive. Metacognitive monitoring refers to students' ongoing control over these learning processes. Monitoring can be used to identify problems and to modify learning behavior when needed (Desoete, 2008). A large number of studies have already been undertaken to show that through metacognitive training, students' ability to solve mathematics problems improves (i.e. Jacobse & Harskamp, 2009). For researchers, as well as teachers, it is important to have an adequate instrument to measure students' metacognition in order to analyze the relationship between growth in metacognition and growth in achievement. However, how to measure metacognition efficiently is still a problem. This problem has been at the heart of a great deal of scientific debate about which instruments are most suitable (Schellings & Van Hout-Wolters, 2011).

One proven effective method to get insight into students' metacognition is asking them to verbalize their thoughts while working on a task. The verbalized thoughts are recorded and fully transcribed or judged by means of systematical observation (Veenman et al., 2005). This measurement technique is called think-aloud. Think-aloud protocols provide rich information on the metacognitive processes used during a learning task and are powerful predictors of test performance (Schraw, 2010; Veenman, 2005). A major strength of the use of think-aloud protocols, is that information about metacognitive behavior is collected directly when it is executed. This makes the information less vulnerable to students' memory distortions. Besides, students do not have to judge the appropriateness of their learning processes themselves (Veenman, 2011b). Although sometimes slowing learning down, when executed correctly think-alouds do not impair students' learning performance (Bannert & Mengelkamp, 2008; Fox, Ericsson, & Best, 2011). However, besides these positive characteristics, there is a major drawback of the method: Gathering and scoring the data of individual students' think-aloud protocols is a complex and time-consuming process which makes this measure inappropriate for test assistants or teachers who lack experience using the method, and for application in larger samples of students (Azevedo et al., 2010; Schellings, 2011). Thus, when using this theoretically grounded measure, it tends to conflict with some more practical constrains of time and effort. Balancing theoretical and practical issues in the measurement of metacognition is a particularly challenging issue (McNamara, 2011). In order to make measurements of metacognition more practical, it is important to explore the use of other instruments.

Researchers have already proposed several alternative measurement instruments to assess metacognition in a more practical manner, such as various self-report questionnaires. However, few of these instruments show convergence with think-aloud measures as predictors of performance. In this study pros and cons of alternative instruments are discussed that may substitute think-aloud protocol analysis. Alternative instruments which are shown in the literature to be valid indicators of metacognition are combined into a new measurement instrument. This measurement instrument can be collected in a paper-and-pencil format for larger groups of students which makes it notably easier to use than think-aloud measures. Explorative analyses comparing the new instrument with think-aloud scores are performed in a grade 5 sample, eventually aiming at the development of a more practical measurement instrument of students' metacognition in mathematics.

5.2 Theoretical Framework

When measuring metacognition, it is important to note that metacognition probably is quite domain-specific (Veenman & Spaans, 2005). The regulation of cognitive activities useful in one domain (e.g. making a summary when reading) may not be directly transferable to another domain (e.g. solving a math problem). It is thus advisable to be specific about the context in which metacognition is measured (McNamara, 2011). One of the domains in which metacognition is a key variable predicting learning performance is the domain of mathematical problem solving (Desoete & Veenman, 2006; Desoete, 2009; Fuchs et al., 2010; Harskamp & Suhre, 2007). In this domain, metacognition is used to monitor solution processes and to regulate the problem solving episodes of analyzing and exploring a task, making a solution plan, implementing the plan and verifying the answer (Schoenfeld, 1992). Such metacognitive processes can be measured off-line or on-line of the learning process. Online methodologies capture any activity that occurs during processing, whereas offline methods capture any activity that happens either before or after processing (Azevedo et al., 2010). Metacognition measured on-line of the learning process typically explains about 37 percent of the variance in learning (Veenman, Van Hout-Wolters & Afflerbach, 2006).

One of the most frequently used categories of off-line measures is self-report questionnaires in which students are asked to report on their own metacognition. Some examples of frequently used questionnaires are the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich & De Groot, 1990), the Learning and Study Strategies Inventory (LASSI; Weinstein, Zimmermann, & Palmer, 1988) and the Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994). These questionnaires typically contain quite general statements about metacognitive monitoring or regulation for which the student is asked to rate the degree to which the statement applies. Statements are used such as: "Before I begin studying I think about the things I will need to do to learn" or "I ask myself questions to make sure I know the material I have been studying" (Pintrich & De Groot, 1990). One notable practical advantage of using questionnaires is that they can easily be administered on a large scale (Schellings & Van Hout-Wolters, 2011). Besides, various studies in mathematical problem solving have shown the practicality and good internal consistency of self-report questionnaires (Kramarski & Gutman, 2006; Mevarech & Amrany, 2008). However, off-line measures do not measure learners' ongoing metacognitive behavior during task processing because they are collected before or after the student processes a learning task (Greene & Azevedo, 2010). This causes some severe problems. Firstly, the fact that self-report questionnaires are collected separate from the learning task means that students have to retrieve earlier processes and performance from their long term memory. Self-report questionnaires thus are susceptible to memory distortion issues (McNamara, 2011; Schellings, 2011; Veenman, 2011b). Secondly, students can differ in their frame of reference as to which situations they have in mind when answering the questions and interpreting the scales (McNamara, 2011; Schellings, 2011). Thirdly, the way students answer self-report questionnaires may be biased by triggers in the questions which prompt them to wrongly label their own behavior or by social desirability (Cromley & Azevedo, 2011; Veenman, 2011a). Therefore, students are typically quite inaccurate in reporting their own metacognitive behavior. Although self-report questionnaires are mostly designed to measure metacognitive regulation, they do not seem to be representative of what students actually do. This is illustrated by the fact that students' self-reported metacognitive behavior has found to be a poor predictor of performance. In a review of 21 studies using self-report questionnaires, the mean variance explained by metacognition in learning performance did not exceed 3% ($r = .17$) (Veenman & Van Hout-Wolters, 2002). Additionally, some studies have shown the convergent validity between different questionnaires, theoretically measuring the same metacognitive processes, to be quite modest (Muis, Winne, & Jamieson-Noel, 2007; Sperling, Howard, Miller, & Murphy, 2002). As some authors argue, off-line, generally formulated metacognitive questionnaires may be more adequate to assess metacognitive knowledge as opposed to metacognition applied during the learning process (Desoete, 2007; Greene & Azevedo, 2010).

On-line measures on the other hand have the advantage of measuring metacognition concurrent with the learning behavior, thus giving more insight in the actual use of metacognition affecting learning behavior. One way to infer on-line information about students' metacognition, apart from using think-aloud protocols as discussed before, is to assess the actions or observable occurrences of events that a student performs such as drawing schemes, taking notes or clicking a button (Winne & Perry, 2000). Although in this case no direct information is gathered about the meta-level processes preceding the event, certain characteristics of the actions can be used to infer this information. In mathematical problem solving, an important cognitive action is making a drawing of the problem situation. Few students in elementary school use this strategy spontaneously. However, instructing students to make a drawing, can clarify how they think about solving a word problem (Van Essen & Hamaker, 1990). Students' problem visualizations in a drawing can be either schematic or pictorial. In schematic visualizations the structural relationships between variables in a problem are provided in a sketch, diagram or schema. In pictorial visualizations the elements in a problem are depicted without any relevant relationships between the elements. Pictorial visualizations show a student does not yet know how to explore the problem towards a useful solution, thus indicating low metacognitive regulation. Visualizations that schematize problem situations on the other hand, are an expression of sophisticated metacognitive regulation in mathematical problem solving, especially giving insight in the episodes of analyzing and exploring a problem (Schoenfeld, 1992; Veenman et al., 2005). Research has shown that schematic versus pictorial visual representations have good predictive validity for students' problem solving performance (Cox, 1999; Edens & Potter, 2007; Hegarty & Kozhevnikov, 1999; Van Essen & Hamaker, 1990; Van Garderen & Montague, 2003). The correlation between the use of schematic visualizations and problem solving in mathematics ranges from about r = .3 (explained variance 9%) (Edens & Potter, 2007) to about r = .7 (explained variance 49%) (Van Garderen & Montague, 2003). So the predictive validity - the relation with problem solving performance as would be expected based on theory – of using the quality of problem visualizations as an indicator of metacognitive regulation seems to be in order. But, using problem visualizations as a metacognitive measure does not cover metacognition over all episodes of problem solving. To avoid underrepresentation of the construct, it is wise to add additional on-line information.

Another way to collect information on metacognitive processes on-line of the learning task is through performance (or calibration) judgments (Schraw, 2009). More specifically, by assessing the accuracy of students' judgments of their own performance. The ability to judge one's performance has been conceptualized as an expression of metacognitive monitoring behavior (Boekaerts & Rozendaal, 2010; Efklides, 2006). When making on-line prediction judgments, that is to say estimations about performance before solving a problem, a student is especially concerned with the question whether he/she can analyze and categorize a problem. This gives the student a general idea whether he/she will be able to solve the problem or not. And a student may already briefly think ahead about a possible solution plan. There are also 'postdiction judgments' made after problem solving. By making a postdiction the student monitors if he/she has solved the problem correctly and adequately (Desoete, 2009). Research has shown the accuracy of performance judgments before and after problem solving to have good predictive validity for mathematics performance. In the literature, correlations between judgments of performance and mathematics performance range from about r = .4 to .6 (explained variance 16% to 36%) (Chen, 2002; Desoete, Roeyers, & Buysse, 2001; Desoete, 2009; Vermeer, Boekaerts, & Seegers, 2000). The relationship is typically stronger when the performance measure is more closely related to the task on which the judgment is based (Pajares & Miller, 1995). But, since accuracy measures give insight into a limited part of metacognitive processes (monitoring by looking forward or looking backward and thinking ahead about a solution plan), it is recommendable to combine them with more measures of metacognitive regulation (Pieschl, 2009), such as the type of visualizations students make.

What do we know about the overlap between different measures of metacognition? Sperling and colleagues (2004) compared the accuracy of performance judgments to the MAI self-report questionnaire. Their findings with college students show correlations around zero or even negative correlations between the accuracy of the performance judgments and the questionnaire. In the same vein, Veenman (2005) reviewed different studies and concluded that there is hardly any correspondence between findings from different on-line measures and self-report questionnaires. This shows that self-report instruments are generally not linked to students' on-line use of metacognition. On the other hand, we have little knowledge about the convergence between on-line performance judgments, problem visualizations and think-aloud scores. Theoretically, we can make some comparisons. As argued above, we expect the quality of problem visualizations to be specifically indicative for the way students analyze and explore a problem towards a solution plan. Such activities are indicators of metacognitive regulation in the first episodes of the problem solving process. Making performance judgments on the other hand primarily draws on students' metacognitive monitoring behavior and possibly on an initial stage of planning a solution. In think-aloud protocols, students' metacognitive regulation and monitoring are recorded over all episodes of problem solving. We would expect a low to moderate correspondence between performance judgments and think-alouds, since monitoring behavior is only a small part of all metacognitive processes executed when solving a problem (compare the findings on off-line performance judgments of Desoete, 2008). And problem visualizations are not expected to cover metacognitive monitoring and regulation in the episodes of setting up and implementing a plan and verifying the solution, which are addressed in think-aloud protocols. So, theoretically, a think-aloud measure in word problem solving should show some overlap with visualizations, but should also have some unique predictive validity because it includes additional information about other problem solving episodes. Some additional differences between performance judgments and think-aloud scores may be caused by the fact that in think-alouds, metacognitive activities are measured which students perform without a specific assignment, while in the other on-line measures information is gathered about the quality of students metacognitive processes when instructed to perform certain actions. When comparing these different types of measures, it is important to use word problems with an adequate level of difficulty so students are enticed to use a varied set of metacognitive activities (Prins et al., 2006).

Since judgments of performance and problem visualizations theoretically measure different aspects of metacognition, but are both practical on-line measurement instruments with sufficient predictive validity, we suggest combining these measures into a new instrument. Collecting a combined measure of prediction judgments, postdiction judgments and visualizations of the problem on-line is meant to provide an indication of the intertwined process of metacognitive monitoring and regulation during problem solving. To study the relation of this newly combined measurement instrument with the other instruments discussed above we have formulated the following research questions:

(1) What is the convergence between an on-line prediction-visualization-postdiction instrument, a self-report questionnaire and an on-line think-aloud instrument measuring metacognition?
(2) Can the on-line prediction-visualization-postdiction instrument predict problem solving on an independent mathematical word problem test just as well as a think-aloud measure?

Based on the theoretical framework, we hypothesize there to be little to no convergence between the off-line, general self-report questionnaire and both on-line measures of metacognition in word problem solving. On the other hand, since the new instrument is collected as a practical on-line instrument measuring monitoring and regulation, we expect this instrument to show moderate convergence with the on-line think-aloud measurement. But, because of the rich information in the think-aloud protocols, this measure is hypothesized to explain the largest proportion of variance in mathematical problem solving.

