Using mini-games for learning multiplication and division:

A longitudinal effect study

This research was funded by a grant from the OnderwijsBewijs program of the Ministry of Education in the Netherlands (project number: ODB 08007).

This research was carried out in the context of the Dutch Interuniversity Centre for Educational Research.

ico

Bakker, Marjoke

Using mini-games for learning multiplication and division: A longitudinal effect study / M. Bakker – Utrecht: Freudenthal Institute for Science and Mathematics Education, Faculty of Science, Utrecht University / FIsme Scientific Library (formerly published as CD-β Scientific Library), no 83, 2014.

Dissertation Utrecht University. With references. Met een samenvatting in het Nederlands.

ISBN: 978-90-70786-24-3
Key words: primary mathematics education / multiplication and division / educational computer games / attitude towards mathematics

Cover design: Vormgeving Faculteit Bètawetenschappen
Printed by: Drukkerij Wilco BV

USING MINI-GAMES FOR LEARNING MULTIPLICATION AND DIVISION: A LONGITUDINAL EFFECT STUDY

HET GEBRUIK VAN MINI-GAMES VOOR HET LEREN VERMENIGVULDIGEN EN DELEN: EEN LONGITUDINALE EFFECTSTUDIE

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G. J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 16 april 2014 des middags te 12.45 uur

door
Marjoke Bakker

geboren op 27 januari 1982
te Deventer

Promotor: Prof. dr. M. H. A. M. van den Heuvel-Panhuizen

Table of contents

Chapter 1

Introduction

Introduction

One of the key domains in primary school mathematics education is the domain of multiplication and division, or multiplicative reasoning. On the one hand, the importance of multiplicative reasoning lies in the fact that many situations encountered in daily life are of a multiplicative nature. One can think of, for example, situations in which amounts of ingredients for cooking for a certain number of persons have to be determined, or situations in which sweets have to be equally divided among a number of children. On the other hand, multiplicative reasoning is foundational for the understanding of many of the mathematical concepts that are met with in the later school career, such as ratio, fractions, and linear functions (see, e.g., Vergnaud, 1983).

When thinking of the domain of multiplication and division, what often comes to mind is the learning of the multiplication tables. Indeed, children should have ready knowledge of multiplication number facts, and skills in quickly computing arithmetic problems. Together, these fact knowledge and skills – also called declarative knowledge and procedural knowledge, respectively (see, e.g., Goldman & Hasselbring, 1997; Miller & Hudson, 2007) – are often referred to as basic skills. Though basic skills are important, it must be noted that skills should not be learned without understanding (see, e.g., Anghileri, 2006; Freudenthal, 1991; Kilpatrick, Swafford, & Findell, 2001). Learning without understanding leads to fragmented pieces of knowledge, and prevents children from connecting new types of mathematics problems to earlier acquired knowledge (e.g., Kilpatrick et al., 2001). Therefore, the third knowledge type important in learning mathematics is conceptual knowledge (e.g., Hiebert & Lefevre, 1986; Miller & Hudson, 2007). When students not only automatize basic skills, but also have conceptual knowledge of, or insight in, the underlying number relations, concepts, and strategies, they can flexibly apply their knowledge and skills in new situations. For the domain of multiplication and division, this means that students should, for example, develop an understanding of multiplicative situations in terms of a number of groups with the same number of items in each group (e.g., 3 bags of 6 apples). In addition, students should gain insight in number relations such as doubles and halves of numbers and divisibility of numbers, and insight in the properties of multiplicative operations, including the commutative property (e.g., $3 \times 8 = 8 \times 3$), the distributive property (e.g., $7 \times 8 = 5 \times 8 + 2 \times 8$), and the associative property (e.g., $2 \times 12 = 4 \times 6$). The combination of fact knowledge and skills on the one hand, and understanding or insight on the other hand, has been a key component in recent recommendations for primary school mathematics education (e.g., Common Core State Standards Initiative, 2010; Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008).

One promising way to promote students' multiplicative fact knowledge and skills, as well as their insight in multiplicative concepts and relations, is through the use of educational computer games. As computer games are often very engaging, they can provide a motivating environment for practicing mathematics fact knowledge and operations, in this way fostering students' automatization of basic skills. Moreover, games can allow for exploration and experimentation (e.g., Kirriemuir, 2002), enabling students to learn new concepts and number relations through experience in the game, and to discover which strategies are useful. Playing such games is, thus, expected to contribute to students' mathematical insight.

The present thesis reports about the research findings gained in a longitudinal research project carried out to investigate the effectiveness of mathematics mini-games in the mathematics domain of multiplicative reasoning. In the following I provide some background related to the topic of the thesis. After that, I introduce our research project and give an overview of the thesis.

1 Background

1.1 Computer games in education

Computer games are very popular among children as well as adults. In the Netherlands, almost all primary school children play computer games on a regular basis (e.g., Bijlsma, 2007; Jeugdpeil, 2010). The appeal of computer games is thought to lie in, among others, the challenging and curiosity-provoking environment provided by games (e.g., Malone, 1981).

Ever since computer games came up, they have been considered a useful tool to be employed for educational purposes (e.g., Egenfeldt-Nielsen, 2005; Malone, 1981; Prensky, 2001). The motivational character of games is seen as a potential catalyst for learning (e.g., Garris, Ahlers, & Driskell, 2002; Malone, 1981): it can lead students to put more attention in a learning activity when it is presented through a game. Also, students who learn with games tend to spend more time on learning, which may positively influence their learning outcomes (e.g., Sandberg, Maris, & De Geus, 2011; Tobias, Fletcher, Dai, & Wind, 2011). Related to this, researchers have indicated that children also play educational games in their free time (e.g., Ault, Adams, Rowland, & Tiemann, 2010; Jonker, Wijers, & Van Galen, 2009). This points to the possibility of extending the learning time through offering educational games for playing at home (e.g., Sandberg et al., 2011).

Another beneficial characteristic of computer games is their potential to give immediate feedback (e.g., Prensky, 2001). Students can often instantly see the consequences of their actions in the game. When games are used for practicing, this immediate feedback is useful because students directly see whether their answers are correct or not. Moreover, this immediate feedback, together with the relatively anonymous, risk-free environment provided by a game, can encourage students to explore and experiment in the game, which may lead to discovering new concepts or strategies (e.g., Kirriemuir, 2002). This learning through exploring and experimenting is often termed experiential learning (e.g., Egenfeldt-Nielsen, 2005; Garris et al., 2002).

Apart from promoting learning, educational computer games may also benefit students' attitude toward the subject matter. Playing a motivating game related to a particular subject matter is likely to evoke positive emotional experiences associated with this subject matter. Such positive experiences may in turn lead to a more positive attitude – e.g., enjoyment or interest – towards the subject in the long run (e.g., Hidi & Renninger, 2006; McLeod, 1992).

Despite the many promises of computer games for education, empirical evidence on their effectiveness is still rather sparse. In a recent meta-analysis, Wouters, Van Nimwegen, Van Oostendorp, and Van der Spek (2013) found an overall positive effect of educational games on learning, but when only randomized studies were taken into account they did not find a significant effect. Other review authors stated that the research base on the effectiveness of educational games is still insufficient and that class-based longitudinal experiments are needed (e.g., Tobias et al., 2011; Young et al., 2012). Several drawbacks of earlier studies have been noted, such as the absence of a control group (e.g., Vogel et al., 2006), no random assignment to conditions (e.g., Slavin & Lake, 2008), and small sample sizes (e.g., Bai, Pan, Hirumi, & Kebritchi, 2012). Regarding the attitudinal effects of games, likewise, little empirical research has been carried out yet (see, e.g., Wouters et al., 2013).

1.2 Computer games for mathematics education

Also for the specific case of mathematics education, games have long been proposed as promising learning tools. A type of game that is often used in mathematics education is the so-called mini-game (e.g., Jonker et al., 2009; Panagiotakopoulos, 2011). Mini-games are short, focused games that are easy to learn (e.g., Frazer, Argles, & Wills, 2007; Jonker et al., 2009). They are clearly different from complex serious games, which usually take hours to play and are often less connected to the curriculum (e.g., Prensky, 2008). Because of their "mini-ness", mini-games are commonly easily accessible (often free of charge) and have low technical requirements. Moreover, they can be played for flexible time durations: often a game can be finished in just a couple of minutes, and can be repeated at will (Jonker et al., 2009). These characteristics are important in implementing games in education (e.g., Kebritchi, 2010).

Many educational games or software for learning mathematics primarily focus on drill-and-practice, offering a motivating environment for performing the repetitive mathematics activities necessary to achieve automaticity in number fact knowledge and operation skills. Examples are handheld games for practicing arithmetic facts (e.g., Miller & Robertson, 2011; Shin, Sutherland, Norris, & Soloway, 2012), and the Dutch mathematics games program Rekentuin (Math Garden, see Jansen et al., 2013).

Next to the possibilities of games in automatizing students' fact knowledge and skills, games can also be used for gaining conceptual understanding of, or insight in, mathematics concepts, relations, and strategies, as was emphasized by, among others, Klawe (1998) and Jonker et al. (2009). Such games are often based on the previously mentioned experiential learning (e.g., Kebritchi, Hirumi, & Bai, 2010). By exploring and experimenting in the game students can discover useful mathematical relations and strategies, and learn new concepts and rules. Van Galen, Jonker, and Wijers (2009), for example, described a game in which students can experiment with divisibility and factors of numbers while decorating a pie.

1.3 Empirical research in the educational practice

In recent years, researchers and public authorities have increasingly stated the importance of evidence-based education (e.g., Baron, 2002; Onderwijsraad, 2006). It is stressed that innovations in the educational practice should be based on sound empirical evidence of what works in education, just as new treatments in medicine should be rooted in rigorous empirical proof. The often suggested approach for gaining evidence for the effectiveness of educational interventions in schools or classes (e.g., learning materials, teaching practices, etc.) is through the use of cluster randomized controlled trials in the school practice (e.g., Onderwijsraad, 2006; Towne & Hilton, 2004). In such experiments, schools or classes are randomly assigned to an experimental or control condition, and pre- and posttests are used to determine the effectiveness of an intervention.

For the case of using computer games – or, more generally, ICT – in education, it is especially important to emphasize the need for empirical evidence of effectiveness in the school practice. One reason is that, because using ICT has become so popular in the recent decades, people often tend to assume it is effective based on some presumably favorable characteristics of ICT, rather than based on sound evidence. This is reflected in the findings from review studies that much research has focused on design or small-scale evaluation of educational games rather than on gaining empirical evidence of their effectiveness (e.g., Tobias et al., 2011; Young et al., 2012). Another major concern is that the use of ICT in the educational practice is often hindered by several practical issues, such as limited available time for teachers to embed ICT in their lessons or to get used to new ICT applications, and infrastructural limitations like limited numbers of computers or technical problems with computers (see, e.g., Bingimlas, 2009). This means that an ICT-based educational intervention may be found effective in a relatively controlled setting, but may fail to lead to the expected learning outcomes when implemented in real educational practice.

2 The BRXXX project

In line with the recent calls for evidence-based research in education, in 2008 the Dutch Ministry of Education launched a research program called OnderwijsBewijs [Evidence-based Education], for which research proposals could be submitted for projects using randomized experiments to test the effectiveness of educational interventions. The BRXXX project[1], proposed by prof. dr. Marja van den Heuvel-Panhuizen as the principal investigator of the project, was of the 112 proposals one of the 18 proposals that was granted. The project, which started in September 2009 and ran till December 2013, was a collaborative undertaking: Besides the principal investigator, the research team consisted of a postdoc researcher, a primary school teacher, a primary school head teacher, a special education teacher, and me as the PhD student. The teachers (two of whom had to leave the project within the first year because of other duties or change of jobs), were involved for one day a week in communicating with the participating schools, trying out tests and games, and preparing guidelines for teachers to implement the intervention. The teacher that kept working in the project till the end also prepared articles about the research results to be published in teacher journals.

The present PhD thesis gives a full overview of the research activities carried out in the project and our findings. The main focus of the BRXXX project was on investigating the effectiveness of online mini-games aimed at developing multiplicative reasoning ability. The games used in the project were mostly adapted versions of multiplicative mini-games selected from the Dutch mathematics games website Rekenweb (www.rekenweb.nl, English version: www.thinklets.nl). The games addressed both practicing multiplicative number facts and operations, and developing insight in multiplicative concepts and number relations. In accordance with the abovementioned requests for gaining educational evidence in the school practice, the project specifically aimed at investigating the effectiveness of the online mini-games when they are integrated in the educational practice of primary school, that is, when they are used as part of the regular educational program for multiplicative reasoning. In a cluster randomized controlled trial, we studied the effectiveness of different ways of deploying multiplicative mini-games in mathematics education, including playing at school, playing at home, and playing at home with afterwards a discussion (debriefing) at school. A large number of students from regular primary schools participated in the project. These students were followed from the end of Grade 1 (Dutch *groep 3*) to the end of Grade 4 (Dutch *groep 6*). Also included were students from special education schools, who were followed from the end of Grade 1 to the end of Grade 2.

Besides studying the learning effects of the multiplicative mini-games, which was the main focus of the BRXXX project, we also examined students' initial knowledge in the domain

[1] Project number ODB 08007. The orginal titel of the project was "Basisvaardigheden leren met RekenXXX-games".

of multiplicative reasoning at the end of Grade 1. Furthermore, we studied students' attitude towards mathematics and the influence of playing the games on this attitude.

3 Thesis overview

Within our large scale longitudinal research project, we performed several separate studies, which are described in the next five chapters. The final chapter of the thesis presents a summary of our findings and a final conclusion. Table 1 gives an overview of the main topics treated in the different chapters.

Table 1
Thesis overview

<table>
<tr><th>Chapter</th><th>Topic</th></tr>
<tr><td>1</td><td>Introduction</td></tr>
<tr><td>2</td><td>What knowledge in the domain of multiplicative reasoning do students already have just before they start receiving formal instruction on this domain?</td></tr>
<tr><td>3</td><td rowspan="2">What are the effects of mini-games on students' multiplicative reasoning ability? – Regular education</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td><td>What are the effects of mini-games on students' multiplicative reasoning ability? – Special education</td></tr>
<tr><td>6</td><td>How does students' attitude towards mathematics develop over time, and what is the influence of playing mathematics mini-games on this attitude?</td></tr>
<tr><td>7</td><td>Summary and conclusion</td></tr>
</table>

In *Chapter 2*, we examined students' scores on the first test of multiplicative reasoning ability administered in the BRXXX project (in regular primary education). Though originally this test was only meant as a baseline test, we found it worthwile to separately report on the test results. Because this test was administered at the end of Grade 1, just before multiplicative reasoning was formally introduced in Grade 2, the test scores provided insight in students' "pre-instructional knowledge" of multiplicative reasoning, that is, the extent to which students are already able to solve multiplication and division problems before they start receiving formal instruction on this domain. Apart from investigating students' test scores, we also looked at whether there were differences in difficulty between different types of problems. Furthermore, we examined whether there were differences in multiplicative reasoning ability between students with different characteristics and between schools using different mathematics textbooks.

In *Chapter 3* we describe a study on the effects of the mini-games on students' multiplicative reasoning ability in Grade 2 of regular primary education. This chapter is, thus, about the first year of the longitudinal study. We report the effects of the three different ways of deploying the mini-games (playing at school, playing at home, and playing at home with debriefing at school) on a combined measure of multiplicative reasoning ability, comprising multiplicative operation skills and insight in multiplicative concepts and relations. Chapter 3 is in Dutch, since it was published in a Dutch scientific journal. An English abstract is included. As this chapter has quite some overlap with Chapter 4 (with Chapter 4 being far more extensive), skipping Chapter 3 will not hamper the reading of this thesis.

Chapter 4 reports on the effects of the full two-year mini-games intervention in regular primary education. We examined the effects of the mini-games – again either played at school, played at home, or played at home and debriefed at school – in both Grade 2 and Grade 3, on the abovementioned three different aspects of multiplicative reasoning ability: number fact knowledge, operation skills, and insight in multiplicative concepts and relations. Furthermore, we looked at the role of students' gender and prior mathematics ability, and the role of the time and effort spent on the games. The retention effect of the games on students' multiplicative reasoning ability at the end of Grade 4 was beyond the scope of the current thesis.

In *Chapter 5*, we report the effects of the mini-games in special education. Here we only examined the effectiveness of the mini-games when they were played at school.

Chapter 6 deals with students' attitude towards mathematics, which we conceptualized as students' liking, or enjoyment, of the subject of mathematics. We examined the development of regular primary education students' mathematics attitude over the entire three-year period of the study. Furthermore, we looked at the relation with gender, mathematics ability, and the time and effort spent on the games.