5.3 Method

5.3.1 Sample

The study reports of a total of 42 students randomly selected from five grade 5 classes in middle sized elementary schools. These students were in the business as usual condition of a larger study. We determined that the sample size is sufficient for detecting moderate correlations (around *r* = .30) (Cohen, 1977). All students were of families with intermediate social economic status. Students scored a mean of 44.82 (*SD* = 5.61) on the Raven Standard Progressive Matrices test, showing them to be well comparable to the norm scores in the Netherlands of 42 for the fiftieth percentile 47 for the seventy-fifth percentile (Raven et al., 1996). The mean age in the sample was 10.91 years old (*SD* = .28). Over the days of testing, three students did not complete all measurements. So the effective sample of the study is 39 students (22 boys, 17 girls).

5.3.2 Instruments

Think-aloud measure. To collect think-aloud protocols, we used a 'type 2' procedure for verbal protocols (Ericsson & Simon, 1993). This means we asked students beforehand to think aloud during execution of the word problems. After students started working on the problems, test leaders only interfered with neutral comments urging students to keep verbalizing ("keep thinking aloud") when students silenced. Test leaders did not help the students to solve the problems in any way. The verbalizations of individual students' thought processes were recorded using a video camera. This way, a detailed report of the verbalizations could be collected, without fully transcribing the protocols. The think-aloud data were gathered as follows.

First each student performed one test problem while thinking aloud. This was intended to help students get used to the procedure and the camera. This problem is not taken up in the analyses. During the actual measurement, students got two word problems (one by one) which they were instructed to solve while thinking aloud. Before starting, students got note paper and a pencil which they could use on their own initiative. Students were instructed in advance to indicate when they thought they were completely ready with the problem to make sure they were not stopped untimely by the test leader.

The two multistep problems used for the think-aloud protocols are presented below. Both problems lend themselves well for a metacognitive approach and they have multiple possible solution paths for reaching the correct answer. Moreover, both problems were judged by three elementary school teachers as being rather difficult for fifth grade students so they specifically require a thoughtful approach (as opposed to atomized behavior).

Hans and Ans are driving on the highway to Amsterdam. *The highway has a gas station every 55 kilometers.* *Their car breaks down after 196 kilometers.* *Which gas station is the nearest, the previous one or the next?*	*Marie has bought a bag with 150 apples.* *She wants to give all children of grade 5 as many apples as possible.* *Grade 5a has 13 children and in grade 5b there are 15 children.* *Marie wants to give each child an equal amount of apples.* *She also wants to give 1 apple to the teacher of grade 5a and 1 apple to the teacher of grade 5b.* *How many apples will Marie have left?*

After having collected students' think-aloud protocols, each videotaped think-aloud session was assessed by two judges. The four judges received two hours of training in scoring the protocols. To rate the think-aloud protocols, a scoring scheme for systematical observation of think-aloud protocols was used (see Table 5.1). The scoring scheme was developed and tested by Veenman and colleagues (Veenman, Kerseboom & Imthorn, 2000; Veenman, Kok, & Blöte, 2005) and consists of activities which are characteristic for mathematical problem solving (Schoenfeld, 1992). Previous research in secondary education has shown the instrument to be reliable and to have high convergent validity with full protocol analysis in which all verbalizations are transcribed.

Each activity in the scoring scheme was judged based on the verbal expressions of students while executing the word problems. Some verbalizations were thoughts preceding an activity, for instance when students verbalized how they were thinking about a plan before starting a calculation. Others thoughts were verbalized during the process, for instance when students verbalized which information they selected from the text while doing so. Following the suggestion of the developers of the systematical observation scheme (Veenman et al., 2005), each activity was given a score ranging from 0 (not executed) to 1 (partially executed) to 2 (executed). An example of activity 6: Students got a score of 1 if they initiated a plan but do not follow through (for instance if a student would say "I am going to subtract" but gets distracted and does not carry on the planning into later solution steps). A score of 2 would be given for students who verbalize a worked out plan which they thought out before solving the problem (for instance saying: "First I need to subtract 13 by 5, and then I am going to divide by 2 to get the right answer"). Another example of activity 2: A student would get a score of 1 if he/she selects some numbers from the text but then quickly moves on (For instance by emphasizing information while reading aloud or by shortly repeating some of the numbers without concretely connecting this to a goal or plan). A score of 2 would be awarded if a student thoughtfully selects information for use in the calculation (For instance saying: "Let's see, what do I need to calculate the answer? I need to know that every person gets 2 eggs and that there are 12 eggs in each box").

Table 5.1.
Scoring scheme for systematical observation of think-aloud protocols in word problem solving

Episode		Activity
Read, analyze /explore (orientation)	1	*Reading carefully*
	2	*Selection of relevant information/ numbers*
	3	Paraphrasing the question
	4	Making a visualization or taking notes to orient on the task
	5	Estimating a possible outcome
Plan and implement (systematical orderliness)	6	*Making a calculation plan*
	7	*Systematically executing the plan*
	8	*Being alert for correctness/ sloppiness (monitoring the calculation)*
	9	*Writing down calculations neatly*
Verify (evaluation and reflection)	10	*Monitoring the process*
	11	Checking calculations and answers
	12	*Drawing a conclusion*
	13	Reflecting on the answer
	14	Reflecting on the learning experience

Note. Items in italic print were used to compute a sumscore.

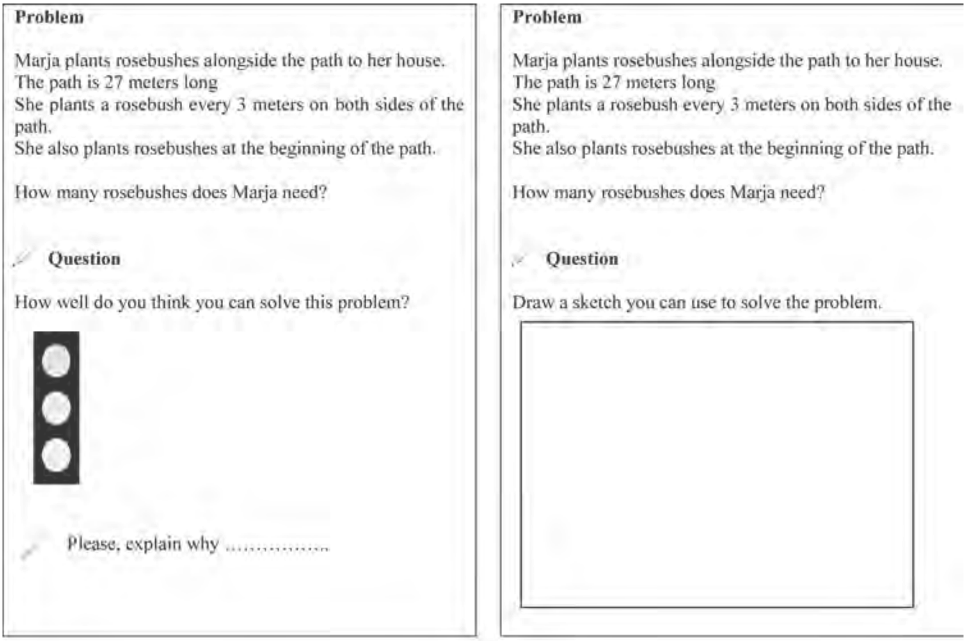
The raters first watched the video of a word problem performed by a student (pausing and rewinding when needed) and individually filled in the scoring scheme. After this, they rewound the video and watched the problem solving of the student a second time, using the video data to explain each other which scores they gave and why. For each activity, the two raters argued until agreement was reached about the definitive scores before moving on to the next activity. This is a common approach in the scoring of think-aloud data (c.f. Elshout, Veenman, & Van Hell, 1993; Veenman et al., 2000; 2004). Observation of students' scores on the items of the instrument shows that some regulation activities were not used by the relatively young students in the sample. For both word tasks, activities 3, 4, 5, 11, 13 and 14 showed little to no variance with almost all students scoring 0 points. These activities refer to sophisticated regulation processes such as reflection which are probably still underdeveloped for students in this early phase of development (Veenman, Van Hout-Wolters, & Afflerbach, 2006). Leaving these items out leads to a maximum score of 16 points for the total instrument. Using the systematical observation scheme for the first twenty think-aloud protocols, a substantial interrater-reliability was found among the judges (κ = .95, p = .00).

VisA instrument. As discussed in the theoretical framework, prediction judgments, postdiction judgments and problem visualizations were combined into one instrument. This instrument assesses a combination of metacognitive monitoring and regulation which are interrelatedly used during problem solving. We call this newly developed instrument the VisA instrument (Visualization and Accuracy). In the VisA instrument, four word problems are presented. For each word problem, students are asked to divide their problem solving over various steps:

- Read the problem and rate your confidence for finding the correct answer (without calculating the answer);
- Make a sketch which can help you solve the problem;
- Solve the problem and fill in the answer;
- Rate your confidence for having found the correct answer;

Four multistep word problems appropriate for using schematic visualizations were selected for the instrument. Students got approximately a maximum of five minutes to solve each problem. The four steps of each word problem are folded in the form of a booklet starting with step 1 as the front-page, step 2 and 3 on the middle two pages, and step 4 on the last page.

Figure 5.1 shows the first part of the instrument. Students are asked to fill in a traffic light with three options: Red (I am sure I cannot solve this problem), orange (I am not sure whether I will solve this problem correctly or incorrectly) and green (I am sure I will solve this problem correctly) and comment on the rationale for their answer. The latter is meant to have students think carefully and ask themselves why they think they can or cannot perform the task. Figure 5.1 also shows the second step of the instrument: Problem visualization. This step is presented on the inside of the booklet and was used to assess the quality of students' problem visualizations.



Problem

Marja plants rosebushes alongside the path to her house.
The path is 27 meters long
She plants a rosebush every 3 meters on both sides of the path.
She also plants rosebushes at the beginning of the path.

How many rosebushes does Marja need?

Question

How well do you think you can solve this problem?

Please, explain why ....

Problem

Marja plants rosebushes alongside the path to her house.
The path is 27 meters long
She plants a rosebush every 3 meters on both sides of the path.
She also plants rosebushes at the beginning of the path.

How many rosebushes does Marja need?

Question

Draw a sketch you can use to solve the problem.

Figure 5.1 Step 1 and 2 of the VisA instrument: Predicting one's performance and visualizing the problem situation.

The scoring procedure for the instrument is designed to be straightforward so it is usable in research and practice. The scoring rules for each step are:

- If students' prediction judgments are correct (i.e. students predicted they could solve the problem correct and indeed did; or they predicted they could not solve the problem and indeed gave the wrong answer) students get 1 point. If students' predictions are uncertain (orange traffic light) or incorrect (i.e. they predicted they could solve the problem correctly but in fact give the wrong answer; or they predicted they could not solve the problem but solved the problem correctly) they score 0 points.
- For the visualization of the problem, students get 0 points if they made a pictorial sketch not depicting any of the important relationships in the problem, 0.5 point is awarded to sketches which are partly pictorial but have some schematic or mathematical features, and 1 point is given to primarily schematic visualizations.
- The postdiction judgments of the students are scored in the same manner as step 1. Thus, students get 1 point when the postdiction is correct and 0 points when the postdiction does not match the answer.

After scoring all four word problems, a sum score was computed for the total instrument. The maximum score is 12 points. The first ten visualizations were scored with two judges arguing until agreement about scoring rules for the visualizations was reached. Internal consistency of the instrument was $\alpha = .70$.

Self-report questionnaire. In this study, the 'metacognitive self-regulation' subscale of the MSLQ (Pintrich & De Groot, 1990) is used. Statements in this subscale best match the metacognitive processes in the other instruments. This subscale contains 12 items in the form of statements about metacognitive behavior such as "Before I study new [mathematics] material thoroughly, I often read it through quickly to see how it is organized" And "When I execute [a math assignment], I set goals for myself in order to direct my activities." General wording such as 'in this course' in the items were replaced by words specifically referring to mathematics.

Students were asked to indicate how much a statement applies to them by checking one out of five boxes ranging from 'not at all true for me' to 'completely true for me'. Scores were coded ranging from no metacognitive regulation (not at all true for me: score 0) to a high amount of self-reported metacognitive regulation (completely true for me: score 4). Some items were stated in a reversed manner in the instrument but were recoded for the analyses. The maximum score on the instrument was 48 points. The internal consistency of the instrument was α = .75.

Mathematical word problem test. As a performance measure, a test of 15 word problems was used. Of the test, two items with negative item-rest correlations were left out of the analyses. A sum score was calculated for the remaining 13 word problems. The test items are multistep word problems based on a national math assessment test (Janssen & Engelen, 2002). Most students were familiar with the computations required to solve the problems. But, the fact that the computations are embedded in text turns them into word problems in which a metacognitive approach can benefit the solution process. Two examples of word problems from the test are presented below.

Hassan already has € 250 in his savings account. *He is saving up for a game computer of € 490.* *He saves € 40 each month.* *In how many months can Hassan buy the game computer?*	*The pet store has a container with 5000 grams of dog food.* *Bart takes 30% out for his dog.* *How many grams of dog food stay in the container?*

Students got 1 point for each correct answer and 0 points for each incorrect answer. On average students in the sample solved 58 percent of the word problems (*SD* = 20). The test had a reliability of α = .65.

5.3.3 Procedure

The word problem test and the self-report questionnaire were collected in the classroom with students filling in all questions individually. Subsequently, data were collected for the think-aloud measure and the VisA instrument. Half of the students completed the think-aloud measurement before the VisA measurement and the other half of the students completed the VisA before the think-aloud measure. Think-aloud protocols were collected individually in a quiet room outside of the classroom. Students completed the VisA measurement in a group setting.

Student responses that were missing after collecting the instruments (varying from 0.4 to 10.9 percent of the responses MCAR) were completed using the Expectation-Maximization Algorithm (Roth, 1994; Schafer & Olsen, 1998) in SPSS.

5.4 Results

5.4.1 Convergence between the instruments

In order to assess the convergence between the three measures aimed at measuring students' metacognition, means and bivariate correlations are presented in Table 5.2.

Table 5.2

Means and bivariate correlations between the different instruments measuring metacognition.

	M (SD)	PS	TA	VisA
TA	9.87 (3.40)	.57**	-	
VisA	4.15 (1.96)	.48**	.29*	-
SQ	25.52 (7.07)	.03	.16	-.20

Note. PS = word problem solving test; TA = Think-aloud measure; VisA = VisA measure; SQ = Student Questionnaire

* $p < .05$ ** <.01

Students in our sample scored relatively low on all metacognitive measures, showing that metacognition is still in an early stage of development in upper elementary school. Concerning the relation with word problem solving, both on-line instruments – the think-aloud and the VisA instrument - were well related to performance with correlations ranging from .57 to .48. This is not the case for the self-report questionnaire which was not related to the mathematics test. Moreover, the self-report questionnaire showed no convergence with on-line metacognitive measures. Scores on the TA measure and the VisA instrument on the other hand were related, although the bivariate correlation is modest. Excluding one outlier with the highest TA score but a low VisA score would have led to a correlation between the two of $r(37) = .35$ and a correlation of VisA and PS of $r(37) = .50$ confirming that in general there is a moderate correlation between the two on-line instruments and that are strongly related to problem solving performance.

5.4.2 Unique and shared predictive validity of think-aloud and VisA

To assess the amount of unique and shared explained variance of the think-aloud measure (TA) and the VisA instrument as predictors of scores on the word problem solving test, a regression commonality analysis was performed. Commonality analysis partitions a regression effect into unique and common effects. Unique effects show the amount of variance uniquely explained by a certain predictor variable. And common effects show how much explained variance two (or more) variables have in common (Nimon & Reio, 2011). Results of the commonality analysis of the think-aloud measure and the VisA measure as predictors of problem solving performance are added in Table 5.3.

Table 5.3

Regression commonality analysis of a think-aloud measure and the VisA measure as predictors of word problem solving performance.

Predictor	*R*	*R²*	*β*	*p*	*Unique*	*Common*	*Total*
Both measures	.66	.43					
TA			.48	.00	.21	.12	.33
VISA			.34	.01	.11	.12	.23

Note. TA = Think-aloud measure; VisA = VisA measure; *Unique* = Unique variance explained by the predictor variable; *Common* = Shared variance explained by both predictor variables; *Total* = Total variance explained by the predictor variable

Table 5.3 shows in the first two columns that together the TA measure and the VisA measure correlated highly with problem solving performance (*r*(37) = .66) and the variance explained by both measures was considerable (43%). The data in columns three and four signify that TA and VisA have their own unique predictive value for performance. The beta coefficients indicate that 1 standard deviation change in TA score respectively VisA will lead to a .48 respectively .34 standard deviation change in students' word problem solving. From the results in column five and six we can derive that both TA and VisA explained some unique variance in the word problem solving test (11 and 21 % respectively), and besides this they communally explained 12 percent of the variance in word problem solving. In total, TA explained 33 percent, and VisA 23 percent of the variance in word problem solving performance.

5.5 Conclusion and Discussion

This study is intended as an exploration towards a more practical way of measuring metacognition to approximate the rich information of think-aloud protocols. Although imperative steps are being made towards new in-depth measures of metacognition supplementary to think-aloud protocol analysis (i.e. trace data; Azevedo et al., 2010; Greene & Azevedo, 2010; Winne, 2010), researchers and practitioners interested in students' metacognition still lack a practical instrument which is less complicated and time-consuming to use than think-aloud protocols or other in-depth measures (McNamara, 2011). We suggest that one of the ways to make a step forwards in this issue, is evaluating findings on various instruments theoretically aimed at measuring metacognitive monitoring and regulation, and comparing their predictive and convergent validity (c.f. Veenman, 2011a). Due to the fairly domain-specific nature of metacognition, we suggest the development of measurement instruments to be specifically molded to fit certain domains.

In this study, metacognition is measured in the domain of mathematical word problem solving. Findings from different measurement instruments were triangulated in an empirical study in grade five. Think-aloud observation was used as a comprehensive measure of metacognitive monitoring and regulation and as a reference point for other metacognitive measures. Think-alouds may not be appropriate for measuring automated processes (McNamara, 2011; Veenman, 2011b). But when collecting the protocols in an appropriate manner (Ericsson & Simon, 1993), with tasks of a suitable level of complexity (Prins et al., 2006), think-alouds provide rich information on consciously used metacognitive processes. In our study, the think-aloud measure explains a total of 33 percent of the variance in mathematics performance, which is comparable to the predictive validity reported in other studies with think-aloud measures (Veenman, et al., 2006). However, in the introduction we pinpointed the issue that collecting and analyzing think-aloud protocols is a very complex and time-consuming process. We reviewed several possible alternative measures which can be used in the design of a more practical, yet valid, measurement instrument. The empirical findings of our study using different measurement instruments are discussed below.

Firstly, based on the literature, it was hypothesized that a general off-line measures collected disconnected from the learning task would show little to no convergence with on-line measures. Findings of this study indeed support this claim. The self-report questionnaire we used shows no convergence with either the think-aloud measure nor the newly developed on-line instrument. Moreover, the questionnaire shows no relation to the problem solving test. This confirms the idea that what students say they do when asking them general self-report questions is not necessarily the same as what they actually do (Veenman, 2011b). As discussed in the theoretical framework, this problem is likely to be caused by memory distortions as well as by variation in interpretation of the questions (McNamara, 2011; Veenman, 2011b). It could be that such issues can be addressed by fitting the formulation of the items more closely to the learning task (Schellings, 2011). However, in our study we found that fitting the formulation of the questions to the learning domain (in this case by adding the word mathematics) does not seem to make the statements specific enough. Until we have more knowledge about how to increase concurrent and predictive validity of self-report questionnaires, we argue them to be more suitable as measures of metacognitive knowledge instead of on-line metacognitive metacognition which would be expected to directly influence performance (c.f. Desoete, 2007; Greene & Azevedo, 2010; Veenman, 2005).

Secondly, we suggested to combine prediction judgments, problem visualizations, and postdiction judgments into a new instrument; the VisA instrument. All of these measures were argued to be indicative of metacognition, as well as having predictive validity for students' word problem solving performance. A large practical benefit of the VisA instrument is that it can be collected in paper and pencil format with groups of students. Teachers or test leaders need to make sure that students fill in every part of the instrument and do not inattentively skip parts. Another practical benefit of the instrument is that the scoring rules are quite straightforward to understand and use. How does the new instrument concur with scores collected with a think-aloud measure? Correlations between the new VisA instrument and think-aloud scores were predicted to be moderate since they both measure on-line metacognition, but the VisA instrument does not cover the whole range of metacognitive activities of the problem solving process captured in the think-alouds. Indeed, a moderate but significant relationship between the two on-line measures was found. The amount of metacognitive activities found with both on-line instruments is relatively low in our elementary sample. But, both instruments are significantly interrelated and are related to word problem solving performance.

Partialing out both instruments unique contribution as predictors of students' word problem solving in a regression commonality analysis, shows a substantial amount of shared predictive variance between the think-aloud measure and the VisA instrument. The overlap between the two instruments accounts for about thirty percent of the total variance explained by both measures as predictors of word problem solving performance. This shows that in combining judgments of performance and problem visualizations, we have made a reasonable step forwards towards finding a valid and efficient instrument which corresponds to the think-aloud measure. Moreover, both instruments have predictive validity for word problem solving performance. As predicted, the think-aloud measure has the greatest predictive validity explaining 33 percent of the variance in problem solving performance. But, the VisA instrument also explains a sound part of 23 percent of the variance in word problem solving. The predictive validity of the VisA instrument for performance is comparable to the correlations reported previous studies of prediction and postdiction judgments (Chen, 2002; Desoete et al., 2001; Desoete, 2009; Vermeer et al., 2000) and problem visualizations (Edens & Potter, 2007; Van Garderen & Montague, 2003). The fact that the VisA instrument also uniquely covers some variance in problem solving which is not covered by the think–aloud measure may be due to the fact that it measures metacognitive monitoring more strongly than the think-aloud measure in which monitoring is only represented in two of the sub-items (see §3.3). Also, the activities of drawing a sketch and making a prediction are not one-to-one related to the activities which students performed in the think aloud measure.

Concluding, our data confirms that think-aloud data gathered on-line of the problem solving process can provide much information about metacognitive processes affecting word problem solving. In searching for a more practical instrument, we found that the VisA instrument shows potential as an instrument for measuring metacognition in mathematical problem solving. The instrument has several benefits which facilitate data collection and scoring. Our empirical study has shown the VisA instrument to have predictive validity for mathematical word problem solving in elementary education. Additionally, the convergence with the think-aloud measure indicates that the instruments partly measure comparable constructs. However, the fact that VisA only partially overlaps with the think-aloud measure is a drawback. Depending on the breadth of the metacognitive construct one aims to measure, there may be more work needed to complete the puzzle.

One of the possible extensions of the present study is to further assess the convergent and predictive validity of performance judgments and visualizations by collecting them separately. Although there is already evidence for the predictive validity of separate prediction judgments and problem visualizations for performance, little is known about the convergence between these measures and other on-line measures such as think-aloud protocols. In VisA, substeps of the instrument are presented as interdependent steps of the problem solving process and can thus not be reliably detangled. But, in a follow-up study, it might be interesting to collect and compare independent measures of performance judgments and problem visualizations. Secondly, the use of think-aloud methods could be strengthened by using factor analysis to determine adequate scoring categories for the specific age group. Moreover, it would be interesting to add additional measures in a follow-up study to control for other variables possibly influencing findings of the think-aloud measure (i.e. verbal abilities; Veenman, 2005) and the VisA instrument (i.e. spatial abilities; Cox, 1999) in order to get a clearer picture of the constructs which are measured. This way, the theory about similarities and dissimilarities of different measures can be expanded. This can facilitate the search for new applied measures.

Although this first exploration of more practical measurement of metacognition in elementary education provides us with ground for further exploration, certain limitations must be kept in mind. Firstly, the measures in our design consist of quite few word problems. It would be well-advised to lengthen the measurement instruments with more word problems to increase their reliability. For instance, to increase the internal consistency of the VisA instrument up to $\alpha = .80$, researchers might consider adding two or three comparable word tasks (Spearman, 1910). Another more general limitation of most on-line measurement instruments is their obtrusive nature which might bias students' responses in a certain direction (Schraw, 2010). The amount of bias caused by the different obtrusive measures is not clear and should be taken into account when interpreting findings from think-aloud protocols and the VisA instrument.

Irrespective of the fact that there are clearly still some hurdles to be taken, we hope this study on more practical measurement of metacognition in word problem solving provides to be an incentive towards the exploration of more efficient, yet valid, measurement instruments in metacognition research. We believe this not only to be a valuable issue for researchers, but also for the community of practitioners interested in stimulating students' metacognitive processes. Especially in schools where teachers have little time for testing individual students, it would be most relevant to have an efficient and valid instrument that shows which teachable metacognitive skills some students lack and others have already acquired. Regarding to the large progress which has already been made in metacognitive theory development in the past decades, making the transition towards more practical use of our knowledge is an imperative - and exiting - step to take.