Finally, in *Chapter 7*, I summarize the findings of this PhD research. I give some practical implications and suggestions for further research, and I end with the main conclusion of the thesis.

References

Anghileri, J. (2006). *Teaching number sense* (2nd ed.). London: Continuum.

Ault, M., Adams, D., Rowland, A., & Tiemann, G. (2010, June). *Targeted educational games: Fun and so much more!* Paper presented at the annual conference of the International Society for Technology in Education, Denver, Colorado.

Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology, 43,* 993-1003.

Baron, J. (2002). *Bringing evidence-driven progress to education: A recommended strategy for the U.S. Department of Education.* Retrieved from http://www.education-consumers.org/research/Evidence-based%20policy%20recommendation.pdf

Bijlsma, J. (2007). *Kan het onderwijs om games heen? Onderzoek naar het computergebruik en beleving van leerlingen in de bovenbouw van het basisonderwijs* [Can education do without games? A study of computer usage and perception in upper primary school students]. Retrieved from http://www.teamonderwijs.nl/download/product/Kan_het_onderwijs_om_games_heen.pdf

Bingimlas, K. A. (2009). Barriers to the successful integration of ICT in teaching and learning environments: A review of the literature. *Eurasia Journal of Mathematics, Science & Technology Education, 5*(3), 235-245.

Common Core State Standards Initiative (2010). Common core state standards for mathematics. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment. Exploring the educational potential of computer games* (Doctoral dissertation). IT-University of Copenhagen, Denmark.

Expertgroep Doorlopende Leerlijnen Taal en Rekenen [Expert group learning trajectories language and mathematics]. (2008). *Over de drempels met rekenen: Consolideren, onderhouden, gebruiken en verdiepen* [Across the tresholds in mathematics: Consolidating, maintaining, applying and deepening]. Enschede, the Netherlands: Author.

Frazer, A., Argles, D., & Wills, G. (2007, September). *Assessing the usefulness of mini-games as educational resources.* Paper presented at the annual conference of the Association for Learning Technology, Nottingham, UK.

Freudenthal, H. (1991). *Revisiting mathematics education: China lectures.* Dordrecht, the Netherlands: Kluwer Academic.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33,* 441-467.

Goldman, S. R., & Hasselbring, T. S. (1997). Achieving meaningful mathematics literacy for students with learning disabilities. *Journal of Learning Disabilities, 30,* 198-208.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41,* 111-127.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analyis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum.

Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24,* 190-197.

Jeugdpeil. (2010). *Jeugdpeil gamen* [Youth monitor gaming]. Retrieved from http://www.jeugdpeil.nl/

Jonker, V., Wijers, M., & Van Galen, F. (2009, October). *The motivational power of mini-games for the learning of mathematics.* Paper presented at the European Conference on Game Based Learning, Graz, Austria.

Kebritchi, M. (2010). Factors affecting teachers' adoption of educational computer games: A case study. *British Journal of Educational Technology, 41,* 256-270.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55,* 427-443.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Kirriemuir, J. (2002). *The relevance of video games and gaming consoles to the higher and further education learning experience.* Techwatch Report. Retrieved from http://tecnologiaedu.us.es/ nweb/ htm/ pdf/ 301.pdf

Klawe, M. M. (1998, June). *When does the use of computer games and other interactive multimedia software help students learn mathematics?* Paper presented at the Technology and NCTM Standards 2000 Conference, Arlington, VA.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4,* 333-369.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). New York: Macmillan.

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research & Practice, 22,* 47-57.

Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology, 42,* 850-864.

Onderwijsraad [Council of Education]. (2006). *Naar meer evidence based onderwijs* [To more evidence-based education]. Den Haag, the Netherlands: Author.

Panagiotakopoulos, C. T. (2011). Applying a conceptual mini game for supporting simple mathematical calculation skills: Students' perceptions and considerations. *World Journal of Education, 1*(1), 3-14.

Prensky, M. (2001). *Digital game-based learning.* New York: McGraw-Hill.

Prensky, M. (2008). Students as designers and creators of educational computer games: Who else? *British Journal of Educational Technology, 39,* 1004-1019.

Sandberg, J., Maris, M., & De Geus, K. (2011). Mobile English learning: An evidence-based study with fifth graders. *Computers & Education, 47,* 1334-1347.

Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology, 43,* 540-560.

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

Tobias, S., Fletcher, J.D., Dai, D.Y., & Wind, A.P. (2011). Review of research on computer games. In S. Tobias & J.D. Fletcher (Eds.), *Computer games and instruction* (pp. 127-222). Charlotte, NC: Information Age.

Towne, L., & Hilton, M. (2004). Implementing randomized field trials in education: Report of a workshop. Washingtion, DC: National Academies Press. Retrieved from http://www.nap.edu/catalog/10943.html

Van Galen, F., Jonker, V., & Wijers, M. (2009). Designing educational mini-games. Paper presented at the 5th Annual Conference of the International Society for Design and Development in Education, Cairns, Australia. Retrieved from http://www.isdde.org/isdde/cairns/pdf/papers/isdde09_vangalen.pdf

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 127–174). New York: Academic Press.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34,* 229-243.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105,* 249-265.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., ... & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61-89.

Chapter 2

First-graders' knowledge of multiplicative reasoning before formal instruction in this domain

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2014). First-graders' knowledge of multiplicative reasoning before formal instruction in this domain. *Contemporary Educational Psychology, 39,* 59-73. doi: 10.1016/j.cedpsych.2013.11.001

First-graders' knowledge of multiplicative reasoning before formal instruction in this domain

1 Introduction

Children usually have already built up a considerable amount of mathematical knowledge before they get their first formal instruction in mathematics (e.g., Aubrey, 1994; Carpenter & Moser, 1984; Ginsburg, Klein, & Starkey, 1998). This knowledge is often referred to as informal knowledge (e.g., Baroody, 1987; Ginsburg et al., 1998; Olivier, Murray, & Human, 1990), and is constructed in response to everyday experiences (e.g., Ginsburg et al., 1998; Leinhardt, 1988). Many mathematics educators have stated the importance of building on children's informal mathematical knowledge when teaching them mathematics (e.g., Baroody, 1987; Ginsburg, 1977; Hiebert, 1984; Leinhardt, 1988). They argue that through their informal knowledge children can give meaning to the formal symbols and procedures of mathematics (e.g., Baroody, 1987; Hiebert, 1984). Not building on the knowledge children bring with them may result in children acquiring superficial knowledge without understanding (e.g., Baroody, 1987; Hiebert, 1984), leading, for example, to the erroneous use of mathematical procedures and difficulties in memorizing them (e.g., Baroody, 1987; Olivier et al., 1990).

This building on children's existing knowledge is not only important when children have their first lessons in mathematics, but is also relevant later in the learning process, when a new mathematics domain, such as multiplication, is introduced (e.g., Kouba & Franklin, 1993; Mack, 1995). In this case, children bring with them informal knowledge about multiplication acquired through everyday experiences, as well as prior knowledge acquired from formal mathematics instruction on the related domain of addition. Also, earlier mathematics instruction may have involved preparatory multiplicative activities. As in this case it is hard to distinguish knowledge that is acquired outside school (informal knowledge) from knowledge that is acquired in earlier mathematics lessons, we prefer to speak of pre-instructional knowledge of a certain mathematics domain, including all the knowledge that children have available before formal instruction on that domain starts, regardless of its source.

Despite the stated importance of connecting the formal mathematics to children's (informal) pre-instructional knowledge, researchers have found that teachers often fail to make these connections (e.g., Aubrey, 1994; Leinhardt, 1988). A possible explanation for this may be that teachers underestimate children's pre-instructional knowledge. Several studies have found evidence for such underestimations of children's mathematics abilities (e.g., Grassmann, Mirwald, Klunter, & Veith, 1995; Lee & Ginsburg, 2009; Selter, 1993; Van den Heuvel-Panhuizen, 1996). For teachers to be able to build on children's prior knowledge, it is at least necessary that they are aware of and acknowledge this knowledge. In fact, it is argued that knowledge of what children already know about a particular mathematics domain should be an important aspect of teachers' didactical knowledge (Carpenter, Fennema, & Franke, 1996). Therefore, it is crucial that the pre-instructional knowledge of children is revealed.

This paper describes a study into children's pre-instructional knowledge in the domain of multiplicative reasoning.[1] [2] We aimed to map children's understanding of multiplication and division just before they start receiving formal instruction on this domain.

2 Theoretical background and research questions

2.1 Multiplicative reasoning

The mathematics domain of multiplicative reasoning, comprising multiplication and division, is clearly distinguished from the domain of additive reasoning, including addition and subtraction (e.g., Clark & Kamii, 1996; Schwartz, 1988; Vergnaud, 1983). In contrast to additive reasoning, in which quantities of the same type are added or subtracted (e.g., 2 *cookies* and 3 *cookies* are 5 *cookies* altogether), multiplicative reasoning involves quantities of different types (e.g., 3 *boxes* with 4 *cookies per box* means 12 *cookies* altogether). Accordingly, Schwartz (1988, p. 41) asserted that addition and subtraction are "referent preserving compositions", whereas multiplication and division are "referent transforming compositions". A multiplicative situation is characterized by a group structure which involves sets (groups, e.g., boxes) of items with in each set the same number of items (e.g., cookies) (see Greer, 1992). This distinction between items and sets of items was emphasized by Nantais and Herscovics (1990, p. 289), stating that "a situation is perceived as being multiplicative when the whole is viewed as resulting from the repeated iteration of a one-to-one or a one-to-many correspondence". In this definition, a one-to-one correspondence refers to the situation where there is one item in each set, whereas in the case of a one-to-many correspondence, the sets contain more than one item. Although multiplication problems can be calculated by repeated addition or counting in groups, which is how they are often introduced to children, multiplication is conceptually different from addition, since one of the operands denotes the number of times a value should be added (the number of sets), instead of a value to be added (see, e.g., Clark & Kamii, 1996).

Multiplicative reasoning has an important place in primary mathematics learning, since it is required as a foundation for the understanding of more complex mathematical concepts in the multiplicative conceptual field (Vergnaud, 1983), such as ratio, fractions, and linear functions. These concepts are all related to proportional reasoning, which Lesh, Post, and

[1] In this chapter, we use the term knowledge to denote all types of knowledge of multiplicative reasoning, that is, declarative, procedural, and conceptual knowledge.
[2] We use the terms *pre-instructional knowledge of multiplicative reasoning* and *pre-instructional multiplicative knowledge* interchangeably.

Behr (1988, p. 94) described as both the "capstone" of primary school mathematics and the "cornerstone" of the mathematics that follows. Besides its importance for later mathematical understanding, multiplicative reasoning is implicitly necessary for understanding place value (e.g., interpreting 63 as 6 tens and 3 ones; see Nunes et al., 2009).

Formal instruction on multiplicative reasoning generally starts in the second grade (e.g., in the Netherlands; see Van den Heuvel-Panhuizen, 2008) or third grade (e.g., in the US; see NCTM, 2006) of primary school, after addition and subtraction have been taught. Often, division is formally introduced after multiplication (see Mulligan & Mitchelmore, 1997; Van den Heuvel-Panhuizen, 2008).

2.2 Previous research on children's pre-instructional knowledge of multiplicative reasoning

Earlier studies have revealed that young children already have some understanding of multiplicative relations before the domain is formally introduced in school (e.g., Anghileri, 1989; Kouba, 1989; Mulligan & Mitchelmore, 1997; Nunes & Bryant, 1996; see also Ter Heege, 1985). In Anghileri's (1989) study, for example, first-grade students could solve an average of 56% of physically presented multiplication tasks, and in Kouba's (1989) study, first graders could already solve some simple multiplication and division word problems (25% correct on average). Furthermore, in a longitudinal study by Mulligan and Mitchelmore (1997), Australian children at the beginning of Grade 2 correctly solved an average of 31% of multiplicative word problems, increasing to 48% at the end of Grade 2 and 55% at the beginning of Grade 3 (all these measurements were before formal instruction on multiplicative reasoning). Carpenter and colleagues found that even many kindergartners were able to solve a variety of multiplication and division word problems (Carpenter, Ansell, Franke, Fennema, & Weisbeck, 1993).

In all previous studies on children's pre-instructional knowledge of multiplicative reasoning, the problems were either presented in a physical context (e.g., Anghileri, 1989) or the children were allowed and encouraged to use physical materials, such as counters and blocks, to construct a physical representation for themselves (e.g., Kouba, 1989; Mulligan & Mitchelmore, 1997). The majority of the children did actually employ these materials (Carpenter et al., 1993; Kouba, 1989). This probably helped them in modeling the problem situation and in keeping track of counting and repeated addition or subtraction activities, and thus made it easier to solve the problems (see, e.g., Ibarra & Lindvall, 1982; Levine, Jordan, & Huttenlocher, 1992). From the previous studies, then, it is not known whether children also show this knowledge when no physical representation is offered or can be created by the child. Moreover, the studies have only focused on problems presented in a context and not on bare number problems, like "$2 \times 4 =$ ___" or "2 times 4 is ___".[3] Furthermore, in the aforementioned studies the children were assessed in individual interviews, in which the interviewer could have encouraged the children in reaching a solution. It has indeed been found that one-to-one interview settings may help students in solving mathematics problems (Caygill & Eley, 2001). In the previous studies it was not investigated whether children also show pre-instructional knowledge of multiplicative reasoning when they are assessed in a more formal setting, in which there is no interviewer sitting next to them. Finally, the previous studies were small-scale studies, which may make results hard to generalize.

2.3 Possible factors influencing children's pre-instructional multiplicative knowledge

Research suggests that there are several factors that may influence the pre-instructional multiplicative knowledge children display. Below we discuss the most important characteristics that we found in the literature. First of all, the characteristics of the problems offered to the children may affect their performance. In addition, children's gender, the educational level of their parents, and the mathematics textbook used in class may have an influence. The latter two can be seen as indicators of the environment in which children have developed their knowledge.

2.3.1 Problem characteristics

Problem format. Arithmetic problems can be presented either as a context problem[4] embedded in a situation, or as a bare number problem without a context. Research has shown that, for students who have had no or only limited formal instruction on a particular mathematics domain, context problems in that domain are often easier to solve than bare number problems (e.g., Koedinger & Nathan, 2004; Levine et al., 1992; Van de Heuvel-Panhuizen, 2005). This can be explained by the fact that context problems, unlike bare number problems, relate to real-life situations and thus can elicit the use of informal mathematical knowledge (Koedinger & Nathan, 2004) and in this way suggest strategies for solving the problem (Van de Heuvel-Panhuizen, 2005). What also may contribute to the relative easiness of context problems is when the problem includes a picture displaying (part of) the multiplicative situation. Especially when context problems include a picture involving countable objects, we expect this to decrease the problems' difficulty level, since

[3] Baroody (1999) did study first-graders' abilities in solving bare number multiplication problems. However, in his study the children were first introduced to the × symbol, which can be considered a first formal instruction on multiplication. Baroody's study, thus, was not performed before formal instruction.

[4] We use the term *context problems*, instead of *word problems* or *story problems*, as it better accounts for the fact that contexts can involve both non-verbal and verbal elements (see also Van den Heuvel-Panhuizen, 2005).

in this case, as when using physical materials, the depicted objects can be used to find the solution by counting. This “countability” aspect of pictures in context problems has, to our knowledge, not been investigated before. Importantly, we note that pictures with countable objects are not the same as the aforementioned physical objects, as pictured objects cannot be moved or manipulated (e.g., Martin & Schwartz, 2005; Moyer, Bolyard, & Spikell, 2002). Indeed, studies showed that children using pictures in solving problems in a new mathematics domain were less successful than children using physical manipulatives (Martin & Schwartz, 2005; Martin, Lukong, & Reaves, 2007). This indicates that context problems with pictures can still be seen as more formal or abstract than the problems with physical representations used in the previous studies on children’s pre-instructional multiplicative knowledge.

Semantic structure. Multiplicative situations can be classified into different categories (e.g., Schwartz, 1988; Vergnaud, 1983), often referred to as semantic structures (e.g., De Corte, Verschaffel, & Van Coillie, 1988; Greer, 1992; Mulligan & Mitchelmore, 1997). The three semantic structures which, according to their lower difficulty level, are relevant for first-graders’ multiplicative reasoning are *equal groups* (e.g., 3 boxes with 4 cookies each), *rectangular array* (e.g., 3 rows of 4 chairs), and *rate* (e.g., 1 cake costs 3 euros, how much do 4 cakes cost?). Of these semantic structures, equal groups has generally been found to be easiest (e.g., Christou & Philippou, 1999; Nesher, 1992). For the case of pre-instructional multiplicative knowledge, though, Mulligan and Mitchelmore (1997) did not find differences in difficulty level between the three abovementioned semantic structures.

Operation (multiplication vs. division). In children who have received formal instruction on multiplication and division, it has generally been found that multiplication problems are easier than division problems (e.g., Christou & Philippou, 1999; Nesher, 1992). This may be a result of the school curriculum, in which multiplication commonly is formally introduced before division. However, in studies specifically focusing on children’s pre-instructional or early abilities in solving multiplicative problems, approximately equal difficulties were found for multiplication and division problems (e.g., Carpenter et al., 1993). Mulligan and Mitchelmore (1997), who also came to this result, explained this by arguing that young children intuitively connect multiplication and division and can use the same strategies for both.

Numbers involved. The kinds of numbers involved in a multiplicative problem can also affect the problem’s difficulty (e.g., Campbell & Graham, 1985). Such number effects can be explained by computational or retrieval efficiency. Because our focus is on revealing the extent to which children understand multiplication and division, rather than on their procedural efficiency, in this study we do not investigate effects of the numbers involved.

2.3.2 Gender

Boys have been found to outperform girls in mathematics from late secondary school on (e.g., Leahey & Guo, 2001) and sometimes already halfway through primary school (e.g., Hop, 2012; Penner & Paret, 2008). However, in kindergarten and first grade, gender differences in mathematics performance have generally been found to be negligible ($d < 0.10$; e.g., Aunio et al., 2006; Penner & Paret, 2008). Nevertheless, some studies showed that, although not differing in their overall mathematics performance, first-grade boys and girls do differ in their use of strategies: Girls tend to employ strategies using manipulatives more often than boys do, whereas boys more often use retrieval or derived-fact strategies (Carr & Davis, 2001; Fennema, Carpenter, Jacobs, Franke, & Levi, 1998). This means that a test setting in which no physical objects are available to be used as manipulatives may disadvantage girls more than boys. However, in a study by Ginsburg and Pappas (2004) focusing on pre-instructional mathematical knowledge, boys and girls did, in general, not differ in their strategy use.

2.3.3 Parental education

Research has consistently shown the importance of parental education as a predictor of children's achievement (see, e.g., Davis-Kean, 2005; Sirin, 2005). Davis-Kean (2005), for example, showed that higher-educated parents have higher expectations of their children's educational outcomes, resulting in more stimulating parenting behavior, which positively predicts child achievement. For the case of mathematics, parents' level of education may influence the time and attention parents spend working with their children on mathematics-related activities, and the complexity of these activities, with higher-educated parents presumably offering their children more and richer mathematical experiences (cf. Saxe, Guberman, & Gearhart, 1987; see also Suizzo & Stapleton, 2007). Given that activities offered by parents have been shown to predict children's early mathematics achievement (e.g., LeFevre et al., 2009), this may imply that students with higher-educated parents display more pre-instructional mathematical knowledge, as has been found, for example, by Entwisle and Alexander (1990). Moreover, studies focusing on effects of students' socio-economic status (SES), of which parental education is an important component (e.g., Sirin, 2005), showed that high SES children outperform low SES children on measures of pre-instructional mathematical knowledge (e.g., Driessen, 1997; Ginsburg & Pappas, 2004; Starkey & Klein, 2008). Similar to the above reasoning for parental education, the found SES differences may be explained by the possibility that higher SES children are offered richer mathematical experiences than are lower SES children (see Starkey & Klein, 2008).

2.3.4 Mathematics textbook

Several studies have shown that students' mathematics achievement is influenced by the textbook that is used in class (see, e.g., Törnroos, 2005). This also applies to the

Netherlands, where primary school mathematics teaching is highly guided by the mathematics textbook series that is used (see Mullis, Martin, Foy, & Arora, 2012). For example, Hop's (2012) analysis of Dutch survey data showed that, in many mathematics content domains including multiplicative reasoning, third graders who were taught with the textbook *Rekenrijk* outperformed children taught with other textbooks. In contrast, in Dutch sixth grade, it has been found that children using the textbook *De Wereld in Getallen* perform better than other children on most mathematics domains (Scheltens, Hemker, & Vermeulen, 2013).

Although in first grade, multiplication and division are not yet formally introduced in the Dutch mathematics curriculum (Van den Heuvel-Panhuizen, 2008), first-grade textbooks may differ in the attention they pay to informal, preparatory activities related to multiplication and division, such as counting in groups and repeated addition. Therefore, the textbook used in class may also play a role for children's multiplicative knowledge in first grade, before this domain is formally introduced.

2.4 Research questions

In earlier research on children's pre-instructional knowledge in the domain of multiplicative reasoning, several possible facets of this knowledge have not been investigated. In our study, we aimed to extend previous research by exploring how far children's pre-instructional multiplicative knowledge reaches when they are assessed in a relatively formal setting, with no teacher or experimenter present to provide help, and no physical objects available to support the finding of an answer. Furthermore, we intended to broaden the scope of the earlier studies by investigating children's performance on a wide range of multiplicative problems, including both problems presented in a context and bare number problems. Finally, we aimed for greater generalizability of findings by employing a large-scale study. In addition to our goal of mapping the pre-instructional multiplicative knowledge of first graders, we aimed to investigate what factors influence this knowledge.

Our research questions were as follows:

1. *To what extent are children, just before they start receiving formal instruction on multiplication and division, able to solve multiplicative problems in a relatively formal setting (without a teacher or experimenter sitting beside them, without physical objects provided, and including context problems as well as bare number problems)?*
2. *In what way is children's pre-instructional performance in solving multiplicative problems influenced by characteristics of the multiplicative problems offered to them?*

3. *In what way is children's pre-instructional performance in solving multiplicative problems influenced by children's gender, their parents' educational level, and the mathematics textbook that is used in class?*

Regarding the first research question, no hypothesis was specified. For this question our study should be considered exploratory in nature. With respect to the second research question, we expected context problems to be easier than bare number problems, especially when the context problems involve pictures offering opportunities to count. For semantic structure and operation, we did not specify a research hypothesis, since earlier studies were not conclusive on this. Finally, regarding our third research question, we hypothesized that students with higher-educated parents would have a higher level of pre-instructional knowledge. For gender and textbook, no specific hypotheses were formulated, again because of the inconclusive findings in literature.

3 Method

To answer our research questions, we carried out a large-scale survey in the Netherlands. Since in the Netherlands multiplication is formally introduced at the beginning of Grade 2 (Van den Heuvel-Panhuizen, 2008), we decided to assess children's knowledge of multiplicative reasoning at the end of Grade 1. The children were assessed by means of an online test on which they had to work individually. The method will be further described in the next sections.

3.1 Participants

3.1.1 Recruitment of schools

When we recruited the schools for this study, we aimed to get a sample of schools that would vary with respect to denomination, urbanization level, average parental education, and school size. For reasons of convenience, we first contacted schools in the center of the Netherlands. Later we sought schools in other parts of the Netherlands to complete our sample. The schools were mainly recruited by telephone (response rate ca. 15%, resulting in 43 participating schools). Additional schools were recruited by e-mail (response rate ca. 2%, resulting in two schools), by an advertisement on a mathematics games website and in a mathematics education newsletter (seven schools altogether), and at a mathematics teachers conference (one school).

3.1.2 The sample

In total, 53 first-grade classes from 53 different primary schools in the Netherlands were involved in our analysis. All classes were single-grade classes. Combined Grade 1-2 classes were excluded, since the first graders in such classes could have received some multiplication instruction when it was given to their second-grade classmates. The 53 participating classes together contained 1225 students. Of these students, 28 did not participate in the data collection because of illness, vacation, or organizational or technical problems in school, or because of expected leave from the class involved in the study due to moving to another school or having to repeat Grade 1. Furthermore, 16 students were excluded because they left 10 or more successive answers in the test blank, which was assumed to have been caused by technical problems (see section 3.4.3). Finally, four students were excluded because of missing student background data, and one student was excluded because of flaws in the data collection. Thus, a total of 1176 students were included in the analyses (580 boys, 596 girls). Their mean age was 7.2 years (SD = 0.4 years).

3.1.3 Representativity

To test the representativity of our sample of students (N = 1176), we compared it to the national data set of the Dutch primary school student population in the 2009-2010 school year (1,548,419 students; CBS, 2012), including information on students' gender and parental education (measured by "student weight", see section 3.2). For both gender and parental education, chi-square tests indicated that our student sample did not significantly differ from the population ($p > .05$), and thus can be considered representative with respect to these characteristics.

For testing the representativity of the sample of schools (N = 53), we used the 2009-2010 national data set of the Dutch population of primary schools (6,882 schools; OCW, 2011) for comparison. Chi-square and t tests indicated that our sample of schools can be assumed to be representative with respect to all the school characteristics we studied – urbanization level, average parental education, school size, and denomination ($p > .05$).

3.2 Background data

Background data of students, and information on the mathematics textbook used in school, were gathered through forms filled in by the teachers. As a measure of parental education we used the so-called student weight, which is a factor used in calculating student-based governmental funding of primary schools in the Netherlands and is determined by the parental level of education. Based on this factor, we distinguish three levels of parental education: Medium-high (at least one of the parents has completed secondary education); Low (both parents have completed vocational education as their highest level of education); and Very low (one parent has only attended primary school, and the other parent has completed primary school or vocational education).

3.3 Checking for previous instruction in multiplicative reasoning

Because the purpose of this study was to investigate students' pre-instructional multiplicative knowledge, the students involved in this study should not have been formally taught multiplication and division problems. Based on the nationally established teaching/learning trajectory and attainment targets for primary school mathematics (Van den Heuvel-Panhuizen, 2008) we can assume that in the Netherlands multiplication and division is not formally introduced until Grade 2. Yet, the first-grade mathematics curriculum may include informal activities preparatory to the formal instruction of multiplication and division.

To investigate the extent to which such preparatory multiplicative activities occur in first grade, and to check whether, indeed, no formal multiplication or division is included yet, we performed a textbook analysis of the textbooks used by the participating schools. Firstly, we examined the contents of the first-grade books of each of the seven textbook series by looking at the assignments per lesson (covering one school day). We found that a weighted average of 30.7% of the textbook lessons contain some preparatory multiplicative activities. In none of the textbooks multiplication and division are formally introduced. In one textbook (*Talrijk*, used by two schools in our sample), some exercises with the × symbol occur, but these are meant only for the more advanced students, and thus are not part of the standard curriculum.

To dig further into the contents of the textbooks, we did a more thorough analysis of the four textbooks that were most commonly used, both in our sample of schools and in the surveys by Hop (2012) and Scheltens et al. (2013): *Alles Telt*, *De Wereld in Getallen*, *Pluspunt*, and *Rekenrijk*. We distinguished five categories of (preparatory) multiplicative activities (see Table 1). For each category, we counted the exercises (both class-wise and individual exercises) occurring in the textbooks.[5] The categories were treated as mutually exclusive, that is, each exercise could be assigned to only one category. The frequencies we found are displayed in Table 1. We see that counting in groups quite often occurs, which can be seen as a preparatory activity that actually belongs to the domain of additive reasoning. Also multiplication and division problems in a context occur, often with the equal groups semantic structure. It should be noted, however, that these multiplicative context problems are presented irregularly (i.e., there is no clear learning trajectory these activities are part of), and many of these problems are presented as part of a class discussion. Next to counting in groups and context multiplication and division problems, doubling and halving problems occur, mostly presented in relation to addition and

[5] An exercise was defined as a problem for which an answer has to be computed. For example, a row of five doubling problems (2 + 2, 3 + 3, etc.) was counted as five exercises. In case exercises were presented together in a set, these exercises were only counted if the set clearly focused on informal multiplicative reasoning, that is, if half or more of the exercises in the set belonged to one of our categories.

subtraction. Groups-of problems and bare number multiplication and division problems only occur in *Alles Telt* (with the bare number problems presented as a context-less "2 groups of 2 is ___"), in the final lessons in the book (which means that they had probably not been treated yet at the time the test was administered).

Table 1
Textbook analysis of first-grade textbook series

		Number of preparatory multiplicative exercises					
Textbook series	*n*	Counting in groups[a]	Context multiplication and division[b]	Doubling and halving[c]	Bare number multiplication and division[d]	Groups-of[e]	Total
Alles Telt	6	37	101	280	6	44	468
WiG[f]	10	88	108	48	0	0	244
Pluspunt	20	36	70	49	0	0	155
Rekenrijk	11	124	75	154	0	0	353

[a]Exercises in which one has to count with jumps (e.g., 5-10-15-...), or in which a number of equal groups have to be added, with all the groups displayed. [b]Multiplication and division problems presented in a context, not belonging to the Counting in groups category. [c]Exercises in which one is asked to double or halve a certain number, either with or without a context. Also included are addition and subtraction exercises focusing on doubles or halves. [d]Multiplication and division problems without a context. [e]Exercises in which a groups-structure in a picture has to be described ("__ groups of __" or "__ groups of __ is __"). [f]*De Wereld in Getallen*.

When we compare the four textbooks in Table 1, we see marked differences. Looking at the total number of exercises, we see that *Alles Telt* includes over 3 times as many preparatory multiplicative exercises as does *Pluspunt*, with the other textbooks lying in between. Furthermore, how much emphasis is put on the different categories differs greatly between textbooks, with *De Wereld in Getallen* and *Rekenrijk* paying more attention to counting in groups than do the other textbooks, *Alles Telt* and *De Wereld in Getallen* including more context multiplication and division problems, and *Alles Telt* and *Rekenrijk* having a much larger focus on doubling and halving.

3.4 Test for assessing students' pre-instructional multiplicative knowledge

Students' pre-instructional knowledge of multiplicative reasoning was measured by an online test consisting of 28 multiplicative items. The use of an online test facilitated our large scale data collection and ensured a standardized test procedure. Our test can be considered a relatively formal test setting, as children had to work on their own without an interviewer or teacher sitting next to them. Furthermore, as explained before, physical objects were not provided to the students as aids in solving the test problems. However, the students were not forbidden to use their fingers as manipulatives (see, e.g., Carr & Davis, 2001). Because in everyday school practice, students have their fingers at their disposal as well, even in formal test settings, using fingers can be considered to be part of formal settings. As reported by Ibarra & Lindvall (1982), although the use of fingers can partly compensate for the absence of physical objects, the unavailability of physical objects still makes problems more difficult to solve, differentiating our study from earlier studies in which physical objects were provided.

3.4.1 Characteristics of the multiplicative items

The 28 multiplicative items of the test varied according to several characteristics, including problem format, semantic structure, operation, and countability level. Table 2 lists the multiplicative items and their characteristics.

Problem format. A first characteristic that was taken into account when constructing the items was the problem format. Of the 28 test items, 18 were *context problems* (see Figure 1) and 10 were *bare number problems* (see Figure 2). In addition, there were four *groups-of problems* (see Figure 3), which were meant to specifically assess students' understanding of the groups-of structure typical of multiplicative situations. For the context and groups-of problems, we chose objects or situations familiar to young children, such as dice, sticker sheets, money, and rabbits. The bare number problems included six times problems (e.g., Figure 2a, 2b) and four doubling problems (Figure 2c). Because in Grade 1 the × symbol is not yet known, in the times problems we used the word *times*.

Semantic structure. The context problems varied by their semantic structure. The test included equal groups problems (e.g., Figure 1a and 1b), rate problems (e.g., Figure 1c), and rectangular array problems (e.g., Figure 1d). Since we aimed to measure the prior multiplicative knowledge that is available to build on when formal multiplication and division are introduced, we decided the majority of the problems to be of the semantic structure that is most common in early formal instruction of multiplication and division, which is equal groups.

Operation. A further problem characteristic was the operation involved. We included both multiplication and division problems in the test. Since in the Netherlands, multiplication is formally introduced earlier than division – which is dealt with only from Grade 3 (Van den Heuvel-Panhuizen, 2008) – the test contained more multiplication problems than division problems. Bare number division problems were not included, since in the Dutch language there is no concise everyday-language translation of the ÷ symbol.