CHAPTER 6

General Conclusion and Discussion

6.1 Introduction

This dissertation started with a quote by Socrates: “I cannot teach anybody anything; I can only make them think.” This quote draws the attention to students’ thinking processes. These processes are still particularly relevant for contemporary education. For students to become active learners, they need academic knowledge, but also they need to be able to monitor and regulate their own thinking. One of the domains in which a structured approach is especially important is word problem solving in mathematics. Word problems are frequently used in education to make students understand the relevance of mathematics and to help them to develop into flexible thinkers. However, many students have difficulties solving complex mathematical problems. They often do not regulate themselves to perform the cognitive actions needed to solve a word problem such as analyzing and exploring the problem, making and executing a solution plan, and verifying answers before they continuing. This regulation of cognitive processes is called metacognition in the literature. Several researchers have argued that in order to improve students’ word problem solving, it is important to include metacognition into instructional models (see chapter 1). In this dissertation it was studied how this can be done for students in elementary education.

To gain knowledge about how to use metacognition as a variable to improve performance, in the introduction it was first discussed which unresolved issues needed to be addressed in reference to the current status quo. First of all, the importance was noted of further defining the relation of metacognition and other variables influencing performance. It was suggested to determine the value of metacognition in addition to another key determinant influencing performance: Intelligence. For elementary school students, there are few empirical studies addressing the question whether metacognition could influence their performance in reading and mathematics apart from the influence of intellectual abilities. Empirically evaluating whether the relationship between metacognition and performance as suggested in theories of metacognition can be validated in an elementary school sample is an important first step towards studying if metacognition can be used to improve performance. Secondly, the major question addressed in this dissertation is whether students can be supported in their word problem solving by metacognitive training. There have already been many training studies using direct instruction to teach students a metacognitive approach. However, to learn students how to flexibly solve different types of word problems, guiding them with hints or prompts seems to be much more effective. The effectiveness of guided metacognitive training has already been studied in secondary education. It was shown that offering students metacognitive hints or questions that support their learning process can largely improve students’ metacognition and word problem solving performance (chapters 3 and 4). However, little was known about the effects of computer supported metacognitive training programs for word problem solving in elementary school. This issue is experimentally studied in two chapters of the dissertation. Lastly, a pressing issue is that practice as well as the research community lacks an easy applicable, yet reliable measure of metacognitive monitoring and regulation. In chapter five, various possible instruments were discussed and a newly developed instrument measuring metacognition in word problem solving was assessed for its psychometric qualities in an explorative study.

6.2 Main Findings

6.2.1 Relationship between metacognition, intelligence and performance.

In chapter 2, the relationship between intelligence, metacognition and academic performance was studied in an observational design to determine whether metacognition has unique predictive validity for performance. In the literature three hypothetical models of the relationship between these variables have been proposed (Minnaert & Janssen, 1999; Veenman & Spaans, 2005; Veenman et al., 2005). The first model, also referred to as the intelligence model, views metacognition as an integral part of intellectual ability. Metacognition according to this view cannot predict performance independent of intellectual abilities because the two are completely interwoven. In a second, contrasting model, the independency model, intelligence and metacognition are seen as completely independent predictors of learning. Finally, the mixed model suggests that metacognition and intellectual abilities are related to some extent, but that metacognition over and above this relationship predicts variance in learning outcomes. It is of relevance for the foundations of metacognition research to study if metacognition is a determinant of learning performance in its own right, or if it is completely interwoven with intelligence. As reported in chapter 2, most observational studies up to date have found results confirming the mixed model, meaning that metacognition has some overlap with intelligence but also explains some unique variance in academic performance (Veenman et al., 2006). However, few observational studies have empirically investigated the relationship between these variables in elementary education. Besides, to our knowledge, no studies report on the question whether metacognition behaves different for native and migrant students. It could be that natives and migrants apply their metacognition differently due to issues such as testwiseness, cultural loading of tests, or novelty (Freeman, 1993; Van de Vijver & Poortinga, 1997). These two issues were studied in an observational design in grade six. As an academic measure, a broad standardized measure was used mainly containing word problems and reading comprehension tasks. Intelligence was measured with both a fluid and a crystallized intelligence measure. And to measure metacognitive monitoring and regulation, think-aloud protocols collected in the context of word problem solving were used.

From the results of this study, it can be concluded that metacognition is significantly related to academic performance for both native and migrant students. The best predictor of learning performance in our sample is the measure of crystallized intelligence. But beyond that, metacognition affects performance of both native and migrants students. The different measures are structured and related in approximately the same way for native and migrant students, implying that students' cultural background is not an essential moderating variable affecting the relation between metacognition and performance. In general, one may conclude that metacognition is a significant determinant of performance for all students. Additionally, a moderate relation between metacognition and intelligence was found. This means our findings are in line with the mixed model in which metacognition and intelligence have some overlap, but metacognition also has its own importance as a predictor of performance. The fact that even non-trained elementary school students benefit from a metacognitive approach is promising. Based on this finding, combined with findings from the literature, there are strong reasons to pursue research focusing on how to use metacognition as an instructional tool in elementary education.

6.2.2 Training students in a metacognitive approach to problem solving

Having established the importance of metacognition for performance, the question arises if metacognitive processes can be trained in order to stimulate problem solving performance. One way to do so is training students' metacognition directly through step-by-step instruction on how to solve problems using a predetermined solution plan. This approach has been extensively studied. It is quite effective in training students to solve certain types of problems through examples and practice (Jacobse & Harskamp, 2011). But, a problem arises when students need to apply problem solving skills to different types of problems. Students often fail to transfer their problems solving skills to new situations. They lack the metacognition to guide them (Jonassen, 2003; Moreno, 2006).

In this dissertation, a different approach is followed. Students received guidance by providing them with metacognitive hints and questions about the problem solving process. It was expected that through choosing instruction and help, students would start to think about how to solve problems (Azevedo, 2005; 2007; Harskamp & Suhre, 2007). Guided metacognitive training programs mostly use some type of computer support. In the literature, there are already several successful examples of computer supported guided metacognitive training in secondary education, but not in elementary education. For instance, in secondary education, computer programs were developed guiding students problem solving with metacognitive questions such as "What is the problem all about?" and "If you had to make an argument to explain your answer, what reason would you give?" (Aleven & Koedinger, 2002; Kramarski & Gutman, 2006; Kramarski & Mizrachi, 2006). These training studies have reported effects on students' metacognitive monitoring and regulation as well as on word problem solving performance on near and far transfer tests. However, there are hardly any guided metacognitive training studies for word problem solving in elementary education. This is striking since national and international evaluations show that elementary school students are generally less than proficient in word problem solving (Janssen et al., 2005; Mullis et al., 2008). Therefore, a metacognitive training for elementary education is developed to test if the positive results found in secondary education could be replicated. The training environment is a computer program with word problems and metacognitive hints. In the hints, metacognitive questions and prompts are connected to the cognitive level of the problem and graphic representations are added to support the construction of a problem model. The computer training is based on success factors found in other metacognitive training programs (see chapters 3 and 4).

From the two experimental studies with the computer program with metacognitive hints, several main findings can be extracted. Firstly, it is found that, when leaving the choice of using the hints with the students, metacognitive hints are used well. As chapter 4 shows, the use of hints is related to students' word problem solving ability, which implies that students monitor when they need the hints. Secondly, as expected, metacognitive training with computer supported hints increases metacognitive behavior. Students practicing with the hints got better in regulating their cognitive strategies (chapter 3). And, in comparison to a control group practicing with word problems, students using metacognitive hints to guide their problem solving were better in monitoring their performance after six weeks of training (*d* = .67).

Evidently, our main objective was to find ways to improve students' word problem solving performance. In the first study presented in chapter 3, it was found that in comparison to a business as usual condition, practice with a computer program with metacognitive hints strongly improved students' word problem solving performance irrespective of their pre-test scores (*d* = .81). Moreover, in the experiment reported in chapter 4, an effect of metacognitive training was found in comparison to a control condition in which students practiced with word tasks in a computer program. Using the metacognitive hints for six weeks largely improved performance in the computer program (*d* = .82), and this effect also transfers to students' performance on an independent posttest (*d* = .54). Thus, providing students with student-controlled metacognitive support during word problem solving affects their performance beyond the effect of just getting more practice in solving word problems. No interactions were found between training effects and students' prior level of metacognition and problem solving. So, it can be concluded that guided metacognitive training with hints is effective for a broad range of students in upper elementary education.

6.2.3 Measuring Metacognition in Word Problem Solving

The last issue addressed in this dissertation is the measurement of metacognition. This topic has been prone to heated scientific debate in articles and conference presentations. Researchers have stressed that the field is still in need of valid and reliable measures to measure students' metacognition (Veenman, 2005; Winne, 2010). However, this is a complex issue. There are measurement instruments which are collected concurrent of the learning process, these are called on-line measures. And other measures are collected separately, off-line, of the learning process. Easy collectible off-line measures such as self-report questionnaires have often been criticized as being invalid measures of what students actually do (Greene & Azevedo, 2010; Veenman, 2005). But most on-line measures on the other hand are complex and time-consuming to collect (Azevedo et al., 2010). One well-known way to measure metacognition concurrent of the learning process is by using think-aloud protocols. In think-aloud protocols, students verbalize their thoughts while executing a task. This way, information can be gathered about students' metacognitive processes without them having to infer information or remember what they have done themselves. However, think-aloud protocols are not easy to collect, or to score. Although think-alouds seem to be valid indicators of metacognitive processes, this makes them quite unappealing for use on a large scale and use in practice. The aim of chapter 5 was to explore a more easily applicable way to measure on-line metacognition in word problem solving.

After summarizing findings on the validity of a number of measurement instruments aimed at measuring metacognition, we suggested combining several measures into a new on-line instrument. The newly combined instrument consists of two performance judgments and a problem visualization. The ability to judge one's performance is an expression of metacognitive monitoring and thinking ahead about an initial solution plan (Boekaerts & Rozendaal, 2010; Efklides, 2006). Research has shown that the accuracy of performance judgments before and after problem solving has good predictive validity for mathematics performance (Chen, 2002; Desoete et al., 2001; Desoete, 2009; Vermeer et al., 2000). The quality of students' visualizations of problem situations on the other hand is expected to be indicative for how well students analyze and explore a word problem towards a solution plan. Several studies have shown that students' sketches of word problems can be judged based on their schematic qualities, which has predictive validity for problem solving performance (Cox, 1999; Edens & Potter, 2007; Hegarty & Kozhevnikov, 1999; Van Garderen & Montague, 2003). By using a combination of these measures, the newly developed instrument measures the intertwined process of metacognitive monitoring and regulation in word problem solving. Although think-aloud protocols were expected to give the most information about metacognition over all episodes of problem solving, the new instrument covers various on-line metacognitive activities. Therefore we expected the new instrument to show reasonable overlap with think-aloud scores. Conversely, it was hypothesized that the off-line self-report questionnaire would show little convergence with the two on-line measures.

In line with our hypothesis, there was no relationship between the self-report measure on the one hand and the think-aloud measure respectively the new instrument on the other. Likewise, the self-report measure was not related to performance on a word problem solving test. This conforms the argumentation in the literature that self-report instruments for metacognition probably measure something else than measures collected during the learning process (Desoete, 2007; Greene & Azevedo, 2010). Between the two on-line instruments on the other hand, a moderate relationship was found. The newly developed instrument had 12 percent shared variance with think-alouds as a predictor of word problem solving. Besides, both instruments had good predictive validity for word problem solving performance. The think-aloud measure explained 33 percent, and the new measure explained 23 percent of the total variance in word problem solving. This shows that a reasonable step forwards towards has been made in finding a valid and efficient instrument which corresponds to the think-aloud measure. Researchers are urged to further explore practical measures to make assessment of metacognition more usable for the research community as well as for practitioners.

6.3 Discussion

6.3.1 Limitations

There are some limitations of our studies which should be kept in mind when interpreting the results. A first limitation has to do with the generalizability of our findings. The experimental studies in chapter 3 and 4 show that effects of using metacognitive hints in the computer program transfer to solving word problems on independent tests without guidance. However, the designs did not include follow-up measures. So, it is not known how long students retain effects of the training on metacognition and word problem solving performance. Additionally, the experiments were executed in classrooms with students of average ability and social economic status. Although within these classrooms students' prior knowledge did not affect effectiveness of the training, it is not clear how the metacognitive training program used in our studies works with subsamples of very low ability or conversely for high achieving students.