Table 2
Multiplicative items and their characteristics

Item no.	Description	Problem format[a]	Semantic structure[b]	Operation[c]	Countability level[d]
1	3 dice with 5 dots, how many dots together?	C	Eq	×	3
2	5 dice with 5 dots, how many dots together?	C	Eq	×	3
3	10 times a die with 5 dots, how many dots together?	C	Eq	×	1
4	1 bear costs 10 euro, how much do 4 bears cost?	C	Rate	×	0
5	1 ball costs 2 euro, how much do 6 balls cost?	C	Rate	×	0
6	1 book costs 5 euro, how many books can you buy for 20 euro?	C	Rate	÷	0
7	Tiled square (3×6) with car covering part of the tiles, how many tiles on the whole square?	C	Rect	×	3
8	Unfinished jigsaw puzzle (3×5), how many pieces does the puzzle have when it is finished?	C	Rect	×	2
9	3 boxes of 4 cookies, how many cookies together?	C	Eq	×	2
10	3 sheets with 10 stickers, how many stickers together?	C	Eq	×	2
11	4 sheets with 4 stickers, how many stickers together?	C	Eq	×	2
12	10 mittens. For how many children?	C	Eq	÷	3
13	8 carrots and 4 rabbits. How many carrots does each rabbit get?	C	Eq	÷	1
14	Each man carries 2 buckets. 20 buckets. How many men are needed to carry them?	C	Eq	÷	1
15	2 times 1 is	BN	-	×	0
16	5 times 2 is	BN	-	×	0
17	2 times 4 is	BN	-	×	0
18	10 times 2 is	BN	-	×	0
19	Which problem belongs to 30? (2 times 10, 6 times 10, or 3 times 10)	BN	-	×	0
20	Which problem belongs to 20? (4 times 5, 7 times 5, or 3 times 5)	BN	-	×	0
21	Make it double: 3 →	BN	-	×	0
22	Make it double: 4 →	BN	-	×	0
23	Make it double: 6 →	BN	-	×	0
24	Make it double: 10 →	BN	-	×	0
25	Cherries: 3 groups of ? (answer is 2)	GO	-	-	-
26	Cherries: ? groups of 2 (answer is 4)	GO	-	-	-
27	Cherries: ? groups of ? (answer is 4 groups of 3)	GO	-	-	-
28	Bananas: ? groups of ? (answer is 3 groups of 4)	GO	-	-	-

[a]C = context problem; BN = bare number problem; GO = groups-of problem. [b]Eq = equal groups; Rate = rate; Rect = rectangular array. [c]× = multiplication; ÷ = division. [d]0 = no terms countable; 1 = one term countable; 2 = both terms countable; 3 = both terms and solution countable.

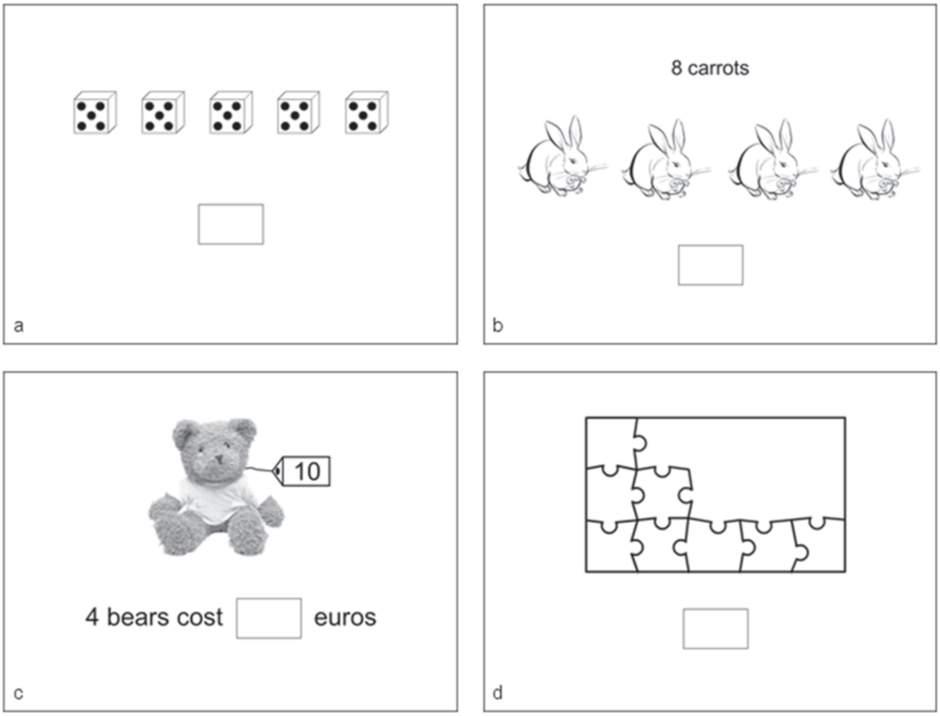


Figure 1. Examples of context problems. a. (Item 2) "How many points together?" b. (Item 12) "Eight carrots. How many carrots does each rabbit get?" c. (Item 4) "How much do four bears cost together?" d. (Item 8) "How many pieces will the puzzle have when it is finished?"

Countability level. A final problem characteristic taken into account was the problems' countability level, that is, the extent to which the picture presented in the problem could be used to find a solution by counting. We distinguished four levels of countability: Level 0, no terms countable (e.g., Figure 1c); Level 1, one term countable (e.g., Figure 1b, here the rabbits are countable, but the carrots are not); Level 2, both terms countable (e.g., Figure 1d, here both the pieces in the length and the width of the puzzle can be counted); and Level 3, both terms and the outcome countable (e.g., Figure 1a, here the dice can be counted, the dots on a dice, and the total number of dots on all dice). The bare number problems, which by definition do not include pictures, have the lowest countability level: no numbers countable.

Other characteristics. The numbers used in the test problems were 1, 2, 3, 4, 5, 6 and 10. All but two test items were constructed-response items in which the students had to type their answers. The two exceptions were multiple choice items (e.g., Figure 2b) which could be answered by clicking on an answer.

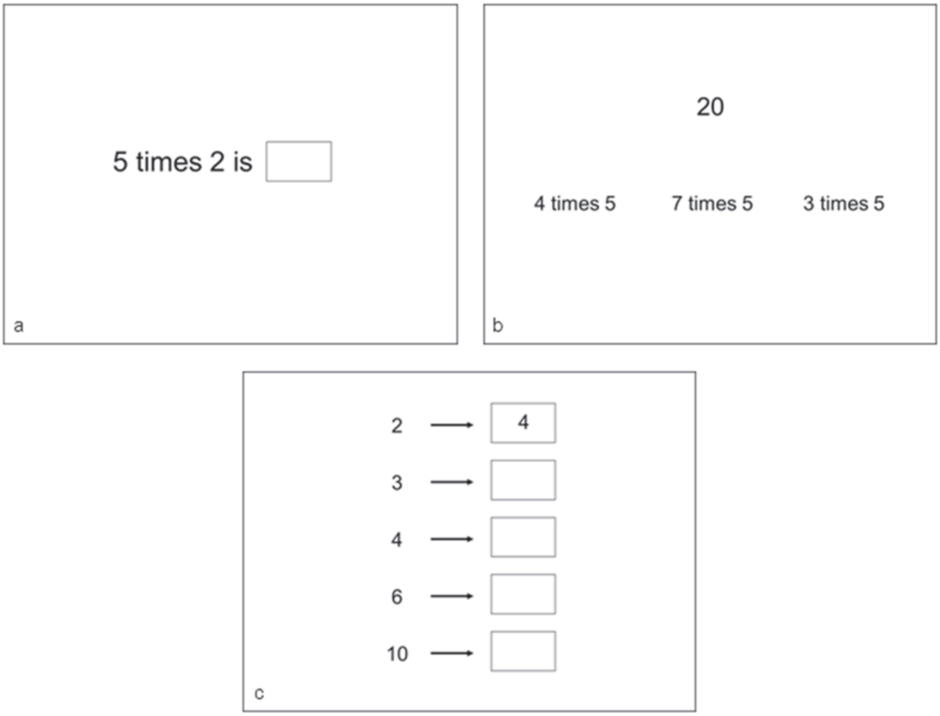


Figure 2. Examples of bare number problems. a. (Item 16) "Five times two is…" b. (Item 20) "What belongs to twenty? Click on the correct problem." c. (Items 21-24) "Make it double. Each time fill in the answer."

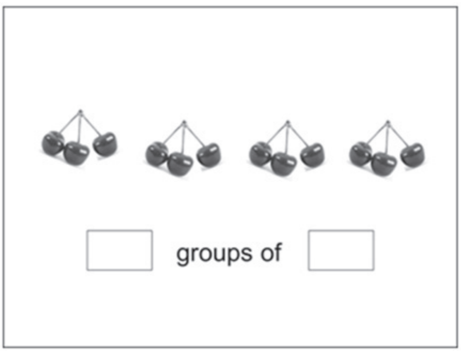


Figure 3. Example of groups-of problem (Item 27). "What sentence fits the picture?"

3.4.2 Test design

To control for order effects, four different versions of the test were constructed. For this purpose the multiplicative items were organized into four item clusters (A, B, C, and D) of seven items each. In the different test versions, the clusters were presented in different orders: ABCD, CDAB, BADC, and DCBA, respectively. The different versions were randomly assigned to the students. In addition to the multiplicative items, each cluster contained three "distractor" items on spatial orientation, which were included to conceal the test's focus on multiplicative reasoning.

3.4.3 Administration of the test

The test was administered at the end of the Grade 1 school year, in June/July 2010. The test was embedded in an online environment, called Digital Mathematics Environment (DME),[6] in which the students' answers to the test items were recorded. The students could reach the test by accessing the test website and logging in with their personal account.

The online test started with a short general instruction, which was read out by the computer. After that, each item was individually displayed on the screen (except for the doubling problems, see Figure 3c). The accompanying question was read aloud by the computer. By clicking on a loudspeaker button, the student could repeat the spoken text. By clicking on an arrow, the child could navigate to the next item or to a previous item.

The teachers were asked to help the students to get online, but not provide support in answering the problems. Each student worked individually on a computer with headphones. In one school, headphones were not available and students used laptop speakers instead. The test was administered as a whole, without breaks in between. The duration of the test was, on average, approximately 20 minutes.

Due to a few technical problems not all responses were recorded by the DME. Therefore, in case 10 or more successive answers were missing for a student, we asked the student to make the test again. For these students, we replaced the missing answers in the first test administration by the corresponding answers from the second administration. In total, 27 students took the test again because more than 10, but not all, successive answers were missing in the first test administration. In addition, 21 students took the test again because none of their answers were recorded.

[6] The Digital Mathematics Environment (DME) has been developed by our colleague Peter Boon at the Freudenthal Institute of Utrecht University.

3.4.4 Data processing

Since the text boxes in which the students had to type their answers accepted all kinds of input, not all responses were in the form of a number. Some students accidentally typed spaces or punctuation marks before, after, or in between the digits, for example “4’0” or “ 40” instead of “40”. Other students typed number words instead of numbers, for example “vier” (Dutch for *four*) instead of “4”. Another typing mistake that occurred was typing the letter *o* instead of the digit *0*, resulting in answers like “1o” instead of “10”. Furthermore, for the items in which more than one answer box had to be filled in on one screen (the doubling problems and two of the groups-of problems, see Figure 2c and Figure 3), some students put more than one answer in a single box. All these typing mistakes were corrected. For 0.59% of the total of the 32,928 cases (1176 students who gave 28 item answers each) an initially incorrect answer was changed to a correct answer.

3.4.5 Psychometric properties of the test

The reliability of the test consisting of 28 multiplicative items was sufficiently high (Cronbach’s alpha of .89). The alpha_{10} reliability, which is based on a hypothetical 10-item-length scale, was .75. The corrected item-total discriminations ranged from .21 to .57 (*M* = .44, *SD* = .10), which indicates that all individual item scores were sufficiently related to the total test score (DeVellis, 2012).

To examine the dimensional structure of the test, we performed an exploratory factor analysis for dichotomous data (Revelle, 2012). Using the scree test for determining the number of factors, we identified a four-dimensional factor structure. The four factors can be interpreted as follows: (a) *context problems* (14 items), (b) *bare number times problems* (six items), (c) *bare number doubling problems* (four items), and (d) *groups-of problems* (four items).This structure coincides with the item characteristic problem format (see Table 2), except that the bare number category is split into two factors (bare number times problems, items 15-20; bare number doubling problems, items 21-24).

Although the scree test pointed towards a four-factor structure, there was also evidence that the performance assessed by our multiplicative ability test can be represented by a unidimensional scoring. In the factor analysis it appeared that about 37.7% of the total variance can be attributed to the first dimension, and there was a large ratio of 4.00 of the first and second eigenvalue. The reliability estimate calculated by the omega total (Revelle, 2012) based on a four-dimensional model (omega = .91) was only slightly larger than the omega total based on the one-dimensional model (omega = .88).[7] We also performed a Schmid Leiman transformation (see, e.g., Revelle, 2012), which resulted in a model with a

[7] The reliability estimate was derived for the sum score based on the model for dichotomous data (Green & Yang, 2009).

general factor and four uncorrelated specific factors. In this model, the variance explained by the general factor was .71. The above findings allowed us, in addition to looking at the four factors separately, to consider the test as a whole and to use the unidimensional scoring in our analyses.

4 Results

4.1 Students' performance

On average, the students ($N = 1176$) answered more than half of the total of 28 items correctly. We found a mean proportion correct of .58 ($SD = .23$) with 2 students (0.2%) having no answers correct and 19 students (1.6%) having all items correct. The distribution of total scores was only slightly skewed (skewness = -.12, $SE = .07$, $p > .05$), indicating that the assumption of normality for the use of parametric tests was not violated.

When zooming in on the four groups of items identified through factor analysis (see section 3.4.5), we found higher scores on context problems ($M = .63$, $SD = .23$) and bare number doubling problems ($M = .63$, $SD = .44$), than on bare number times problems ($M = .52$, $SD = .35$) and groups-of problems ($M = .47$, $SD = .38$).

To get a better view of the pre-instructional knowledge displayed by students at different levels of performance, we applied a Rasch model, which places student abilities and item difficulties on a common scale. This Rasch analysis was performed using the ConQuest software (Wu, Adams, Wilson, & Haldane, 2007). We classified the students' Rasch scores (the WLE person parameter estimates; Warm, 1989) into four ability levels, based on ability score quartiles. For each level we examined which items the students could solve. Figure 4 shows the difficulty levels of all test items, indicating which items belong to each level. If a group of students is located at a certain ability level, this means that this group of students has "mastered" all items from the previous levels, meaning that these items can be solved with a probability of at least 62.5% (cf. OECD, 2009). Depending on the exact position of groups of students within a level (e.g., whether students belong to the more able or to the less able students of this level), some of the items of that particular level can also be considered to be mastered.

The items belonging to each level are described in the right part of Figure 4. At the lowest ability level, Level 1 ($n = 327$), students were on average not yet able to solve any of the test items with a probability of at least 62.5%. The more competent students at this level had, however, already mastered some of the problems at Level 1: small number multiplication problems of the equal groups semantic structure, with both numbers countable, and a division problem in a sharing context. Students at Level 2 ($n = 269$) had mastered all Level 1 problems. More competent students at this level could already solve

	Item numbers	Level	Problem descriptions: Context problems	Doubling problems	Times problems	Groups-of problems
100%	3 28, 27, 20 8, 14	4 (22-28 items correct)	- multiplication problem with the equal groups semantic structure, with one term countable, with large numbers (product 50) - division problem with the equal groups semantic structure, with one term countable - multiplication problem with the rectangular array semantic structure, with both terms countable	-	MC times problem with factor 5	__ groups of __
73.8%	18 16, 25, 6 5 17, 15	3 (17-21 items correct)	- multiplication and division problems with the rate semantic structure, with factors other than 10 (no terms countable)	-	Constructed-response times problems (product up to 20)	X groups of __
50.7%	24, 23, 26 7, 12 22, 4, 19 21	2 (12-16 items correct)	- multiplication problem with the rate semantic structure, with factor 10 (no terms countable) - multiplication problem with the rectangular array semantic structure, with countable solution - division problem with the equal groups semantic structure, with countable solution	doubling problems up to double 10	MC times problem with factor 10	__ groups of X
27.8%	10, 13 2, 11 9 1	1 (0-11 items correct)	- multiplication problems with the equal groups semantic structure, with at least both terms countable, with small numbers (product up to 25) - division problem with the equal groups semantic structure, in a sharing context (one term countable)	-	-	-
0%						

Figure 4. Distribution of items over Rasch score ability levels (quartiles) and descriptions of items at each level. See Table 2 for descriptions of individual items. Because there are only 29 possible test scores (0-28 items correct), the actual percentage of students in each quartile deviates somewhat from 25%. MC = multiple choice.

some of the Level 2 items – easy rate and rectangular array multiplication problems, an equal groups division problem with countable solution, doubling problems, a multiple choice bare number multiplication problem, and a groups-of problem of the form "__ groups of X". Students at Level 3 (n = 272) had mastered all problems belonging to Level 1 and Level 2. Some of them could also solve some of the items at Level 3: more difficult rate multiplication and division problems, constructed-response times problems, and a groups-of problem of the form "X groups of __". Finally, students at the highest ability level, Level 4 (n = 308), could solve all items at the previous levels, meaning that they could solve almost all equal groups multiplication and division problems, all rate multiplication and division problems, a rectangular array multiplication problem with countable solution, all constructed-response times problems and doubling problems, and part of the groups-of problems. Many of the Level 4 students had not yet mastered the most difficult items, including an equal groups multiplication problem with large numbers and only one term countable, an equal groups division problem with one term countable, a rectangular array multiplication problem with both terms (but not the solution) countable, a multiple choice times problem, and the groups-of problems of the form "__ groups of __".

4.2 Influence of problem characteristics

To study the influence of problem characteristics on problem difficulty, we looked at the different problem characteristics as displayed in Table 2. For the characteristic problem format, instead of a three-category variable (as in Table 2), we employed a four-category variable, based on the test's factor structure (see section 3.4.5). As a first analysis, we studied the marginal effects of the problem characteristics, that is, the raw effects of the characteristics of the problems on their difficulty level, without taking into account the other problem characteristics or student and class characteristics. We performed a student-level within-subjects Wald chi-square test ($N = 1176$) for each problem characteristic, comparing the mean scores for the different values of the characteristic. To account for the nested data structure (students within classes), cluster-robust standard errors were employed (see Angrist & Pischke, 2009), using the Mplus software (Muthén & Muthén, 1998-2010). As an effect size measure we used the ω^2 estimate of proportion of explained variance (see, e.g., Grissom & Kim, 2012), for which a value of .010 can be interpreted as a small effect, .059 as a medium sized effect, and .138 as a large effect (e.g., Kirk, 1996). As is shown in Table 3, results were significant for all problem characteristics. Looking at the ω^2 effect sizes, we see non-trivial effect sizes (i.e. $\omega^2 \geq .010$) for problem format ($\chi^2(3) = 111.73$, $p < .001$, $\omega^2 = .038$), semantic structure ($\chi^2(2) = 420.40$, $p < .001$, $\omega^2 = .041$), and countability level ($\chi^2(3) = 544.76$, $p < .001$, $\omega^2 = .090$). The effect of operation can be considered negligible ($\chi^2(1) = 20.59$, $p < .001$, $\omega^2 = .007$).

4.3 Influence of student- and class-related characteristics

As we did for the problem characteristics, for the student- and class-related characteristics gender, parental education, and mathematics textbook we first analyzed the marginal effects, i.e., the effects without taking other factors into account. For gender and parental education, we performed a between-subjects ANOVA (total $N = 1176$), with the test score as the dependent variable and the possible values of the student characteristic as the between-subject factors. Again, cluster-robust standard errors were employed to account for the nested data structure. The results of the ANOVA are presented in Table 4. We found that parental education significantly predicted students' test score, with students with lower-educated parents performing less well ($\chi^2 = 35.05$, $p < .001$, $\omega^2 = .033$). Gender did not have a significant effect on test score ($p > .05$).

For the class characteristic mathematics textbook, the marginal effect was investigated using a class-level between-subjects ANOVA ($N = 53$ classes), with the class mean test score as the dependent variable. As is shown in Table 5, we found no effect of mathematics textbook ($p > .05$).

Table 3

Marginal effects of problem characteristic on mean proportion correct

Problem characteristic	#items[a]	*M*	*SD*	Wald χ^2	*df*	ω^2
Problem format	28			111.73***	3	.038
Context	14	.63	.23			
Bare number times	6	.52	.35			
Bare number doubling	4	.63	.44			
Groups-of	4	.47	.38			
Semantic structure	14			420.40***	2	.041
Equal groups	9	.68	.23			
Rectangular array	2	.53	.37			
Rate	3	.54	.37			
Operation	24			20.59***	1	.007
Multiplication	20	.61	.24			
Division	4	.56	.32			
Countability level[b]	24			544.76***	3	.090
0	13	.56	.30			
1	3	.50	.32			
2	4	.69	.27			
3	4	.73	.26			

Note. $N = 1176$. Cluster-robust standard errors were employed. ω^2 is based on regular within-subjects ANOVA results (see Grissom & Kim, 2012).

[a]For each item characteristic, the total number of items for which this characteristic is applicable is given, as well as the number of items in each category of the characteristic. [b]0 = no terms countable; 1 = one of the terms countable; 2 = both terms countable; solution not countable; 3 = both terms and solution countable.

*** $p < .001$.

Table 4
Marginal effects of gender and parental education on test scores

Student characteristic	*n*	*M*	*SD*	χ^2	*df*	ω^2
Gender				0.75	1	0[a]
Male	580	.58	.24			
Female	596	.59	.23			
Parental education				35.05***	2	.033
Medium-high	1009	.60	.23			
Low	104	.49	.23			
Very low	63	.45	.22			

Note. Cluster-robust standard errors were employed. χ^2 = Satorra-Bentler scaled difference chi-square statistic, testing the goodness-of-fit difference between the model with and without the predictor (Satorra & Bentler, 2001). ω^2 is based on regular ANOVA results.
[a]The value was negative and therefore set to 0.
*** $p < .001$.

Table 5
Marginal effect of mathematics textbook on class mean test score

Mathematics textbook	*n*[a]	*M*	*SD*	*F*	*df1*	*df2*	ω^2
Alles Telt	6	.57	.11	0.27	4	48	0[c]
De Wereld in Getallen	10	.61	.12				
Pluspunt	20	.58	.08				
Rekenrijk	11	.56	.09				
Other[b]	6	.59	.11				

Note. The effect was not significant.
[a]Number of classes. [b]*Talrijk* (2 classes), *Wis en Reken* (2 classes), and *Wizwijs* (2 classes). [c]The value was negative and therefore set to 0.

4.4 Cross-classified multilevel analysis

In the analyses presented in the previous sections, the effect of each item (problem), student, and class characteristic was investigated separately, without accounting for effects of the other characteristics. To study all characteristics simultaneously, we performed a logistic multilevel regression analysis with the dichotomous item score (correct or incorrect) as the dependent variable. This item score is nested within the item level on the one hand, and on the other hand within the student level, which is in turn nested within the class level. This cross-classified logistic multilevel regression analysis was performed using the MCMCglmm package in R (Hadfield, 2010). Next to the item, student, and class variables mentioned earlier, we controlled for some additional variables at the student and class level: students' age (grade-appropriate age vs. delayed; cf. Hop, 2012), school average parental education (obtained from OCW, 2011; cf. Driessen, 1997, 2002), and the number of weeks of Grade 1 education the class had received before the test was administered.[8]

4.4.1 Model specification

We specified five cross-classified multilevel models (see Table 6), following a hierarchical inclusion of predictor groups, which were compared using variance explanation measures (R^2). From the empty model (Model 0), it can be concluded that on the student side, about 15% of the total student variance can be attributed to variance between classes (Intra-class correlation = class variance / (class variance + student variance) = .15). In Model 1, item predictors explained 22% of the item level variance ($R^2 = .22$). Introducing student level predictors in Model 2a explained 2% of the student level and 15% of the class level variance. The explained class level variance may be due to the fact that students with different demographic backgrounds tend to attend different schools, and thus different classes. When class level predictors were added in Model 2b, the proportion of explained class level variance increased to 32%. Nearly the same variance proportions of the item level in Model 1 and the student and class level in Model 2b were observed in Model 3, which contained all predictors.

To be able to take all predictors into account, we decided to continue with Model 3. The regression coefficients of the predictors in Model 3 are displayed in Table 7. For the continuous predictors, the regression coefficients represent the contrast to predictor value 0. For the categorical predictors, which were dummy coded, the regression coefficients contrast the predictor's categories to the reference category. For predictors with more than two categories, a Wald *F* test was performed to test the overall effect of the predictor (see, e.g., Lehmann & Romano, 2005). We did not add effect size measures to the model, because measures that can be easily calculated and interpreted in logistic multilevel models seem to be rare in literature (for example, the often cited book by Hox, 2010, is almost silent about this topic).

[8] Since the participating classes varied somewhat in the date on which the test was administered (with a range of 4½ weeks), and because in the Netherlands, schools vary in the date at which the school year starts (with a range of 3 weeks), classes differed in the amount of Grade 1 education they received before the test was taken (ranging from 32 to 38 weeks).

Table 6
Comparison of cross-classified multilevel models

	Model 0: No predictors			Model 1: Item level predictors			Model 2a: Student level predictors			Model 2b: Student and class level predictors			Model 3: All predictors		
Random effects	Var	*SE*	%Var	Var	*SE*	R^2	Var	*SE*	R^2	Var	*SE*	R^2	Var	*SE*	R^2
Item	0.78	0.22	31.7	0.62	0.22	.21	0.77	0.22	.00	0.78	0.22	.00	0.63	0.23	.19
Student	1.43	0.08	58.4	1.42	0.08	.01	1.40	0.07	.02	1.40	0.07	.02	1.41	0.07	.01
Class	0.24	0.06	9.9	0.25	0.07	0[a]	0.21	0.06	.15	0.17	0.05	.32	0.17	0.06	.31
Total	2.45		100	2.28		.07	2.38		.03	2.35		.04	2.21		.10

Note. Var = variance; R^2 = proportion of explained variance.
[a]The value was negative and therefore set to 0.

Table 7
Regression coefficients of cross-classified multilevel Model 3

Predictor	(dummy) variable	Regression coefficient *B*	*SE*	*t*	Wald test *F*	*df1*	*df2*
Item level							
Problem format[a]	Prob_bare	-0.17	0.58	-0.30	3.00	3	21
	Prob_double	0.48	0.63	0.77			
	Prob_groupsof	-0.45	0.62	-0.73			
Operation[b]	Division	-0.16	0.55	-0.29			
Countability level[c]	Count_level1	-0.11	0.66	-0.17	5.29**	3	21
	Count_level2	0.85	0.62	1.38			
	Count_level3	1.11	0.59	1.88			
Student level							
Gender[d]	Female	0.04	0.08	0.54			
Parental education[e]	ParEd_low	-0.36	0.14	-2.54*	8.84***	2	1175
	ParEd_verylow	-0.39	0.19	-2.03*			
Age[f]	AgeDelayed	-0.41	0.15	-2.76**			
Class level							
Mathematics textbook[g]	Book_AllesTelt	0.00	0.26	0.00	5.36**	4	40
	Book_Pluspunt	-0.30	0.20	-1.53			
	Book_Rekenrijk	-0.36	0.22	-1.64			
	Book_Other	0.06	0.25	0.26			
School average parental education	AvgParEd	-2.17	0.67	-3.24**			
Weeks of Grade 1 education	EdWeeks	0.02	0.06	0.36			

[a]Reference category (Ref): Context problems. [b]Ref: Multiplication. [c]Ref: Level 0 (no terms countable). [d]Ref: Male. [e]Ref: Medium-high. [f]Ref: Grade-appropriate age. [g]Ref: *De Wereld in Getallen*.
* $p < .05$. ** $p < .01$. *** $p < .001$.

4.4.2 Item level predictors

At the item level, we included the predictors problem format, operation, and countability level. Semantic structure was not included, because this variable only applies to half (14) of the items. As is shown in Table 7, none of the item level regression coefficients were significantly different from 0 ($p > .05$). However, in the Wald test we found a significant value for countability level ($F(3, 21) = 5.29$, $p = .007$). Thus, for countability level, although no significant effects were found when the dummy categories were contrasted to the reference category, effects were significant when all contrasts between categories were taken together. An analysis of contrasts revealed a marginally significant contrast of countability level 3 as compared to countability level 1 ($B = 1.22$, $t = 1.88$, $p = .059$). Also the regression coefficient of countability level 3, contrasting countability level 3 with 0, was close to significance ($B = 1.11$, $t = 1.88$, $p = .060$). Thus, we see that countability did play a role in predicting item difficulty, with items offering the opportunity to count up to the solution tending to be easier than problems in which no or only one factor can be counted. For problem format, the absence of a significant effect may be caused by a confounding with countability level (only context problems can have a countability level other than 0). To check for this possible confounding, we specified an alternative cross-classified multilevel model with all predictors except countability level included. In this model we found a significant Wald test F value for problem format ($F(3, 21) = 6.85$, $p = .002$) indicating that, when countability level was excluded from the model, problem format did affect problem difficulty. An analysis of contrasts revealed a significant contrast for groups-of problems vs. context problems, with groups-of problems being more difficult ($B = -1.12$, $t = -2.22$, $p = .026$), and a marginally significant contrast for times problems vs. context problems, with times problems tending to be more difficult ($B = -0.84$, $t = -1.88$, $p = .059$).

4.4.3 Student and class level predictors

At the student level, we found a significant effect of parental education ($F(2, 1175) = 8.84$, $p < .001$). As expected, students with lower-educated parents performed worse than students with higher-educated parents. In an analysis of contrasts, significant differences were found for the comparison of medium-high parental education with low parental education ($B = -0.36$, $t = -2.54$, $p = .011$) and with very low parental education ($B = -0.39$, $t = -2.03$, $p = .043$). For gender, we did not find a significant effect ($p > .05$).

At the class level, the Wald test showed a significant effect for mathematics textbook ($F(4, 40) = 5.36$, $p = .001$). From the regression coefficients in Table 7 it seems that the students in classes using the textbook *Pluspunt* or *Rekenrijk* performed somewhat worse than students using the other textbooks, but an analysis of contrasts did not reveal significant pair-wise differences between textbooks (comparisons of *Pluspunt* and *Rekenrijk* with the other textbooks yielded B values ranging from -0.42 to -0.30, with t values from -1.64 to -1.18 and p values from .108 to .238). Taken together, the mathematics textbook used in class influenced students' multiplicative reasoning performance at the end of Grade 1, but we do not have clear evidence of which textbooks were related to better performance. The fact that we did find an effect of textbook in our multilevel analysis but not in our marginal class-level ANOVA (Table 5) may be explained by classes having different textbooks also differing with respect to other characteristics, which in the analysis of marginal effects may have cancelled out the textbook effects. When we compared classes with different textbooks, it appeared that this was indeed the case: Classes using *Pluspunt* or *Rekenrijk* had a higher school average parental education (measured by student weight, with lower values indicating higher levels of parental education; $M = 0.06$, $SD = 0.08$) than classes using other textbooks ($M = 0.14$, $SD = 0.16$, $t(51) = -2.49$, $p = .016$).

5 Conclusions and discussion

5.1 Children's pre-instructional multiplicative knowledge

The current study reveals that first-grade children can already solve a considerable amount of multiplicative problems. On average, the children correctly answered more than half (58%) of the test items. Thus, even when first graders are assessed in a relatively formal online test setting, with no teacher or experimenter present to provide support and no physical objects available as an aid in solving the problems, they show a rather high performance in solving multiplicative problems, despite the fact that multiplicative problems have not yet been formally introduced to them. This extends the findings of Anghileri (1989), Kouba (1989), Mulligan and Mitchelmore (1997), and others, who found young children being able to solve several multiplicative problems in a one-to-one interview setting, with the help of physical objects. Interestingly, in our more formal setting, we found a higher percentage correct than did Kouba (29%, first graders) and Mulligan and Mitchelmore (48%, second graders). In contrast to earlier studies, which were rather small-scale, our study was a large-scale study, which made our findings more robust.

Percentages correct varied for the different problem formats in our test. Next to context problems (63% correct), also bare number doubling problems (63%) and bare number times problems (52%) could be solved by more than half of the children. This further adds to previous findings (e.g., Anghileri, 1989; Kouba, 1989; Mulligan & Mitchelmore, 1997) by showing that in addition to the ability to solve context problems, pre-instructional multiplicative knowledge for many children also includes the ability to solve bare number multiplication problems in the form of doubling or in the form of problems in which the × symbol is replaced by the word *times*. The groups-of problems in our study appeared the hardest (47% correct). Thus, although quite a number of children can solve multiplicative problems, this does not necessarily mean that they have an explicit understanding of the groups structure typical of multiplicative situations.

An analysis of performance levels revealed large variations in children's performance. The lowest performing 25% of the children were about to be able to solve small number multiplication problems of the equal groups semantic structure with both terms countable, and a division problem in a sharing context, whereas the highest performing 25% could, generally, solve almost all equal groups multiplication and division problems, all rate problems, a rectangular array multiplication problem with countable solution, all constructed-response times problems and doubling problems, and part of the groups-of problems. Thus, although on average first graders are quite able in solving simple multiplicative problems, large differences exist between individual children.

5.2 Problem characteristics influencing students' achievement

Our hypothesis that context problems would be easier to solve than bare number problems was only supported by a marginally significant difference between these two problem formats (when countability level was excluded from the model). This is in line with earlier research findings (e.g., Koedinger & Nathan, 2004; Van den Heuvel-Panhuizen, 2005). We found no difference in difficulty between bare number doubling problems and context problems, indicating that the doubling problems were relatively easy, possibly because of their close relation to addition.