Another limitation of our intervention studies is that, although data about students' metacognitive and cognitive outcomes were collected, no fine-grained data were gathered about students' learning processes while working with the hints. In chapter 3 students were observed while working with the computer program and it seemed that they used the hints well. However, more in-depth process data is needed to evaluate how students use the information from the hints and fit this into their thinking about the solution of problems. This way the interaction of metacognition and cognition affecting performance could be further explicated. The lack of such fine-grained information on students' learning processes is a shortcoming which should be taken into account in future studies.

6.3.2 Recommendations for further research and improvement of practice

Defining and measuring metacognition. As noted in the introduction, metacognition research is often labeled as being rather 'fuzzy' or vague (Dinsmore et al., 2008; Hacker, 1998; Schunk, 2008). Although much progress has been made in developing theories about metacognition, there are still some unresolved issues which add to this problem. One of the most pressing issues has to do with empirically validating the theory behind the concept. In theoretical models, metacognition is defined as a meta-level process affecting performance (Flavell, 1979; Nelson, 1996; Veenman et al., 2006). To evaluate the theory, researchers have executed correlational studies to evaluate the relationship between metacognition and performance (i.e. Van der Stel & Veenman, 2008). However, there are few studies which have experimentally studied this assumed causal relationship. For instance, as the review study of Dinsmore, Alexander and Loughlin (2008) makes clear, in articles on metacognition over the past couple of years, only five percent of the articles explicitly address behavioral components in their definitions. And half of the studies on metacognition did not at all study any causal effects of prompting or scaffolding metacognitive processes. For the field to further develop and to become less fuzzy and more concrete, it is time to move forward and study how the theory can actually be applied in practice.

One important finding from this dissertation is that when metacognition and cognition are used in interaction to improve the way students learn this can largely influence their performance. In accordance with this finding, I would recommend researchers studying metacognition in a particular domain to not merely assess metacognitive processes, but to measure these processes in a specific learning context and to relate them to learning outcomes. The VisA instrument described in chapter 5 might be a good example of such an applied measure. Contradictory to general metacognitive measures such as most self-report questionnaires, in the VisA instrument metacognitive actions are recorded during the problem solving process. Judgments of performance and problem visualizations were used as on-line indicators of metacognitive monitoring and regulation. As expected, the instrument has good predictive validity for problem solving performance. In line with the findings, I strongly encourage researchers to move on along this path and try to further concretize the concept of metacognition by developing applied measurement methods.

Metacognitive training. The theoretical choices one makes, also influence the way interventions are designed. In this dissertation the studies were structured around the theory that metacognition is explicitly linked to the use of cognitive strategies influencing performance (see above). In line with this assumption, metacognitive processes were trained embedded in cognitive content. However, some other studies have trained metacognition as an isolated variable using fact-sheets or checklists with metacognitive activities which are not directly connected to a learning task (i.e. Schraw, 1998). To further evaluate which type of metacognitive guidance is best to improve performance in different situations, one needs to know which type of metacognitive support works best when it is situated in specific tasks and which metacognitive processes may be more general across tasks. Overall, there is no consensus about the question if metacognition is typically domain-specific or if it also has some general features (Veenman et al., 2006). I would recommend further investigating this issue both within the domain of mathematics as across other domains.

A first suggestion in this respect is to study if metacognition has as much predictive validity for other mathematical topics as it has for word problem solving. Most studies on metacognition in mathematics focus on problem solving. However, there are other mathematical domains in which students seem to need more guidance. For instance, in the Netherlands, eighty to almost ninety percent of the students were found to perform under the average benchmark for complex calculations and geometry (Van der Schoot, 2008). For such complex subdomains other than problem solving, teaching students to regulate how they solve a geometry task or a complex calculation may also influence performance. To our knowledge, there are no studies showing how metacognition can benefit performance in geometry, except a study on help-seeking in a geometry tutor environment (Roll et al., 2007). For the subdomain of calculations, there are some intervention studies which show that determining the type of calculation and choosing an appropriate calculation plan can improve performance on computation measures and standardized mathematics tests (Fuchs et al., 2009; Jitendra et al., 2007; Tournaki, 2003). However, these studies were performed using direct metacognitive training and were undertaken only with low ability students. To gain insight in which problem solving strategies are transferable to mathematics in a broader sense, more research is needed using other math topics.

A second recommendation to gain insight in the domain-specificity versus the generality of metacognition, is designing interventions which assess effects of metacognitive training within- and across domains. There have already been some intervention studies showing that metacognition can improve performance in other domains than mathematics such as comprehensive reading, writing and science (Baker, 2002; Harris, Graham, Brindle, & Sandmel, 2009; Peters & Kitsantas, 2010). However, there are few studies experimentally studying if metacognitive training in one domain can improve performance in another domain, or if simultaneously training metacognition in different domains is effective. Only the findings of an experiment by Kramaski, Mevarech and Lieberman (2001) suggest that training metacognition in multiple domains at once may increase effectiveness of the training. It would be interesting to further study this topic. A first step to do so could be to train students in one domain and to test if the effects crossover to metacognitive behavior and performance in other related domains. Secondly, following the suggestion by Kramarski and colleagues, researchers could simultaneously train students' metacognition in different domains. It would be interesting to study if metacognitive training in different domains which have some commonalities (such as word problem solving and comprehensive reading, or comprehensive reading and other domains in which reading texts is important), can increase the overall effectiveness of metacognitive interventions.

Metacognition for different groups of students. A next recommendation would be to further study how metacognition influences performance for different subsamples of students. Our studies show that the effects of guided metacognitive training are not different for lower ability and higher ability students (chapters 3 and 4). But, students with low pretest scores used the metacognitive hints more than students with high problem solving scores on the pretest. Using more hints that contain cognitive and metacognitive support helped them improve their metacognition and problem solving performance just as much as other students (chapter 4). Although the low achieving students in our samples generally profited from the training, we did not yet study how guided metacognitive training works for specific subgroups of students with learning difficulties or contrariwise for high achieving students.

When aiming to train students with learning difficulties, a first requirement for the development of metacognition would be that students have enough mathematical knowledge to understand the tasks. Since metacognition is about regulating cognition, students should at least have basic mathematics skills. To develop this knowledge and to connect it to metacognitive regulation and monitoring, some direct instruction may be needed (c.f. Chang et al., 2006; Teong, 2003). Then, when moving on to a guided instructional phase, students with learning difficulties may need considerable individual support. In this sense, working with a computer program can be very helpful because students can work in their own pace and use hints to fit their own needs. Researchers can select appropriate problems or even use a computer script which automatically adapts tasks to students' level of performance. Working on their own level and individually receiving support can help low achievers stay motivated (Yeh, 2010). Also, receiving metacognitive support may strengthen low achievers' confidence for solving problems (Caprara et al., 2008; Kleitman & Stankov, 2007). But, in order to effectively fit metacognitive hints to the needs of low achievers, more in depth information is needed about how students use the hints. To gain insight in this issue, I suggest collecting fine-grained measures of students' learning processes. One way to do so is by recording students' verbalizations of thoughts while they are working with the computer program (Moos & Azevedo, 2008). An alternative could be to collect so called trace-data. Traces are observable sequences of events which can make causal relations between computer support and the actions by the student following this support more clear (Winne, 2010). In the computer program developed for our studies, adding monitoring question-popups (such as: How well do you think you can solve the problem?) or a digital sketchpad recording the notes and sketches students make to the interface could help to unravel more information. For example: The student answers in the popup that he/she is not able to solve the problem without help, and consequently clicks the metacognitive hints. Or, the student clicks a planning hint, writes down and completes a mathematical model on the digital sketchpad, and then fills in the correct answer. Such traces of events, eventually supplemented with think-aloud data, could be used to clarify how students use the information from the hints towards finding an answer. Additionally, when designing metacognitive training for students with learning difficulties, we recommend taking into account students' limited working memory capacity. Research on cognitive load has shown that especially for novice students complex tasks like solving a mathematical problem can load heavily on their working memory (Sweller, 1988; Van Gog, Paas, & van Merrienboer, 2004). When providing students with instructional support in the form of metacognitive hints, researchers should keep an eye on whether students are able to process the information from the hints and to apply this to solve the problem. In line with this recommendation, I would suggest studying whether incorporating a gradual shift from computer-controlled to student-controlled hints as students develop their problem solving abilities (or abilities in another domain) is more effective for low achievers than immediately giving them full control.

For high achievers on the other hand, it would be interesting to study if they efficiently monitor their use of hints according to their level of proficiency. When students monitor their hint-use effectively, after some practice they will need fewer hints, at least if tasks do not increase in difficulty (compare findings in chapter 3). Fading hints when students become more proficient can positively affect their performance (i.e. Atkinson, Renkl, & Merrill, 2003). This causes students to shift from relying on guidance by a teacher or a computer tool to independently structuring their learning. Such a shift towards more independent agency of the learning process may be particularly feasible for high achievers. It would be interesting to study how high achieving students can be supported and motivated to take on this responsibility.

Metacognitive training and the teacher. A final recommendation is to include teachers in further studies on metacognitive training in order to connect findings from research to daily practice in the classrooms. The studies in this dissertation have shown that when researchers supervise if the computer program is implemented well, metacognitive training can largely improve elementary school students' word problem solving performance. However, it was not yet studied how teachers can use the program. Studies leaving more responsibility to the teacher will probably take some time and effort. Teachers were generally found to struggle with constructivist forms of teaching (Davis & Sumara, 2002; Kirschner et al., 2006). More specifically, when it comes to instructing metacognitive behavior, teachers are typically relatively unsuccessful and non-explicit in their instruction (Dignath & Buttner, 2008; Kistner et al., 2010). To support teachers to incorporate metacognitive instruction in their daily routines, I would recommend designing interventions consisting of a few careful steps: A first step is supporting teachers to think more metacognitively themselves (Kramarski & Revach, 2009; Wilson & Bai, 2010). Then, principles of effective teacher training should be applied to instruct them about how to use metacognitive instruction in their lessons. Teacher training on using metacognitive instruction can be designed in several ways, for instance by using workshops or personal coaching in the classroom. Lastly, supporting and observing teachers who integrate metacognitive training in their lessons can provide information about how to effectively transfer experimental metacognitive interventions into the classroom.

SAMENVATTING (Dutch summary)

In het huidige onderwijs delen veel partijen de opvatting dat leerlingen zouden moeten leren om actief kennis te construeren en hun eigen leerproces te sturen in plaats van slechts passief kennis te vergaren. Deze constructivistische opvatting vraagt structurele begeleiding van de leerkracht en een actieve rol van de leerling. Veel leerkrachten proberen leerlingen echter kennis te laten construeren door ze volledig zelfstandig te laten werken (Davis & Sumara, 2002). Onderzoek laat zien dat dit ineffectief is (Kirschner, Sweller & Clark, 2006; Mayer, 2004). Om leerlingen actief te laten leren, zullen leerkrachten leerlingen basiskennis moeten aanreiken en ze tevens moeten begeleiden in het reguleren en monitoren van hun eigen denkprocessen. Er zijn al veel onderzoeken waaruit blijkt dat leerlingen die goed in staat zijn hun leren te reguleren, betere prestaties behalen (Dignath & Buttner, 2008; Swanson, 2001; Wang, Haertel, & Walberg, 1990). Dit laat zien dat een combinatie van cognitieve, en proces-georiënteerde instructie erg belangrijk is. In de praktijk wordt er echter weinig aandacht besteed aan de procedurele kant van het leren (Inspectie van het onderwijs [Inspectorate of Education], 2008; Kistner et al., 2010). Er is meer onderzoek nodig naar hoe leerlingen effectief kunnen worden begeleid om hun leerprocessen aan te sturen.

Eén van de subdomeinen in het onderwijs waarbij procesregulatie erg belangrijk is, is het oplossen van toepassingsopgaven bij rekenen. Toepassingsopgaven zijn opgaven waarbij de rekenkundige inhoud is verwerkt in tekst. Dergelijke opgaven worden veelvuldig gebruikt in methodes en toetsen om het rekenen in een min of meer realistische context aan te bieden. Echter, uit verschillende nationale en internationale peilingen blijkt dat veel leerlingen in het basisonderwijs moeite hebben met het oplossen van toepassingsopgaven (Janssen, Van der Schoot, & Hemker, 2005; Mullis, Martin & Foy, 2008). Omdat toepassingsopgaven anders in elkaar zitten dan ‘kale sommen’, vereisen ze specifieke vaardigheden. Daarbij rijzen de vragen: Wat zijn dit voor vaardigheden en hoe kun je vaststellen of leerlingen deze ontwikkeld hebben? En kunnen er instructietechnieken worden ontwikkeld om vaardigheid in probleemoplossen te stimuleren?

Uit de literatuur is al het een en ander bekend over de vaardigheden die leerlingen nodig hebben om toepassingsopgaven succesvol op te lossen. Ten eerste moeten leerlingen de rekenkennis hebben die nodig is om de berekeningen uit te voeren. Maar dat is niet de crux. Een specifieke moeilijkheid bij toepassingsopgaven is dat leerlingen een mentaal model moeten construeren door informatie uit te tekst te extraheren (Fuchs et al., 2008; Kintsch & Greeno, 1985). Om dit te kunnen doen, moeten zij cognitieve strategieën gebruiken. Uit onderzoek blijkt dat succesvolle probleemoplossers hun leren structureren over verschillende episodes van het oplossingsproces, namelijk: Goed lezen en analyseren van de opgave, exploreren van het type opgave, maken en uitvoeren van een rekenplan, en verifiëren van het antwoord. Zij reguleren het gebruik van de verschillende episodes en monitoren of tussentijdse aanpassing van hun leergedrag nodig is. Aan de andere kant gaan zwakke probleemoplossers vaak al berekeningen maken voor ze zorgvuldig hebben geanalyseerd wat de bedoeling van de opgave is (Schoenfeld, 1992; Verschaffel et al., 1999a). Hieruit kan worden geconcludeerd dat het verschil tussen leerlingen die succesvol versus onsuccesvol zijn in het oplossen van toepassingsopgaven grotendeels wordt bepaald door het reguleren en monitoren van cognitieve strategieën. Dit wordt in de literatuur metacognitie genoemd.

Metacognitie gaat over het aansturen van denkprocessen op meta-niveau (Flavell, 1979; Nelson, 1996). Dit kan een leerling doen door metacognitieve regulatie en door de eigen leerprocessen te monitoren. Metacognitieve regulatie refereert naar het aansturen van de cognitieve strategieën die nodig zijn on een taak op te lossen (Brown & DeLoache, 1978; Efklides, 2006). En monitoren gaat over de voortdurende controle over leerprocessen die wordt gebruikt om waar nodig de eigen werkwijze bij te sturen (Desoete, 2008). Dus als een leerling er bijvoorbeeld voor kiest om een berekening op te schrijven is dat een voorbeeld van metacognitieve regulatie, maar het schrijven zelf is cognitief. Als de leerling vervolgens een fout in zijn oplossingsstrategie opmerkt en aanpast, monitort hij/zij het leerproces en reguleert het maken van een nieuw rekenplan. Zoals dit voorbeeld laat zien zijn metacognitieve en cognitieve processen sterk met elkaar verweven. Het één kan niet los gezien worden van het ander.

Hoewel veel onderzoekers het er over eens zijn dat metacognitie erg belangrijk is voor het oplossen van toepassingsopgaven zijn er nog enkele hiaten in de huidige kennisbasis. Zo zou men zich allereerst kunnen afvragen of metacognitie wel voorspellende waarde heeft voor de prestaties van basisschoolleerlingen. Daarbij is het belangrijk onderscheid te maken tussen metacognitie en een andere voorspeller van leren: Intelligentie. Er zijn voor adolescenten al enkele onderzoeken die laten zien dat metacognitie voorspellende waarde heeft voor academische prestaties naast intelligentie. Maar er is nog weinig bekend over het effect van metacognitie op prestaties in taal en rekenen voor basisschoolleerlingen. Voor zowel theorievorming als voor de ontwikkeling van interventiestudies is het van belang om te onderzoeken of de veronderstelling dat metacognitie een unieke determinant is van prestaties ook opgaat voor relatief jonge leerlingen. Tevens is nog niet onderzocht of metacognitie en prestaties zich op dezelfde manier verhouden voor allochtone en autochtone leerlingen.

Ten tweede volgt de vraag hoe leerlingen kunnen worden gestimuleerd om metacognitie te gebruiken bij het oplossen van toepassingsopgaven. Een manier die uitermate geschikt is om leerlingen hierover te instrueren is door middel van begeleide training. Een begeleidende manier van trainen wordt gekarakteriseerd door een combinatie van een actieve rol van de leerling en begeleiding van de leerkracht of door een computerprogramma (Mayer, 2004). Meestal wordt de begeleiding aangeboden in de vorm van leerling-gestuurde vragen of hints in een computeromgeving (Azevedo, 2005; 2007). Begeleide training waarbij de leerling keuzevrijheid heeft, heeft een aantal voordelen ten opzichte van training waarbij een docent vaste oplossingswijzen aanleert. Zo leren leerlingen zelf te bepalen hoe ze de opgaven oplossen en wanneer ze hulp nodig hebben. Bovendien zorgt het feit dat leerlingen leren om actief verschillende procedures toe te passen voor transfer van het geleerde naar andere typen toepassingsopgaven (Jonassen, 2003; Moreno, 2006; 2009). Onderzoeken op de middelbare school hebben al getoond dat het mogelijk is om leerlingen aan de hand van computergestuurde metacognitieve hints en vragen te helpen hun metacognitie en prestaties in het oplossen van toepassingsopgaven te verbeteren (i.e. Aleven & Koedinger, 2002; Harskamp & Suhre, 2007; Kramarski & Gutman, 2006). Maar, het was tot op heden nog onduidelijk of een dergelijke aanpak ook kon werken voor leerlingen in de basisschoolleeftijd.

Een laatste hiaat in de literatuur is de onduidelijkheid over hoe metacognitie efficiënt kan worden gemeten (Schellings & Van Hout-Wolters, 2011). Er zijn enkele meetmethodes die over het algemeen worden gezien als valide manieren om metacognitie te meten, waaronder het gebruik van hardop-denk protocollen waarbij leerlingen hun gedachten verbaliseren. Maar het gebruik van dergelijke meetmethodes is gecompliceerd en tijdrovend. Meetinstrumenten die makkelijker af te nemen zijn, zoals zelfrapportagevragenlijsten voor het meten van metacognitie, worden echter vaak bekritiseerd als slechte voorspellers van het gedrag van leerlingen (Veenman, 2005). Het is belangrijk om te onderzoeken of er een meer efficiënte meetmethode kan worden gevonden die de voordelen van eenvoudige data-verzameling combineert met kenmerken van valide instrumenten.

Om de voorspellende waarde, de trainbaarheid en het meten van metacognitie verder te onderzoeken zijn er vier onderzoeken uitgevoerd. De eerste studie (hoofdstuk 2) is een observationeel onderzoek naar de relatie tussen metacognitie, intelligentie en cognitieve prestaties op een gestandaardiseerde prestatiemaat (de Cito eindtoets). In dit onderzoek is onderzocht of metacognitie voorspellende waarde heeft voor de prestaties van allochtone en autochtone leerlingen. De tweede en derde studie rapporteren over de ontwikkeling van een metacognitief trainingsprogramma voor basisschoolleerlingen. In hoofdstuk 3 wordt gerapporteerd over de effecten van metacognitieve training versus een controlegroep zonder training. En in hoofdstuk 4 wordt een experimentele studie beschreven naar het effect van metacognitieve hints naast het effect van oefening met toepassingsopgaven. Ten slotte gaat hoofdstuk 5 in op de vraag hoe metacognitie bij het oplossen van toepassingsopgaven praktischer kan worden gemeten.

Wat leren de onderzoeken uit deze dissertatie ons? Ten eerste kan op basis van de bevindingen van de observationele studie in hoofdstuk 2 worden geconcludeerd dat de theorie dat metacognitie unieke waarde heeft voor schoolprestaties kan worden gevalideerd in een steekproef met basisschoolleerlingen. Metacognitie en intelligentie zijn gerelateerd maar hebben ook unieke waarde voor prestaties op de cito eindtoets. Bovendien is metacognitie gerelateerd aan prestaties van zowel autochtone als allochtone leerlingen. We concludeerden dat het relevant is om te onderzoeken of deze relatie kan worden versterkt door middel van metacognitieve training. Hiervoor ontwikkelden we een computerprogramma met toepassingsopgaven en leerling-gestuurde metacognitieve hints voor verschillende cognitieve episodes van het probleemoplossingsproces zoals analyseren, plannen en verifiëren van het antwoord. Het programma is gebaseerd op kenmerken van effectieve trainingsprogramma's bij oudere leerlingen die uitgaan van begeleide training door middel van vragen, hints of hulpaanwijzingen (zie hoofdstuk 3 en 4). Metacognitieve training had bij adolescenten effect op zowel de metacognitie als op hun vaardigheid voor het oplossen van toepassingsopgaven. Maar aangezien er weinig bekend is over begeleide training van metacognitie bij basisschoolleerlingen zou men zich af kunnen vragen of het wel mogelijk is om metacognitive vaardigheden aan jonge leerlingen aan te leren zonder expliciete instructie door de leerkracht. Onze twee experimentele studies in groep 7 laten zien dat begeleide training van basisschoolleerlingen haalbaar en effectief is. Er werden effecten gevonden op maten van metacognitie en op rekenprestaties. Zo vonden we dat leerlingen na het oefenen met de metacognitieve hints hun probleemoplossen beter reguleren (hoofdstuk 3) en beter in staat zijn om te beoordelen hoe goed ze een opgave kunnen oplossen (hoofdstuk 4). Bovendien laat de studie in hoofdstuk 4 zien dat leerlingen niet alleen vooruitgaan doordat ze oefenen met toepassingsopgaven, maar dat het gebruik van de metacognitieve hints hun prestaties extra bevordert. Tot slot wordt in hoofdstuk 5 een onderzoek gepresenteerd naar verschillende manieren om metacognitie te meten. Dit onderzoek laat zien dat door het vergelijken van de predictieve validiteit en de overeenkomsten tussen verschillende instrumenten een stap voorwaarts kan worden gezet naar meer praktische meting van metacognitie. We ontwikkelden een instrument dat inschatten van eigen prestaties en het maken van een schets voor het oplossen van toepassingsopgaven combineert. De resultaten van deze eerste exploratieve studie wijzen uit dat het nieuwe instrument overlapt met hardop-denk protocollen en goede voorspellende waarde heeft voor de rekenprestaties van leerlingen.

Bij het interpreteren van deze resultaten moeten enkele beperkingen in acht worden genomen. Ten eerste laten de experimentele studies zien dat de training effect heeft op het oplossen van toepassingsopgaven op onafhankelijke toetsen. Echter, de designs hadden geen retentiemeting waardoor het niet mogelijk is uitspraken te doen over langetermijneffecten van metacognitieve training. Bovendien is er nog niet onderzocht of effecten kunnen worden gegeneraliseerd naar specifieke subgroepen zoals leerlingen met leerproblemen of hoog presterende leerlingen. Ten tweede is er geen fijnmazige data verzameld over hoe leerlingen de metacognitieve hints gebruikten om hun leren te monitoren en reguleren. Afgezien van informatie over het hintgebruik zoals verzameld in de hoofdstukken 3 en 4, hebben we geen procesgegevens over hoe de denkprocessen en gedragingen van leerlingen tijdens het maken van toepassingsopgaven werden beïnvloed door het gebruik van hints.

Voor verder onderzoek beveel ik aan om studies te richten op vier onderwerpen: Het verder definiëren en meer toegepast meten van metacognitie; Onderzoek naar generaliseerbaarheid van de effecten van metacognitieve training naar andere rekendomeinen en andere vakken; Onderzoek naar hoe metacognitieve training kan worden gebruikt voor het stimuleren van prestaties van leerlingen met leerproblemen of juist van hoog presterende leerlingen; En onderzoek naar hoe leerkrachten kunnen worden begeleid om metacognitieve instructie te gebruiken in hun dagelijkse lespraktijk. In hoofdstuk 6 worden enkele suggesties gedaan voor verder onderzoek naar deze onderwerpen waardoor de kennis over en toepasbaarheid van metacognitie als een belangrijke variabele in het onderwijs verder kan worden uitgediept.

REFERENCES

Aleven, V. A. W. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101-128.

Alexander, J. M., Carr, M., & Schwanenflugel, P. J. (1995). Development of metacognition in gifted children: Directions for future research. *Developmental Review, 15*, 1-37.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*, 774-783.

Azevedo, R., Greene, J. A., & Moos, D. C. (2007). The effect of a human agent's external regulation upon college students' hypermedia learning. *Metacognition and Learning,* 2, 67-87.

Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist, 40*, 193-197.

Azevedo, R. (2007). Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning, 2*, 57-65.

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*, 210-223.

Baker, L. (2002). Metacognition in comprehension instruction. what we've learned since NRP. In C. Collins Block, & S. R. Parris (Eds.), *Comprehension instruction: Research-based best practices* (pp. 65-79). New York, NY: Guilford Press.

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research, 35*, 359-375.

Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior, 25*, 829-835.

Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning, 3*, 39-58.