As expected, we found that context problems were especially easy when they involved a picture with countable objects: multiplicative problems offering pictures with more opportunities for counting were easier to solve. This indicates that, although pictured objects are more abstract than real-world physical objects (e.g., Martin & Schwartz, 2005; Moyer et al., 2002), the former can also assist in solving problems. This effect of countability mediated the effect of problem format: when countability was included in the multilevel model, the effect of problem format was no longer significant.

For operation, no significant effect was found. This indicates that, when other problem, student, and class characteristics are controlled for, multiplication and division problems are equally difficult for children who have not yet received formal instruction on multiplicative reasoning. This finding confirms earlier findings by Mulligan and Mitchelmore (1997), and Carpenter et al. (1993).

Semantic structure was not included in the multilevel model because it applied to only half of the items. However, in the analysis of marginal effects we found a significant effect, with equal groups problems being easier than rate and rectangular array problems. This result is in line with Christou and Philippou's (1999) and Nesher's (1992) findings that equal groups problems are easiest for children who have received formal instruction on multiplicative reasoning. However, our finding contradicts Mulligan and Mitchelmore's (1997) conclusion that before formal instruction of multiplicative reasoning, there is no difference in difficulty between the semantic structures equal groups, rate, and rectangular array. The finding in our study that equal groups problems were easier than rate and rectangular array problems may partly be explained by the fact that, as we found out in our textbook analysis (section 3.3), the informal and preparatory multiplicative activities that occur in the Dutch first-grade curriculum primarily focus on equal groups situations.

5.3 Student- and class-related characteristics influencing students' achievement

We did not find a difference between boys' and girls' pre-instructional multiplicative knowledge, which is in accordance with earlier studies that showed no gender differences in mathematics performance before Grade 3 (e.g., Aunio et al., 2006; Penner & Paret, 2008). Although the absence of physical objects in our test setting could have disadvantaged girls more than boys – because of girls' preferences for strategies using manipulatives (Carr & Davis, 2001; Fennema et al., 1998) – our results do not confirm this reasoning. This is in line with Ginsburg and Pappas' (2004) finding that in the case of pre-instructional knowledge, boys and girls overall do not differ in their strategy use. Possibly, prior to formal introduction of a mathematics domain, girls and boys both tend to use manipulatives a lot, which would make the absence of physical objects equally disadvantageous for both.

With respect to the educational level of the students' parents, our results show that, as we expected, students with higher-educated parents had higher levels of pre-instructional multiplicative knowledge. This finding corresponds to earlier research findings indicating the importance of parents' educational level as a predictor of children's achievement in general (e.g., Davis-Kean, 2005; Sirin, 2005) and early mathematics achievement in particular (e.g., Entwisle & Alexander, 1990). Our results are in line with the idea that higher-educated parents spend more time and attention on mathematics-related activities with their children, and offer their children higher quality mathematics experiences (cf. Saxe et al., 1987; see also Suizzo & Stapleton, 2007). Further research is needed to investigate whether at-home mathematics activities do indeed play a role here. Moreover, other factors that might mediate the influence of parental education, such as parents' expectations of their children's educational outcomes (e.g., Davis-Kean, 2005), should be further investigated for the case of pre-instructional mathematical knowledge.

Regarding the mathematics textbook used in class, our multilevel analysis revealed a significant effect on students' multiplicative knowledge at the end of Grade 1, when formal instruction of multiplication and division problems has not yet started. This finding is in accordance with other studies in which the textbook was found to affect students' mathematics performance (Hop, 2012; Scheltens et al., 2013; Törnroos, 2005). The differential performance in our study of children taught with different textbooks may be related to differences between the textbooks with respect to including preparatory multiplicative activities (see Table 1). However, as we could not identify any pair-wise differences in the effect of textbooks, it remained unclear which of the textbooks were related to higher student performance.

5.4 Limitations of the study

Some limitations of our study should be noted. A first shortcoming concerns the composition of the test. Since our study primarily aimed to explore the extent of children's pre-instructional multiplicative knowledge, we used an item set in which we included a variety of problem characteristics. However, due to testing time restrictions our test had a limited number of items. For this reason, and as some combinations of problem characteristics were not used because of their assumed too high difficulty level, we could not counterbalance all the problem characteristics. This means that, in our test, certain problem characteristics co-occurred and others did not. In order to more thoroughly study the effects of the different problem characteristics, such as semantic structure and operation, an item set is needed in which all characteristics are combined with all other characteristics. Furthermore, to rule out effects of possible peculiarities in individual items, a larger number of items for each category of a problem characteristic is needed.

A second weakness of our study, which is hard to avoid in a large-scale study such as ours, is that we do not really know the conditions in which the children took the test. Our goal was to study children's abilities in solving multiplicative problems in a relatively formal setting, without the use of physical objects. However, although the teachers were not told to give the students physical objects, we cannot be sure that such objects were indeed not employed. Nevertheless, since the test setting at a computer is quite formal, and since in formal tests in Dutch Grade 1 it is generally not allowed to use physical materials (e.g., the Dutch national student monitoring system, Janssen, Scheltens, & Kraemer, 2005), we assume that in general the teachers in our study will not have provided physical materials to their students.

A last note concerns our use of a computer-based test (which enabled our large-scale data-collection). Although in the Netherlands digital versions of standardized tests do exist for first graders (e.g., Janssen et al., 2005), the participating children were not very familiar with computer-based testing, which is reflected in the typing mistakes that were made (see section 3.4.4). While we think that a computer-based test is equally "formal" as a paper-and-pencil test and thus adequately served our aim of testing children in a relatively formal test setting, one should bear in mind that the children might have scored somewhat differently had they been assessed using a paper-and-pencil test. In a paper-and-pencil test it might, for example, be easier to make auxiliary notes (see, e.g., Johnson & Green, 2006).

5.5 Practical implications

Our study shows that children at the end of Grade 1, before they have received formal instruction on multiplicative reasoning, already display a considerable amount of knowledge in this domain. They even show this knowledge in a relatively formal setting, working on their own at a computer, without having an interviewer sitting next to them and without the use of physical objects. This implies that, when instruction of multiplication starts in Grade 2, there is a lot of prior knowledge to build on, which can even be tapped in more formal situations. In such situations, context problems including a picture with countable elements can help children to apply their existing knowledge. But also in bare number problems in the form of doubling or with the word *times*, children can make use of their pre-instructional knowledge.

Since we found, in accordance with earlier studies (Carpenter et al., 1993; Mulligan & Mitchelmore, 1997), that prior to formal instruction on multiplication and division these two operations appear equally difficult (other factors controlled for), one may question the usual approach of introducing division later than, and separated from, multiplication. Mulligan and Mitchelmore found that young children use the same strategies for both multiplication and division, indicating that children intuitively see connections between the two operations. Possibly, simultaneous introduction of multiplication and division would better exploit these informal insights.

Apart from showing that, in general, much pre-instructional multiplicative knowledge exists in first graders, our findings indicate that this knowledge varies a lot between individual children. For example, we found that children with lower-educated parents generally show a lower level of pre-instructional multiplicative knowledge. For instruction to maximally build on the prior knowledge children bring with them, teachers should, in addition to knowing about the pre-instructional knowledge of the average child, also be aware of individual children's prior knowledge (e.g., Carpenter et al., 1996). To obtain knowledge of individual children's prior knowledge of multiplicative reasoning, we advise teachers to assess their children's multiplicative knowledge before starting formal instruction on the domain. Our study showed that a computer-based test can be a useful way of assessing this knowledge.

References

Anghileri, J. (1989). An investigation of young children's understanding of multiplication. *Educational Studies in Mathematics, 20,* 367-385.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Aubrey, C. (1994). An investigation of children's knowledge of mathematics at school entry and the knowledge their teachers hold about teaching and learning mathematics, about young learners and mathematical subject knowledge. *British Educational Research Journal, 20,* 105-120.

Aunio, P., Niemivirta, M., Hautamäki, J, Van Luit, J. E. H., Shi, J., & Zhang, M. (2006). Young children's number sense in China and Finland. *Scandinavian Journal of Educational Research, 50,* 483-502.

Baroody, A. J. (1987). *Children's mathematical thinking: A developmental framework for preschool, primary, and special education teachers.* New York: Teachers College Press.

Baroody, A. J. (1999). The roles of estimation and the commutativity principle in the development of third graders' mental multiplication. *Journal of Experimental Child Psychology, 74,* 157-193.

Campbell, J. I. D., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology, 39,* 338-366.

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education, 24,* 428-441.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal, 97,* 3-20.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15,* 179-203.

Carr, M., & Davis, H. (2001). Gender difference in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology, 26,* 330-347.

Caygill, R., & Eley, L. (2001, September). *Evidence about the effects of assessment task format on student achievement.* Paper presented at the Annual Conference of the British Educational Research Association, Leeds, UK.

CBS [Central Bureau of Statistics]. (2012). *StatLine Basisonderwijs; Leerlingen in het basis- en speciaal basisonderwijs* [StatLine Primary education; Students in primary and special primary education] [Data File]. Retrieved March 24, 2012, from http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=37846sol&D1=a&D2=1&D3=a&D4=18&HDR=G2,G3,G1&STB=T&VW=T

Christou, C., & Philippou, G. (1999). Role of schemas in one-step word problems. *Educational Research and Evaluation, 5,* 269-289.

Clark, F. B., & Kamii, C. (1996). Identification of multiplicative thinking in children in grades 1-5. *Journal for Research in Mathematics Education, 27,* 41-51.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology, 19,* 294-304.

De Corte, E., Verschaffel, L., & Van Coillie, V. (1988). Influence of number size, problem structure, and response mode on children’s solutions of multiplication word problems. *Journal of Mathematical Behavior, 7,* 197-216.

DeVellis, R. F. (2012). *Scale development.* Newbury Park, NJ: Sage Publications.

Driessen, G. W. J. M. (1997). Pre-reading and pre-arithmetic instruction in infant education in the Netherlands: A multilevel analysis approach. *Early Child Development and Care, 134,* 1-21.

Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Studies in Educational Evaluation, 28,* 347-368.

Entwisle, D. R., & Alexander, K. L. (1990). Beginning school math competence: Minority and majority comparisons. *Child Development, 61,* 454-471.

Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L. W. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher, 27*(5), 6-11.

Ginsburg, H. P. (1977). *Children’s arithmetic: The learning process.* New York: Van Nostrand.

Ginsburg, H. P., Klein, A., Starkey, P. (1998). The development of children’s mathematical thinking: Connecting research with practice. In W. Damon, I. E. Sigel, & K. A. Renninger (Eds). *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed.) (pp. 401-476). Hoboken, NJ: John Wiley & Sons.

Ginsburg, H. P., & Pappas, S. (2004). SES, ethnic, and gender differences in young children's informal addition and subtraction: A clinical interview investigation. *Applied Developmental Psychology, 25,* 171-192.

Grassmann, M., Mirwald, E., Klunter, M., & Veith, U. (1995). Arithmetische Kompetenz von Schulanfängern – Schlußfolgerungen für die Gestaltung des Anfangsunterrichtes – mehr Fragen als Antworten [The arithmetical compentence of beginning first graders – conclusions for education in the early grades – more questions than answers]. *Sachunterricht und Mathematik in der Primarstufe, 23*(7), 302-321.

Green, S. A., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika, 74*, 155-167.

Greer, B. (1992). Multiplication and division as models of situations. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 276-295). New York: Macmillan.

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York: Routledge.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software, 33*(2), 1-22. Retrieved January 16, 2012, from http://www.jstatsoft.org/v33/i02/

Hiebert, J. (1984). Children's mathematics learning: The struggle to link form and understanding. *The Elementary School Journal, 84,* 496-513.

Hop, M. (Ed.). (2012). *Balans van het reken-wiskundeonderwijs halverwege de basisschool 5: Uitkomsten van de vijfde peiling in 2010* [Audit of the mathematics education halfway through primary school 5: Results of the fifth sounding in 2010]. Arnhem, the Netherlands: Cito.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

Ibarra, C. G., & Lindvall, C. M. (1982). Factors associated with the ability of kindergarten children to solve simple arithmetic story problems. *Journal of Educational Research, 75,* 149-155.

Janssen, J., Scheltens, F., & Kraemer, J.-M. (2005). *Leerling- en onderwijsvolgsysteem rekenen-wiskunde groep 3.* [Student and education monitoring system mathematics grade 1]. Arnhem, the Netherlands: Cito.

Johnson, M., & Green, S. (2006). On-line mathematics assessment: The impact of mode on performance and question answering strategies. *Journal of Technology, Learning, and Assessment, 4*(5). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1652/1494

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56,* 746-759.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences, 13,* 129-164.

Kouba, V. L. (1989). Children's solution strategies for equivalent set multiplication and division word problems. *Journal for Research in Mathematics Education, 20,* 147-158.

Kouba, V. L., & Franklin, K. (1993). Multiplication and division: Sense making and meaning. In R. J. Jensen (Ed.), *Research ideas for the classroom: Early childhood mathematics* (pp. 103-126). New York: MacMillan.

Leahey, E., & Guo, G. (2001). Gender differences in mathematical trajectories. *Social Forces, 80,* 713-732.

Lee, J. S., & Ginsburg, H. P. (2009). Early childhood teachers' misconceptions about mathematics education for young children in the United States. *Australasian Journal of Early Childhood, 34*(4), 37-45.

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science, 41,* 55-66.

Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. New York: Springer.

Leinhardt, G. (1988). Getting to know: Tracing students' mathematical knowledge from intuition to competence. *Educational Psychologist, 23,* 119-144.

Lesh, R., Post, T., & Behr, M. (1988). Proportional reasoning. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 93-118). Hillsdale, NJ: Lawrence Erlbaum.

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of experimental child psychology, 53,* 72-103.

Mack, N. K. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education, 26,* 422-441.

Martin, T., Lukong, A., & Reaves, R. (2007). The role of manipulatives in arithmetic and geometry tasks. *Journal of Education and Human Development, 1*(1).

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29,* 587-625.

Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathematics, 8*, 372-377.

Mulligan, J. T., & Mitchelmore, M. C. (1997). Young children's intuitive models of multiplication and division. *Journal for Research in Mathematics Education, 28,* 309-330.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *Timss 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's Guide. Sixth edition.* Los Angeles, CA: Muthén & Muthén.

Nantais, N., & Herscovics, N. (1990). Children's pre-concept of multiplication: Logico-physical abstraction. In G. Booker, P. Cobb, & T. N. de Mendicuti (Eds.), *Proceedings of the 14th Conference of the International Group for the Psychology of Mathematics Eduation* (vol. 3, pp. 289-296). Oaxtapec, Mexico.

NCTM (National Council of Teachers of Mathematics). (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence.* Reston, VA: Author.

Nesher, P. (1992). Solving multiplication word problems. In G. Leinhardt, R. Putnam, & R. A. Hattrup (Eds.), *Analysis of arithmetic for mathematics teaching* (pp. 189-219). Hillsdale, NJ: Lawrence Erlbaum.

Nunes, T., & Bryant, P. (1996). *Children doing mathematics.* Oxford, UK: Blackwell.

Nunes, T., Bryant, P., Burman, D., Bell, D., Evans, D., & Hallett, D. (2009). Deaf children's informal knowledge of multiplicative reasoning. *Journal of Deaf Studies and Deaf Education, 14,* 260-277.

OCW (Ministry of Education, the Netherlands). (2011). CFI database school year 2009-2010 [Data file]. Received by email, December 2, 2012.

OECD. (2009). *PISA 2006 – Technical report.* Paris: Author.

Olivier, A., Murray, H., & Human, P. (1990). Building on young children's informal arithmetical knowledge. In G. Booker, P. Cobb, & T. N. de Mendicuti (Eds.), *Proceedings of the 14th Conference of the International Group for the Psychology of Mathematics Education* (vol. 3, pp. 297-304). Oaxtapec, Mexico.

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research, 37,* 239-253.

Revelle, W. (2012). *An introduction to psychometric theory with applications in R.* Retrieved November 28, 2012, from http://personality-project.org/r/book/. In progress.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507-514.

Saxe, G. B., Guberman, S. R., & Gearhart, M. (1987). Social processes in early number development. *Monographs of the Society for Research in Child Development, 52*(2, Serial No. 216).

Scheltens, F., Hemker, B., & Vermeulen, J. (2013). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 5: Uitkomsten van de vijfde peiling in 2011* [Audit of the mathematics education at the end of primary school 5: Results of the fifth sounding in 2011]. Arnhem, the Netherlands: Cito.

Schwartz, J. L. (1988). Intensive quantity and referent transforming arithmetic operations. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 41-52). Hillsdale, NJ: Lawrence Erlbaum.