Blom, S. V., Hoek, D. J., & Ten Dam, G. T. M. (2007). Metacognitieve zelfregulatie, motivatie en perceptie van klassenklimaat. Zijn er sociaal-culturele verschillen? [Metacognitive self-regulation, motivation and perception of classroom culture. Are there socio-cultural differences?] *Pedagogische Studiën, 84*, 20-36.

Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning & Instruction, 20*, 372-382.

Borkowski, J. G., Carr, M., & Pressley, M. (1987). 'Spontaneous' strategy use: Perspectives from metacognitive theory. *Intelligence, 11*, 61-75.

Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. In R. S. Siegler, & R. S. Siegler (Eds.), *Children's thinking: What develops?* (pp. 3-35). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Burt, C. (1949). The structure of the mind; A review of the results of factor analysis. *British Journal of Educational Psychology, 19*, 176-199.

Campione, J. C., & Brown, A. L. (1978). Toward a theory of intelligence: Contributions from research with retarded children. *Intelligence, 2*, 279-304.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology, 100*, 525-534.

Centraal Bureau voor Statistiek [Central Statistics Agency]. (2003). *In 2020 één op de drie leerlingen allochtoon. [in 2020 one in three students are migrant].* Retrieved July 20, 2006, from http://www.cbs.nl/nl-nl/menu/themas/dossiers/allochtonen/publicaties/artikelen/archief/2003/2003-1164-wm.htm

Chang, K. E., Sung, Y. T., & Lin, S. F. (2006). Computer-assisted learning for mathematical problem solving. *Computers and Education, 46*, 140-151.

Chen, P. P. (2002). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*, 77-90.

Citogroep Primair Onderwijs (2007). *Terugblik en resultaten eindtoets basisonderwijs 2007. [Reflection and results primary school final test 2007].* Arnhem: Cito bv.

Clark, R. C., & Mayer, R. E. (2008). Who is in control? Guidelines for e-learning navigation. In R. C. Clark, & R. E. Mayer (Eds.), *E-learning and the science of instruction* (pp. 289-316). San Fransisco, CA: Pfeiffer.

Cognition and Technology Group at Vanderbilt. (1997). *The jasper project: Lessons in curriculum, instruction, assessment, and professional development*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (rev. ed.)*. Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Conway, A.R.A., Cowan, N., Bunting, M.F., Therriault, D.J., Minkoff, S.R.B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence 30*, 163–183.

Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction, 9*, 343-363.

Cromley, J. & Azevedo, R. (2011). Measuring Strategy Use in Context with Multiple-Choice Items. *Metacognition and Learning*, *6*, 155-177.

Dagevos, J., Gijsberts, M., & van Praag, C. (2003). *Rapportage minderheden 2003. Onderwijs, arbeid en sociaal-culturele integratie. [Minority report 2003. Education, work and socio-cultural integration.].* Den Haag: Sociaal en Cultureel Planbureau.

Davis, B., & Sumara, D. (2002). Constructivist discourses and the field of education: Problems and possibilities. *Educational Theory, 52*, 409-28.

Desoete, A., & Veenman, M. V. J. (2006). Metacognition in mathematics: Critical issues on nature, theory, assessment and treatment. In A. Desoete, & M. V. J. Veenman (Eds.), *Metacognition in mathematics education* (pp. 1-10). New York: Nova Science Publishers, Inc.

Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology, 5*, 705-730.

Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. *Metacognition and Learning, 3*, 189-206.

Desoete, A. (2009). Metacognitive prediction and evaluation skills and mathematical learning in third-grade students. *Educational Research and Evaluation, 15*, 435-446.

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities, 34*, 435.

Dignath, C., & Buttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*, 231-264.

Dinsmore, D., Alexander, P., & Loughlin, S. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*, 391-409.

Edens, K., & Potter, E. (2007). The relationship of drawing and mathematical problem solving: “draw for math” tasks. *Studies in Art Education: A Journal of Issues and Research in Art Education, 48*, 282-298.

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review, 1*, 3-14.

Elshout, J. J., Veenman, M. V. J., & Van Hell, J. G. (1993). Using the computer as a help tool during learning by doing. *Computers & Education, 21*, 115-122.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (rev. ed.)*. Cambridge, MA, US: The MIT Press.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. & Mascolo, J. (2002). *The Achievement Test Desk Reference (ATDR): Comprehensive Assessment and Learning Disabilities*. Boston, MA: Allyn & Bacon.

Flavell, J. H. (1976). The nature of intelligence. In L. B. Resnick (Ed.), *Metacognitive aspects of problem solving.* (pp. 231-235). Hillsdale N.J.: Erlbaum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. A new area of cognitive developmental inquiry. *American Psychologist, 34*, 906-911.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*, 316-344.

Freeman, J. (1993). Parents and families in nurturing giftedness and talent. In K. A. Heller, F. J. Mönks, A. H. Passow, K. A. Heller, F. J. Mönks & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent.* (pp. 669-683). Elmsford, NY US: Pergamon Press.

Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology, 100*, 30-47.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., Hamlett, C.L., Zumeta, R. O. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology, 101*, 561-576.

Fuchs, L. S., Zumeta, R. O., Schumacher, R. F., Powell, S. R., Seethaler, P. M., Hamlett, C. L., & Fuchs, D. (2010). The effects of schema-broadening instruction on second graders’ word-problem performance and their ability to represent word problems with algebraic equations: A randomized control study. *Elementary School Journal, 110*, 440-463.

Fuson, K. C. (2003). Developing mathematical power in whole number operations. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 68-94). Reston, VA: National Council of Teachers of Mathematics.

Graesser, A. C., Wiley, J., Goldman, S. R., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning, 2*, 89-105.

Gravemeijer, K. (1997). Mediating between concrete and abstract. In T. Nunes, P. Bryant, T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective.* (pp. 315-345). Hove England: Psychology Press/ Erlbaum (UK) Taylor & Francis.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*(4), 203-209.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice.* (pp. 1-23). Mahwah New Jersey: Lawrence Erlbaum Associates, Publishers.

Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and children's writing. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 131-153). New York, NY: Routledge.

Harskamp, E., & Suhre, C. (2007). Schoenveld's problem solving theory in a student controlled learning environment. *Computers & Education, 49*, 822-839.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*, 99-136.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*, 684-689.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87*, 18-32.

Helms-Lorenz, M., & Jacobse, A. E. (2008). Metacognitive skills of the gifted from a cross-cultural perspective. In F. Shaughnessy, M.V.J. Veenman & C. Kleyn-Kennedy (Eds.), *Metacognition: A recent review of research, theory, and perspectives* (pp. 3-43) Hauppage: Nova Science Publishers.

Helms-Lorenz, M., van de Vijver, F. J. R., & Poortinga, Y. H. (2003). Cross-cultural differences in cognitive performance and Spearman's hypothesis: g or c? *Intelligence, 31*, 9-29.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 253-270. doi:10.1037/h0023816

Inspectie van het onderwijs [Inspectorate of Education]. (2008). *Basisvaardigheden rekenen-wiskunde in het basisonderwijs [Basic math skills in elementary education].*

Jacobse, A. E. (2007). *De takentrap. Training van metacognitieve vaardigheden bij leerlingen in het basisonderwijs. [The task stairs. Training of metacognitive skills for students in primary education].* Groningen: GION.

Jacobse, A. E., & Harskamp, E. G. (2011). *A meta-analysis of the effects of instructional interventions on students' mathematics achievement.* Groningen: University of Groningen, GION.

Jacobse, A. E., & Harskamp, E. G. (2009). Student-controlled metacognitive training for solving word problems in primary school mathematics. *Educational Research & Evaluation, 15*, 447-463.

Janssen, J., & Engelen, R. (2002). *Verantwoording van de toetsen rekenen-wiskunde 2002 [Account of the mathematics tests 2002].* Arnhem: Citogroep.

Janssen, J., Van der Schoot, F., & Hemker, B. (2005). *Balans [32] van het reken-wiskundeonderwijs aan het einde van de basisschool 4. Periodieke peiling van het onderwijsniveau. [Report (32) of mathematics education at the end of elementary school 4. Periodical survey of educational level].* Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

Jitendra, A. (2002). Teaching students math problem-solving through graphic representations. *TEACHING Exceptional Children, 34*, 34-38.

Jitendra, A. K., Griffin, C. C., Haria, P., Leh, J., Adams, A., & Kaduvettoor, A. (2007). A comparison of single and multiple strategy instruction on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 99*, 115-127.

Jonassen, D. H. (2003). Designing research-based instruction for story problems. *Educational Psychology Review, 15*, 267-296.

Kapa, E. (2007). Transfer from structured to open-ended problem solving in a computerized metacognitive environment. *Learning and Instruction, 17*, 688-707.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109-129.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86.

Kistner, S., Rakoczy, K., Otto, B., Dignath-van Ewijk, C., Buttner, G., & Klieme, E. (2010). Promotion of self-regulated learning in classrooms: Investigating frequency, quality, and consequences for student performance. *Metacognition and Learning, 5*, 157-171.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*, 161-173.

KNAW. (2009). *Rekenonderwijs op de basisschool. analyse en sleutels tot verbetering. [Mathematics education in elementary school. Analysis and keys to improvement].* Alkmaar: Bejo druk & print.

Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning, 22*, 24-33.

Kramarski, B., Mevarech, Z. R., & Lieberman, A. (2001). Effects of multilevel versus unilevel metacognitive training on mathematical reasoning. *Journal of Educational Research, 94*, 292.

Kramarski, B., & Mizrachi, N. (2006). Online discussion and self-regulated learning: Effects of instructional methods on mathematical literacy. *Journal of Educational Research, 99*, 218-230.

Kramarski, B., & Revach, T. (2009). The challenge of self-regulated learning in mathematics teachers' professional training. *Educational Studies in Mathematics, 72*, 379-399.

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say.* Arlington, VA: SRI International.

Marsh, H. W., Roche, L. A., Pajares, F., & Miller, D. (1997). Item-specific efficacy judgments in mathematical problem solving: The downside of standing too close to trees in a forest. *Contemporary Educational Psychology, 22*, 363-377.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist, 40*, 257-265.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*, 49-63.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*, 14-19.

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg, T. Ben-Zeev, R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29-53). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*, 419-425.

McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1–10.

McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: Challenges and successes. *Metacognition and Learning, 6*, 195-203.

Mevarech, Z., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*, 365-394.

Mevarech, Z. R., & Amrany, C. (2008). Immediate and delayed effects of meta-cognitive instruction on regulation of cognition and mathematics achievement. *Metacognition and Learning, 3*, 147-157.

Minnaert, A. E. M. G., & Janssen, P. J. (1999). The additive effect of regulatory activities on top of intelligence in relation to academic performance in higher education. *Learning and Instruction, 9*, 77-91.

Moos, D. C., & Azevedo, R. (2008). Monitoring, planning, and self-efficacy during learning with hypermedia: The impact of conceptual scaffolds. *Computers in Human Behavior, 24*, 1686-1706.

Moreno, R. (2006). When worked examples don't work: Is cognitive load theory at an impasse? *Learning and Instruction, 16*, 170-181.

Moreno, R. (2009). Constructing knowledge with an agent-based instructional program: A comparison of cooperative and individual meaning making. *Learning and Instruction, 19*, 433-444.

Muis, K. R., Winne, P. H., & Jamieson-Noel, D. (2007). Using a multitrait-multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *The British Journal of Educational Psychology, 77*, 177-195.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report. findings from IEA's trends in international mathematics and science study at the fourth and eighth grades.* Lynch School of Education, Boston College: TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O' Sullivan, C. Y., Arora, A., & Erberber, E. (2005). *TIMMS 2007 assessment frameworks.* Boston College: TIMMS & PIRLS international study center, Lynch school of education.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102-116.

Nimon, K., & Reio, T. G., Jr. (2011). Regression commonality analysis: A technique for quantitative theory building. *Human Resource Development Review, 10*, 329-340.

OECD. (2003). *The PISA 2003 assessment framework.* Paris, France: OECD Publishing.

OECD. (2006). *Where immigrants succeed - a comparative review of performance and engagement in PISA 2003.* Paris: Organization for economic co-operation and development.

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology, 42*, 190-198.

Peters, E. E., & Kitsantas, A. (2010). Self-regulation of student epistemic thinking in science: The role of metacognitive prompts. *Educational Psychology, 30*, 27-52.

Pieschl, S. (2009). Metacognitive calibration - an extended conceptualization and potential applications. *Metacognition and Learning, 4*, 3-31.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33-40.

Pol, H. J., Harskamp, E. G., Suhre, C. J. M., & Goedhart, M. J. (2008). The effect of hints and model answers in a student-controlled problem-solving program for secondary physics education. *Journal of Science Education & Technology, 17*, 410-425.