Selter, C. (1993). Die Kluft zwischen den arithmetischen Kompetenzen von Erstkläszlern und dem Pessimismus der Experten [The gap between the arithmetic competence of first graders and the pessimism of experts]. In *Beitrage zum Mathematikunterricht 1993* (pp. 350-353). Hildesheim, Germany: Franzbecker.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417-453.

Starkey, P., & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 253-276). Charlotte, NC: Information Age Publishing.

Suizzo, M. A., & Stapleton, L. M. (2007). Home-based parental involvement in young children's education: Examining the effects of maternal education across US ethnic groups. *Educational Psychology, 27,* 533-556.

Ter Heege, H. (1985). The acquisition of basic multiplication skills. *Educational Studies in Mathematics, 16,* 375-388.

Törnroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation, 31,* 315-327.

Van den Heuvel-Panhuizen, M. (1996). The MORE Entry-test – what a paper-and-pencil test can tell about the mathematical abilities of beginning first-graders. In Van den Heuvel-Panhuizen, M., *Assessment and Realistic Mathematics Education* (pp. 189-232). Utrecht, the Netherlands: CD-β Press.

Van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the learning of mathematics, 25*(2), 2-9.

Van den Heuvel-Panhuizen, M. (Ed.). (2008). *Children learn mathematics: A learning-teaching trajectory with intermediate attainment targets for calculation with whole numbers in primary school.* Rotterdam, the Netherlands: Sense.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 127–174). New York: Academic Press.

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

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). ACER ConQuest (Version 2.0) [Computer Software]. Camberwell, Australia: Australian Council for Educational Research.

Chapter 3

Effecten van online mini-games op multiplicatieve vaardigheden van leerlingen in groep 4

Effects of online mini-games on second-graders' multiplicative reasoning ability

Bakker, M., Van den Heuvel-Panhuizen, M., Van Borkulo, S., & Robitzsch, A. (2013). Effecten van online mini-games op multiplicatieve vaardigheden van leerlingen in groep 4. *Pedagogische Studiën, 90*(3), 21-36.

Effecten van online mini-games op multiplicatieve vaardigheden van leerlingen in groep 4

Effects of online mini-games on second-graders' multiplicative reasoning ability

Abstract

Review articles have shown that there is still insufficient evidence for the learning effects of educational computer games. This study used a large-scale randomized experiment (N = 1005; 46 schools) to investigate the effects of mathematics mini-games on second-graders' multiplicative reasoning ability (multiplication and division). Four conditions were included: playing multiplicative mini-games at school, integrated in a lesson (E1), playing multiplicative mini-games at home without attention at school (E2), playing multiplicative mini-games at home with afterwards debriefing at school (E3), and, in the control group, playing at school mini-games on other mathematics topics (C). The mini-games were played during two periods of 10 weeks (16 mini-games in total). Students' multiplicative reasoning ability was measured using a pretest and a posttest, comprising items measuring skills in calculating multiplicative problems as well as items measuring insight in, or understanding of, multiplicative concepts and number relations. Regression-analysis showed that, overall, the intervention with multiplicative mini-games (the three experimental conditions together) did not affect students' learning gains. However, when we separately compared the three experimental groups to the control group, we found a marginally significant effect (p = .07, d = 0.23) for the E3 group. Thus, although hard evidence is lacking, it appears that playing the mini-games at home with debriefing at school has the highest potential in promoting multiplicative reasoning ability.

1 Inleiding

1.1 Effectiviteit van educatieve computerspelletjes

Sinds de opkomst van de computerspelletjes is vaak geopperd dat deze goed in het onderwijs kunnen worden ingezet (bijv. Egenfeldt-Nielsen, 2005; Prensky, 2001). Een belangrijke reden hiervoor is de motiverende werking van computerspelletjes (bijv. Garris, Ahlers, & Driskell, 2002; Malone & Lepper, 1987; Prensky, 2001). Studies hebben laten zien dat het spelen van educatieve computerspelletjes daadwerkelijk de motivatie, en daarmee de leeruitkomsten, kan verhogen (bijv. Cordova & Lepper, 1996; Bai, Pan, Hirumi, & Kebritchi, 2012). Behalve de motiverende werking is een andere belangrijke eigenschap van computerspelletjes dat ze directe feedback kunnen geven. Leerlingen krijgen vaak meteen te zien wat het gevolg is van hun acties in het spel (bijv. Prensky, 2001). Door deze directe feedback en de risico-vrije omgeving die de computer biedt, worden leerlingen gestimuleerd om te exploreren en experimenteren, waardoor ze nieuwe regels en strategieën kunnen ontdekken (Kirriemuir, 2002). In dit verband wordt vaak gesproken over ervaringsleren (*experiential learning*, bijv. Egenfeldt-Nielsen, 2005; Garris et al., 2002).

Door deze veronderstelde voordelen maken computerspelletjes op Nederlandse basisscholen steeds vaker onderdeel uit van het onderwijs (bijv. Kennisnet, 2009). Ook in het reken-wiskundeonderwijs worden regelmatig computerspelletjes of andere educatieve software gebruikt (bijv. Hop, 2012).

Echter, ondanks de verwachte leereffecten van computerspelletjes, laten recente overzichtsartikelen zien dat er nog onvoldoende experimenteel bewijs is voor de effecten van educatieve computerspelletjes in de onderwijspraktijk (bijv. Tobias, Fletcher, Dai, & Wind, 2011; Vogel et al., 2006), en dat grootschalige praktijkexperimenten nodig zijn (bijv. Tobias et al., 2011). Ook op het gebied van rekenen-wiskunde is er onvoldoende bewijs voor de effecten van educatieve computerspelletjes (bijv. Bai et al., 2012). Weliswaar is uit meta-analyses van Li en Ma (2010) en Slavin en Lake (2008) gebleken dat in het algemeen het gebruik van ict in het reken- en wiskundeonderwijs een positief effect heeft op de leerprestaties, maar hier werden spelletjes niet als een aparte categorie onderscheiden. Als problematisch punt bij uitgevoerde studies naar effecten van computerspelletjes en andere educatieve software wordt genoemd dat de onderzoeken nogal eens methodologische mankementen vertonen: vaak is er bijvoorbeeld geen controlegroep (bijv. Vogel et al., 2006), wordt er gebruikt gemaakt van een kleine steekproef (Bai et al., 2012), is er geen sprake van random toewijzing aan condities (Slavin & Lake, 2008), of wordt er in de analyses geen rekening gehouden met de geneste structuur van de data (leerlingen binnen klassen of scholen; Slavin & Lake, 2008).

Om meer duidelijkheid te krijgen over de mogelijkheid en effectiviteit van het inzetten van computerspelletjes in het reken-wiskundeonderwijs op de basisschool, is in 2009, in het kader van het door het ministerie van OC&W opgezette onderzoeksprogramma OnderwijsBewijs, het BRXXX-onderzoeksproject gestart. Indachtig de doelstelling van OnderwijsBewijs hebben we in dit project gekozen voor een grootschalig gerandomiseerd experiment. De in dit artikel beschreven studie vormde hiervan een onderdeel waarin we hebben onderzocht of reken-mini-games kunnen bijdragen aan de multiplicatieve vaardigheden (vermenigvuldigen en delen) van leerlingen in groep 4, de groep waarin over het algemeen wordt gestart met vermenigvuldigen en delen.

1.2 Computerspelletjes in het reken-wiskundeonderwijs

1.2.1 Mini-games

Een veelgebruikt type computerspel in het reken-wiskundeonderwijs is de zogenaamde mini-game (bijv. Jonker et al., 2009; Panagiotakopoulos, 2011). Mini-games zijn korte, gefocuste spelletjes, die gemakkelijk te leren zijn (bijv. Frazer, Argles, & Wills, 2007; Jonker et al., 2009). Ze kunnen vaak gemakkelijk (en veelal gratis) online toegankelijk worden gemaakt, en hebben meestal een flexibele tijdsduur: één spelletje kost vaak maar enkele minuten, en kan eindeloos herhaald worden (bijv., Jonker et al., 2009). Eerdere studies laten zien dat mini-games potentie hebben voor het reken-wiskundeonderwijs (bijv. Jonker et al.; Panagiotakopoulos, 2011).

1.2.2 Automatiseren en inzicht

De meeste van de computerspelletjes en andere educatieve software die in het reken-wiskundeonderwijs worden gebruikt, richten zich op het automatiseren van rekenfeiten (bijv. Mullender-Wijnsma & Harskamp, 2011). Behalve het kennen van rekenfeiten is het voor het kunnen oplossen van complexere opgaven ook belangrijk dat leerlingen inzicht ontwikkelen in getalrelaties en eigenschappen van operaties (bijv. Anghileri, 2006; Nunes, Bryant, Barros, & Sylva, 2012). Voor het vermenigvuldigen en delen betekent dit dat leerlingen naast het paraat hebben van de tafelfeiten, inzicht moeten hebben in de factoren van getallen en de eigenschappen van vermenigvuldigen, zoals de commutatieve eigenschap (bijv. $3 \times 7 = 7 \times 3$) en distributieve eigenschap (bijv. $6 \times 7 = 5 \times 7 + 1 \times 7$). Ook voor het ontwikkelen van dit inzicht kunnen computerspelletjes worden ingezet (bijv. Jonker et al., 2009; Klawe, 1998; Van Borkulo, Van den Heuvel-Panhuizen, Bakker, & Loomans, 2012). Zulke spelletjes zijn vaak gebaseerd op het eerder genoemde ervaringsleren. Door het opdoen van concrete ervaringen in het spel en door het experimenteren met verschillende (reken)strategieën leert de leerling ‘vanzelf’ bepaalde concepten en regels (kennisconstructie, zie Mullender-Wijnsma & Harskamp, 2011), en ontdekt de leerling welke strategieën handig zijn. Reflectie is hierbij cruciaal. Door middel van reflectie kan de leerling het geleerde generaliseren, zodat er *transfer* plaatsvindt en het geleerde ook buiten het spel kan worden toegepast (bijv. Leemkuil & de Jong, 2004; Tobias et al., 2011). Veel onderzoekers zijn van mening dat deze reflectie niet spontaan plaatsvindt bij de leerling (bijv. Leemkuil & De Jong, 2004). Door na afloop van het spelen van een computerspelletje klassikaal of in groepjes hierover na te praten, kan reflectie worden bevorderd (bijv. Egenfeldt-Nielsen, 2005; Klawe, 1998). In zo’n nabespreking, ook wel *debriefing* genoemd (bijv. Garris et al., 2002), worden de leerpunten uit het spel benadrukt en worden verschillende mogelijke strategieën met elkaar vergeleken (bijv. Klawe, 1998). Ook ondersteuning voor en tijdens het spel kan bevorderend werken (bijv. Leemkuil & De Jong, 2004).

1.2.3 Intrinsieke integratie van rekenstof

Rekencomputerspelletjes (en educatieve computerspelletjes in het algemeen) kunnen variëren in de mate waarin de leerstof in het spel geïntegreerd is. Uit onderzoek van Habgood en Ainsworth (2011) bleek dat het rekencomputerspelletje “Zombie Division” een groter leereffect had wanneer de rekenstof geïntegreerd was in de hoofdactiviteit van het spel (*intrinsieke integratie*) – en daarmee echt onderdeel van het spel was – dan wanneer dezelfde rekenstof tussen het spelen door werd aangeboden en dus meer los stond van het spel.

1.2.4 Op school vs. thuis spelen

Mini-games kunnen zowel op school als thuis worden gespeeld. Door de betrokkenheid van de leraar heeft het op school spelen van mini-games het voordeel dat alle leerzame aspecten van de spelletjes kunnen worden benut door ze in de les te bespreken. Echter, het thuis spelen, wat ook veel gebeurt (bijv. Ault, Adams, Rowland, & Tiemann, 2010; Jonker et al., 2009), heeft evenzeer voordelen. Jonker e.a. (2009) rapporteerden bijvoorbeeld dat de website van Rekenweb met name na schooltijd wordt bezocht, hetgeen voor de betreffende leerlingen in feite een uitbreiding inhoudt van de op school bestede tijd aan rekenen, die mogelijk een positieve invloed heeft op hun rekenprestaties. Samenhangend hiermee stellen onderzoekers als Kamil en Taitague (2011) en Tobias e.a. (2011) dat een belangrijke eigenschap van educatieve computerspelletjes is dat hun motiverende werking ervoor zorgt dat leerlingen langer dan gewoonlijk in een leeractiviteit geïnvolveerd zijn.

Wat betreft motivatie heeft het thuis spelen van reken-computerspelletjes mogelijk nog een bijkomend voordeel ten opzichte van het op school spelen. Als een van de motiverende factoren van educatieve computerspelletjes wordt vaak *control* genoemd: de mate waarin de leerling controle heeft over de activiteit (bijv. Malone & Lepper, 1987). In een studie van Cordova en Lepper (1996) leidde een hogere mate van *control* in een reken-computerspelletje, onder andere in de vorm van door de leerlingen te kiezen avatars en spelernamen, inderdaad tot een hogere mate van motivatie bij de leerlingen, en daarmee tot betere leeruitkomsten. Ook keuzevrijheid wat betreft welk spel gespeeld wordt, en wanneer en hoelang dit gespeeld wordt (bijv. Ault et al., 2010), kan worden gezien als een aspect van *control*. Wanneer educatieve spelletjes in de vrije tijd gespeeld worden, is deze keuzevrijheid in grotere mate aanwezig dan wanneer ze op school gespeeld worden, wat mogelijk leidt tot een grotere mate van motivatie bij leerlingen, en daardoor een groter leereffect.

De keuzevrijheid bij het thuis spelen kan naast een voordeel ook een nadeel zijn, omdat de leraar geen controle heeft over welke spelletjes gespeeld worden, en óf ze gespeeld worden. Een ander nadeel van het thuis spelen is dat de leerlingen geen ervaringen kunnen uitwisselen samen met de leraar. Een mogelijke tussenvorm is dat de spelletjes thuis worden gespeeld, maar dat er op school wel een nabespreking plaatsvindt, zodat op de ervaringen uit de spelletjes kan worden gereflecteerd. Deze vorm van het inzetten van computerspelletjes in het onderwijs bleek bijvoorbeeld effectief te zijn in een experiment van Kolovou, Van den Heuvel-Panhuizen en Köller (2013).

1.3 Onze studie

In de hier beschreven studie onderzochten we het op verschillende manieren inzetten van online multiplicatieve mini-games in groep 4, en de effecten hiervan op de multiplicatieve vaardigheden van de leerlingen. De gebruikte mini-games waren zowel gericht op kennisconstructie – via exploreren en experimenteren – als op het automatiseren van rekenfeiten en rekenstrategieën (zie Van Borkulo et al., 2012). In de meeste van de mini-games was sprake van intrinsieke integratie van de rekenstof.

Met ons onderzoek wilden we de volgende onderzoeksvragen beantwoorden:

1. *Wat zijn de effecten van het spelen van multiplicatieve mini-games op de multiplicatieve vaardigheden van leerlingen in groep 4?*
2. *In welke setting zijn multiplicatieve mini-games het meest effectief: wanneer ze op school worden gespeeld, wanneer ze thuis worden gespeeld zonder aandacht op school, of wanneer ze thuis worden gespeeld met een nabespreking op school?*

Onze verwachting was dat multiplicatieve mini-games, in vergelijking met het gewone reken-wiskundecurriculum zonder deze mini-games, een positief effect hebben op het leren van multiplicatieve vaardigheden, omdat ze een motiverende omgeving bieden waarin leerlingen vrij kunnen experimenteren en zo zelf concepten en strategieën kunnen ontdekken. Verder verwachtten we dat de mini-games het meeste effect hebben wanneer ze thuis worden gespeeld en op school worden nabesproken, omdat hier de voordelen van het thuis spelen (extra rekentijd, meer *control*) worden gecombineerd met de voordelen van het op school spelen (nabespreking).

2 Methode

2.1 Onderzoeksopzet

Om de onderzoeksvragen te kunnen beantwoorden, hebben we gebruik gemaakt van een onderzoeksopzet bestaande uit vier condities:

E1 Op school spelen van multiplicatieve mini-games, geïntegreerd in een les.

E2 Thuis spelen van multiplicatieve mini-games, met minimale aandacht op school.

E3 Thuis spelen van multiplicatieve mini-games, gevolgd door een nabespreking op school.

C Op school spelen van mini-games over andere rekenonderwerpen (pseudo-interventie).

Door de controle-conditie (C) te vergelijken met de drie experimentele condities samen (E), kon het effect van de multiplicatieve mini-games worden gemeten (onderzoeksvraag 1). De pseudo-interventie in de controlegroep voorkwam dat het vaststellen van dit effect werd verstoord door het mogelijke positieve effect dat deelname aan een project met computerspelletjes op zichzelf al kan hebben (Hawthorne-effect; zie Parsons, 1974). De drie experimentele condities waren bedoeld om verschillende manieren te vergelijken waarop de spelletjes in het reken-wiskundeonderwijs kunnen worden ingezet (onderzoeksvraag 2).