Pressley, M., Wharton-McDonald, R., Allington, R., Collins Block, C., Morrow, L., Tracey, D., Woo, D. (2001). Strategy instruction for elementary students searching informational text. *Scientific Studies of Reading, 5*, 35-58.

Prins, F. J., Veenman, M. V. J., & Elshout, J. J. (2006). The impact of intellectual ability and metacognition on learning: New support for the threshold of problematicity theory. *Learning and Instruction, 16*, 374.

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for raven's progressive matrices and vocabulary scales. Section 3: Standard progressive matrices*. Oxford: Oxford Psychologist Press.

Raven, J. (2000). The raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1-48.

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology, 79*, 749-767.

Roeschl-Heils, A., Schneider, W., & van Kraayenoord, C. E. (2003). Reading, metacognition and motivation: A follow-up study of German students in grades 7 and 8. *European Journal of Psychology of Education, 18*, 75-86.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition: Applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning, 2*, 125-140.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*, 537-560.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545-71.

Schellings, G., & Van Hout-Wolters, B. H. A. M. (2011). Measuring strategy use with self-report instruments: Theoretical and empirical considerations. *Metacognition and Learning, 6*, 83-90.

Schellings, G. (2011). Applying learning strategy questionnaires: Problems and possibilities. *Metacognition and Learning, 6*, 91-109.

Schoenfeld, A. H. (Ed.). (1985). *Mathematical problem solving*. New York: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching.* (pp. 224-270). New York: McMilan Publishing.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31*, 13-25.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113-125.

Schraw, G. (2007). The use of computer-based environments for understanding and improving self-regulation. *Metacognition and Learning, 2*, 169-179.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*, 33-45.

Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist, 45*, 258-266.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-75.

Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review, 20*, 463-467.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*, 427-515.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271-295.

Spearman, C. (1927). *The abilities of man*. Oxford England: Macmillan.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children’s knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*, 51-79.

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation, 10*, 117-139.

Stanic, G., & Kilpatrick, J. (1989). Historical perspectives on problem solving in the mathematics curriculum. In R. I. Charles, & E. A. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 1-22). Reston, VA: National Council of Teachers of Mathematics.

Sternberg, R. J. (1984). What should intelligence tests test? Implications of a triarchic theory of intelligence for intelligence testing. *Educational Researcher, 13*, 5-15.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly, 47*, 1-41.

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology, 82*, 306-314.

Swanson, H. L. (1992). The relationship between metacognition and problem solving in gifted children. *Roeper Review, 15*, 43-48.

Swanson, H. L. (2001). Research on interventions for adolescents with learning disabilities: A meta-analysis of outcomes related to higher-order processing. *Elementary School Journal, 101*, 331-348.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96*, 471-491.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science: A Multidisciplinary Journal, 12*, 257-285.

Teong, S. K. (2003). The effect of metacognitive training on mathematical word-problem solving. *Journal of Computer Assisted Learning, 19*, 46-55.

Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *Journal of Learning Disabilities, 36*, 449-58.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies (personnel research section report no. 984).* Washington, DC: Departement of the Army.

Uiterwijk, H., & Vallen, T. (2003). Test bias and differential item functioning: A study of the suitability of the CITO primary education final test for second generation immigrant students in the Netherlands. *Studies in Educational Evaluation, 29*, 129-43.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment, 13*, 29-37.

Van den Heuvel-Panhuizen, M., & Bodin-Baarends, C. (2004). All or nothing: Problem solving by high achievers in mathematics. *Journal of the Korea Society of Mathematical Education, 8*, 115-121.

Van der Schoot, F. (2008). *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON. [Education on track? A summative review of 20 years PPON].* Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

Van der Stel, M., & Veenman, M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences, 18*, 128-134.

Van der Stel, M., & Veenman, M. V. J. (2010). Development of metacognitive skillfulness: A longitudinal study. *Learning & Individual Differences, 20*, 220-224.

Van Dijk, H., & Tellegen, P. J. (2004). *Handleiding en verantwoording NIO, Nederlandse intelligentietest voor onderwijsniveau. [Manual and account of NIO, Dutch intelligence test for educational level].* Amsterdam: Boom test uitgevers.

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

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

Van Gog, T., Paas, F., & Van Merrienboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science: An International Journal of Learning and Cognition, 32*, 83-98.

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*, 246-260.

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In Artelt, C.,Moschner,B. (Ed.), *Lernstrategien und metakognition: Implikationen für forshung und praxis.* (pp. 77-99). Münster: Waxmann.

Veenman, M. V. J. (2011a). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning, 6*, 205-211.

Veenman, M. V. J. (2011b). Learning to self-monitor and self-regulate. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). New York: Routledge.

Veenman, M. V. J., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction, 7*, 187-209.

Veenman, M. V. J., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and meta-cognitive skills in early adolescence. *Instructional Science: An International Journal of Learning and Cognition, 33*, 193.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning & Individual Differences, 15*, 159-176.

Veenman, M. V. J., & Van Hout-Wolters, B. (2002). Het meten van metacognitieve vaardigheden [Measuring metacognitive skills]. In F. Daems, R. Rymenans, G. Rogiest (Eds.). *Onderwijsonderzoek in Nederland en Vlaanderen. Proceedings 29e ORD 2002*. (pp. 102-103). Wilrijk: Universiteit van Antwerpen.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*, 3-14.

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production defiencies. *Anxiety, Stress & Coping, 13*, 391.

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning & Instruction, 14*, 89.

Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology, 92*, 308-315.

Vernon, P. E. (1950). *The structure of human abilities*. Oxford England: Wiley.

Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999a). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning, 1*, 195-229.

Verschaffel, L., De Corte, E., & Vierstraete, H. (1999b). Upper elementary school pupils' difficulties in modeling and solving nonstandard additive word problems involving ordinal numbers. *Journal for Research in Mathematics Education, 30*, 265-285.

Vilenius-Tuohimaa, P., Aunola, K., & Nurmi, J. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology, 28*, 409-426.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research, 84*, 30.

Wegerif, R., Myhill, D., Vickers, N., Goodall, O., & Allan, A. (2011). *EARLI conference 2011; education for a global networked society.* Exeter, UK.

Weinstein, C. E., Zimmermann, S. A., & Palmer, D. R. (1988). Assessing learning strategies: The design and development of the LASSI. In C. E. Weinstein, E. T. Goetz, P. A. Alexander, C. E. Weinstein, E. T. Goetz & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 25-40). San Diego, CA US: Academic Press.

Wilson, N. S., & Bai, H. (2010). The relationships and impact of teachers' metacognitive knowledge and pedagogical understandings of metacognition. *Metacognition and Learning, 5*, 269-288.

Windschitl, M. (1999). The challenges of sustaining a constructivist classroom culture. *Phi Delta Kappan, 80*, 751-55.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*, 267-276.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, M. Zeidner, M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531-566). San Diego, CA US: Academic Press.

Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review, 20*, 429-444.

Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education, 33*, 153-69.

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *Journal of Special Education, 39*, 181-192.

Yeh, S. S. (2010). Understanding and addressing the achievement gap through individualized instruction and formative assessment. *Assessment in Education: Principles, Policy & Practice, 17*, 169-182.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*, 166-183.

CURRICULUM VITAE

Annemieke Jacobse (1982) attended the Teachers' Training College for Primary Education from 2000 – 2004. After receiving her teaching degree, she worked as an elementary school teacher in different schools. While working as a teacher, she studied for her bachelor's degree in educational science. She received her bachelor degree in 2005. Moving on to the master's program at the University of Groningen, Annemieke specialized in instructional design and learning disabilities and counseling. Her master thesis was about metacognitive skills of gifted students. The thesis was awarded with the thesis-prize of the Faculty of Behavioral and Social Sciences in 2006. During her master program, Annemieke worked as a research assistant at the Groningen Institute for Educational Research (GION) of the University of Groningen. After graduation, she continued this position as a researcher. She executed an experimental study and co-organized an educational conference (ORD) in Groningen. After finishing these projects, she started her PhD project at the GION in 2007. Besides working on the studies in this dissertation, Annemieke has been project leader for multiple short research projects, she executed a meta-analysis about mathematical interventions with her supervisor for the Dutch Programme Council for Educational Research and she was part of the organizing committee of the Junior Researchers of Earli conference in Frankfurt. In 2010 she co-wrote a research proposal for Kennisnet and received funding for a two-year project on metacognitive training in comprehensive reading. Annemieke is currently employed as a teacher and researcher at the GION.

ICO DISSERTATION-LIST

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the following universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Groningen, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University (and formerly Radboud University Nijmegen and Tilburg University).

ICO-dissertations 2009/2010:

220. Bruin-Muurling, G. (21-12-2010). *The development of proficiency in the fraction domain: Affordances and constraints in the curriculum*. Eindhoven: Eindhoven University of Technology.

219. Kostons, D.D.N.M. (05-11-2010). *On the role of self-assessment and task-selection skills in self-regulated learning*. Heerlen: Open University of the Netherlands.

218. Vos, M.A.J. (30-09-2010). *Interaction between teachers and teaching materials: On the implementation of context-based chemistry education*. Eindhoven: Eindhoven University of Technology.

217. Bonestroo, W.J. (24-09-2010). *Planning with graphical overview: Effects of support tools on self-regulated learning*. Enschede: University of Twente.

216. Groenier, M. (10-09-2010). *The decisive moment: Making diagnostic decisions and designing treatments*. Enschede: University of Twente.

215. De Bakker, G.M. (08-09-2010). *Allocated online reciprocal peer support as a candidate for decreasing the tutoring load of teachers*. Eindhoven: Eindhoven University of Technology.

214. Endedijk, M.D. (02-07-2010). *Student teachers' self-regulated learning*. Utrecht: Utrecht University.

213. Kessels, C.C. (30-06-2010). *The influence of induction programs on beginning teachers' well-being and professional development*. Leiden: Leiden University.

212. Duijnhouwer, H. (04-06-2010). *Feedback effects on students' writing motivation, process, and performance*. Utrecht: Utrecht University.

211. Moolenaar, N.M.(01-06-2010). *Ties with potential: Nature, antecedents, and consequences of social networks in school teams*. Amsterdam: University of Amsterdam.

210. Mittendorff, K. M. (12-03-2010). *Career conversations in senior secondary vocational education*. Eindhoven: Eindhoven University of Technology.

209. Platteel, T. (11-02-2010). *Knowledge development of secondary school L1 teachers on concept-context rich education in an action-research setting*. Leiden: Leiden University.

208. Koopman, M. (11-02-2010). *Students' goal orientations, information processing strategies and knowledge development in competence-based pre-vocational secondary education*. Eindhoven: Eindhoven University of Technology.

207. Zitter, I.I. (04-02-2010). *Designing for learning: Studying learning environments in higher professional education from a design perspective*. Utrecht: Utrecht University.

206. Wetzels, S.A.J. (18-12-2009). *Individualized strategies for prior knowledge activation*. Heerlen: Open University of the Netherlands.

205. Kirschner, F.C. (11-12-2009). *United brains for complex learning*. Heerlen: Open University of the Netherlands.

204. Rohaan, E.J. (09-12-2009). *Testing teacher knowledge for technology teaching in primary schools*. Eindhoven: Eindhoven University of Technology.

203. Seezink, A. (18-11-2009). *Continuing teacher development for competence-based teaching*. Tilburg: Tilburg University.

202. Van Ewijk, R. (20-10-2009). *Empirical essays on education and health*. Amsterdam: University of Amsterdam.

201. Mainhard, M.T. (25-09-2009). *Time consistency in teacher-class relationships*. Utrecht: Utrecht University.

200. Van der Rijst, R.M. (23-09-2009). *The research-teaching nexus in the sciences: Scientific research dispositions and teaching practice*. Leiden: Leiden University.

199. Visser-Wijnveen, G.J. (23-09-2009). *The research-teaching nexus in the humanities: Variations among academics*. Leiden: Leiden University.

198. Nievelstein, F.E.R.M. (18-09-2009). *Learning law: Expertise differences and the effect of instructional support*. Heerlen: Open University of the Netherlands.

197. Handelzalts, A. (17-09-2009). *Collaborative curriculum development in teacher design teams*. Enschede: University of Twente.

196. Van Borkulo, S.P. (26-06-2009). *The assessment of learning outcomes of computer modelling in secondary science education*. Enschede: University of Twente.

195. Du Chatenier, E. (09-09-2009). *Open innovation competence: Towards a competence profile for inter-organizational collaboration in innovation teams*. Wageningen: Wageningen University.

194. Radstake, H. (14-05-2009).*Teaching in diversity: Teachers and pupils about tense situations in ethnically heterogeneous classes*. Amsterdam: University of Amsterdam.

thinking
about thinking
solves more problems
than you would have thought