In alle condities werd de totaal op school bestede tijd per rekenonderdeel hetzelfde gehouden als wanneer niet aan het onderzoek zou zijn deelgenomen. Wanneer er spelletjes, lessen of nabesprekingen op school werden gedaan, gebeurde dat dus als onderdeel van de beschikbare rekentijd voor het betreffende rekenonderdeel: in de experimentele groepen als onderdeel van het onderwijs in multiplicatieve vaardigheden, en in de controlegroep als onderdeel van het onderwijs in andere rekenonderwerpen. Op deze manier konden we het reguliere lesprogramma voor multiplicatieve vaardigheden (in de controlegroep) vergelijken met een lesprogramma waar het spelen van mini-games deel van uitmaakte (in de experimentele groepen).

De hier beschreven studie is gestart met een voortoets over multiplicatieve vaardigheden aan het eind van groep 3 (Toets 1). Toen dezelfde leerlingen in groep 4 zaten, hebben zij gedurende twee periodes van 10 weken met mini-games gespeeld, volgens een van bovenstaande condities. In juni 2011, aan het eind van groep 4, werd een natoets over multiplicatieve vaardigheden afgenomen (Toets 2). Beide toetsen waren online toetsen die op school werden afgenomen.

2.2 Onderzoeksgroep

Bij het werven van scholen voor het onderzoek hebben we, om een voor Nederland representatieve steekproef van scholen te krijgen, scholen benaderd die varieerden wat betreft de schoolkenmerken gemiddeld leerlinggewicht, schoolgrootte, verstedelijking en richting (denominatie). Ter voorkoming van een oververtegenwoordiging in de steekproef van scholen met een specifieke onderwijsvisie, zoals Montessori- en Jenaplanscholen, of scholen met een minder vaak voorkomende denominatie, zoals islamitische en gereformeerde scholen, hebben we dergelijke scholen uitgesloten van deelname. Dit betreft 13.3% van de Nederlandse basisscholen (berekening o.b.v. CFI, 2011). Verder hebben we alleen scholen in de steekproef opgenomen waarvan de onderwijskwaliteit door de Onderwijsinspectie als voldoende was beoordeeld.

Om de geworven scholen gelijkmatig over de verschillende onderzoekscondities te verdelen is gebruik gemaakt van *blocking*. Hierbij zijn de scholen op basis van schooleigenschappen in groepjes van vier of vijf aan elkaar gekoppeld. Vervolgens werd uit elk groepje random één school aan elk van de experimentele condities (E1, E2, E3) toegewezen, en één of twee scholen aan de controlegroep.[1]

In totaal zijn 66 Nederlandse basisscholen met 81 klassen en 1661 leerlingen in juni 2010 begonnen met het eerste deel van het onderzoek, de afname van Toets 1. In Tabel 1 (linkerzijde) is weergegeven hoe deze scholen, klassen en leerlingen over de vier condities waren verdeeld. Door diverse redenen is een aantal scholen voortijdig gestopt met hun deelname aan het onderzoek. In september 2010 zijn 61 scholen (67 klassen) met 1463 leerlingen gestart met de interventie met mini-games, en op het moment van Toets 2 (juni 2011) deden nog 54 scholen (58 klassen) mee, met in totaal 1233 leerlingen (zie Tabel 1, midden).

Tabel 1
Aantallen deelnemende scholen, klassen en leerlingen in juni 2010, september 2010, en juni 2011, en in de analyse

	juni 2010		september 2010		juni 2011		in analyse	
Conditie	scholen (klassen)	lln	scholen (klassen)	lln	scholen (klassen)	lln	scholen (klassen)	lln
C	21 (25)	519	21 (23)	498	20 (22)	461	19 (20)	415
E1	15 (18)	381	13 (15)	336	9 (9)	206	8 (8)	176
E2	15 (19)	394	14 (14)	342	13 (13)	307	11 (11)	254
E3	15 (19)	367	13 (15)	287	12 (14)	259	8 (9)	160
Totaal	66 (81)	1661	61 (67)	1463	54 (58)	1233	46 (48)	1005

Noot. Bij juni 2011 staan de aantallen leerlingen die over waren gebleven van de deelnemende leerlingen in september 2010. lln = leerlingen.

Om de effecten van de interventies in de verschillende condities zo zuiver mogelijk te kunnen meten, hebben we in de hier gepresenteerde analyse alleen de scholen meegenomen waar de interventie voor meer dan de helft is uitgevoerd (zie paragraaf 2.3.3). Dit zijn 46 scholen met samen 48 klassen (zie Tabel 1, rechterzijde). Van deze scholen hebben we die

[1] Om de controlegroep te kunnen vergelijken met zowel de afzonderlijke experimentele groepen (E1, E2, en E3) als met alle experimentele groepen samen, hebben we ervoor gekozen om de controlegroep groter te maken dan de drie afzonderlijke experimentele groepen (1.5 keer zo groot).

leerlingen meegenomen die de gehele interventieperiode leerling waren in de deelnemende klas, en minstens één van de multiplicatieve vaardigheidstoetsen hebben gemaakt ($N = 1005$).

We hebben de representativiteit van onze uiteindelijke steekproef van leerlingen ($N = 1005$) getoetst middels een vergelijking met de dataset van de Nederlandse populatie van basisschoolleerlingen in het schooljaar 2009-2010 (1,548,419 leerlingen; CBS, 2012). Deze dataset bevat informatie over het geslacht en het leerlinggewicht van de leerlingen. Uit Chi-kwadraat toetsen bleek dat, voor zowel geslacht als leerlinggewicht, onze steekproef niet significant verschilde van de populatie ($p > .05$), en dus, wat betreft deze leerlingkenmerken, als representatief kon worden beschouwd. Verder hebben we de representativiteit van onze steekproef van scholen ($N = 46$) onderzocht door deze te vergelijken met een selectie uit het databestand van Nederlandse basisscholen (CFI, 2011) volgens de bovengenoemde selectiecriteria wat betreft onderwijsvisie en denominatie (6,035). Chi-kwadraat en t toetsen wezen uit dat onze steekproef van scholen voor alle onderzochte schoolkenmerken – denominatie, verstedelijking, schoolgrootte, en gemiddeld leerlinggewicht – representatief is ($p > .05$) voor de genoemde selectie van Nederlandse basisscholen.

2.3 Interventieprogramma

In groep 4 speelden de leerlingen gedurende twee periodes van 10 weken met mini-games: van eind september tot begin december en van februari tot april. In elke spelletjesperiode werden acht verschillende mini-games aangeboden: elke week een nieuw spelletje, behalve in de vijfde en de tiende week, waarin eerdere spelletjes konden worden herhaald.

2.3.1 De spelletjes

De mini-games die in de experimentele groepen werden gebruikt, waren zowel gericht op kennisconstructie – via exploreren en experimenteren – als op het automatiseren van rekenfeiten en rekenstrategieën (zie Van Borkulo et al., 2012). In de spelletjes kwamen verschillende concepten en strategieën aan de orde, zoals de commutatieve, associatieve en distributieve eigenschap, en rekenstrategieën zoals één meer/minder en verdubbelen en halveren. In de meeste van de gebruikte mini-games was sprake van intrinsieke integratie van de rekenstof: de rekenstof was onderdeel van de hoofdactiviteit van het spelletje (zie Habgood & Ainsworth, 2011).

De gebruikte spelletjes waren aangepaste versies van mini-games van Rekenweb (www.rekenweb.nl). De aanpassingen hadden zowel betrekking op de moeilijkheidsgraad van de achterliggende vermenigvuldig- en deelopgaven als op de leermogelijkheden van de spelletjes. Zo kwamen in de eerste spelletjesperiode voornamelijk opgaven met 2, 5 en 10 voor, terwijl in de tweede periode ook opgaven met 3 en 4 en met 6, 7, 8, en 9 voorkwamen. Verder werden in de aangepaste spelletjes bijvoorbeeld meer en duidelijkere verbindingen gelegd tussen opgaven en representaties (bijv. formele notatie en rechthoek-structuur). Ook hebben we aan elk spelletje een scoringsmechanisme toegevoegd, waarbij de score hoger werd naarmate de leerling het spelletje vaker met succes had afgerond. Voor de controlegroep hebben we bestaande, niet door ons aangepaste mini-games van Rekenweb gebruikt, gericht op ruimtelijke oriëntatie, optellen en aftrekken.

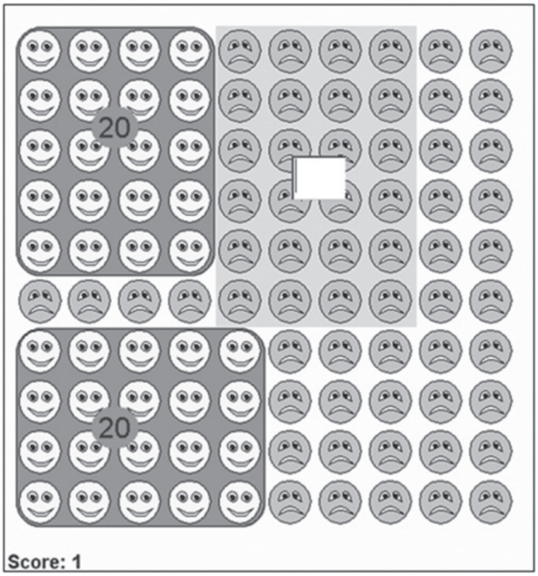
Figuur 1 toont twee spelletjes uit het programma voor de experimentele groepen.[2] In het spelletje “Groepjes maken” (Figuur 1a) moest de leerling steeds een rechthoekig groepje van gezichtjes maken en vervolgens bepalen hoeveel gezichtjes het groepje had. In dit spel oefende de leerling met het berekenen van keersommen (dan wel als gememoriseerde keersom, dan wel door middel van herhaald optellen). Ook kon het spelletje bijdragen aan inzicht in relaties tussen keersommen (kennisconstructie); bijvoorbeeld, 4 rijen van 5 is evenveel als 5 rijen van 4 (commutatieve eigenschap) en 5 rijen van 4 is samen 20, dan is 6 rijen 4 meer, dus 24 (één meer/minder strategie). In het spelletje “Kikker” (Figuur 1b) werd de leerling gevraagd zelf een keersom te bedenken, waarna de kikker het antwoord op een gerelateerde keersom vroeg. Ook in dit spel werd het uitrekenen van keersommen geoefend en kreeg de leerling inzicht in de relaties tussen keersommen.

De experimentele en controle-spelletjes werden online beschikbaar gesteld, via de Digitale Wiskunde Omgeving (DWO).[3] In de eerste week van een spelletjesperiode bevatte de DWO alleen het eerste spelletje van die periode; in latere weken werden latere spelletjes toegevoegd.

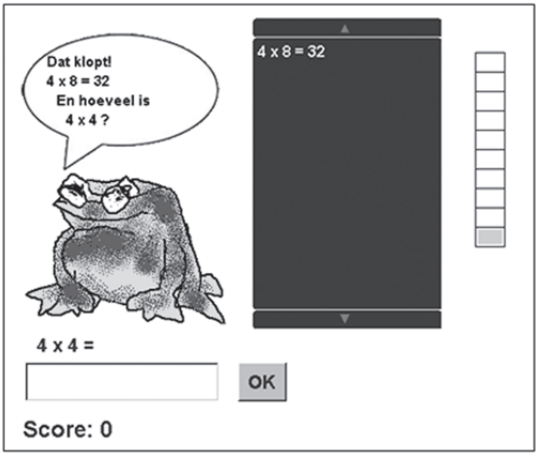
Omdat in de E2- en E3-conditie de kinderen thuis zelfstandig de spelletjes speelden, moesten we een manier vinden om de spelletjes zonder uitleg van de leerkracht te kunnen spelen. Een schriftelijke instructie over de werking van de spelletjes leek ons niet geschikt voor de betrokken leeftijdsgroep. Daarom hebben we bedacht om instructiefilmpjes toe te voegen bij de spelletjes. In deze filmpjes liet iemand die het spelletje speelt de leerlingen door hardop te denken en te zeggen wat zij doet, de fijne kneepjes van het spel en verschillende mogelijke strategieën zien. Dit is in lijn met de suggestie van Leemkuil en De Jong (2004) dat ondersteuning vóór het spelen bevorderend kan werken voor het leren. De filmpjes, die op de openingspagina van de spelletjeswebsite aangeklikt konden worden, duurden elk ca. 3 minuten.

---

[2] In de appendix van dit proefschrift zijn afbeeldingen en beschrijvingen opgenomen van alle spelletjes in het programma voor de experimentele groepen (Game period 1 en Game period 2).
[3] De DWO is ontwikkeld door Peter Boon van het Freudenthal Instituut.



a



b

Figuur 1. Voorbeeld-spelletjes uit het programma van de experimentele groepen. a. "Groepjes maken" (spelletje 2 uit spelletjesperiode 1), b. "Kikker" (spelletje 3 uit spelletjesperiode 2).

2.3.2 Instructies voor de leerkrachten

Voorafgaand aan elke spelletjesperiode kregen de leerkrachten een handleiding, waarin per week stond beschreven welk spelletje die week aan bod kwam en hoe dit spelletje moest worden aangeboden. Samengevat gaven de handleidingen voor de verschillende condities de volgende aanwijzingen:

E1 De leerkracht introduceert het spelletje in een klassikale les (ca. 20 minuten), aan de hand van een werkblad. Hierna bekijken de leerlingen het instructiefilmpje bij het spelletje en gaan ze het spelletje spelen. Nadat alle leerlingen ongeveer 10 minuten hebben gespeeld wordt het spelletje klassikaal nabesproken (ca. 15 minuten), gebruikmakend van het digibord of een klassikale computer. In de handleiding staat aangegeven welke onderwerpen hierbij aan bod moeten komen. Het gaat er hierbij om dat er een discussie ontstaat over welke strategieën handiger of sneller zijn in het spelletje. Hierna spelen de leerlingen nogmaals 10 minuten met het spelletje, waarbij ze de strategieën die in de klassendiscussie zijn besproken kunnen uitproberen.

E2 De leerkracht kondigt aan dat het spelletje op de spelletjeswebsite staat en dat de leerlingen dit thuis mogen spelen. Verder wordt er op school geen aandacht aan het spelletje besteed. De leerkracht controleert niet expliciet of de leerlingen het spelletje hebben gespeeld.

E3 De leerkracht kondigt aan het begin van de week aan dat het spelletje op de spelletjeswebsite staat en dat de leerlingen dit thuis mogen spelen, en dat dit spelletje aan het eind van de week in de klas zal worden nabesproken. De klassikale nabespreking (ca. 15 minuten) ziet er hetzelfde uit als de bespreking in de E1-conditie: Er wordt besproken wat de leerlingen hebben ontdekt en welke strategieën ze handig vinden. Net als in de E2-conditie controleert de leerkracht niet expliciet of de leerlingen het spelletje hebben gespeeld. Ook leerlingen die niet thuis gespeeld hebben doen gewoon mee met de nabespreking.

C De leerkracht introduceert het spelletje in een klassikale les (ca. 10 minuten), op het digibord of op een computer. Hierna spelen de leerlingen het spelletje, in één of twee sessies van 10 minuten (afhankelijk van de beschikbare tijd).

Vóór de start van de eerste spelletjesperiode is voor de leerkrachten van de experimentele groepen (E1, E2 en E3) een informatiebijeenkomst georganiseerd. De leerkrachten werd uitgelegd dat er drie verschillende condities waren en dat het belangrijk was om zich aan de instructies van de eigen conditie te houden, om zo de effecten van de verschillende condities goed te kunnen meten. De leerkrachten uit de controlegroep werden geïnformeerd door middel van een informatiepakket. Deze leerkrachten werd niet verteld dat het onderzoek nog andere condities bevatte. Ook werd hen niet verteld dat het onderzoek ging om multiplicatieve vaardigheden. Er werd enkel gezegd dat het ging om computerspelletjes om de rekenvaardigheden te bevorderen.

Zoals genoemd in paragraaf 2.1 werd de leerkrachten in alle condities gevraagd om ervoor te zorgen dat de totale lestijd die zij op school besteedden aan de verschillende rekenonderdelen hetzelfde was als wanneer zij niet zouden meedoen aan het onderzoek. Verder werd er in alle condities een informatiebrief meegegeven voor de ouders. In de condities E2 en E3 gaf deze brief uitleg over de rol van de ouders bij het thuis spelen: Het was de bedoeling dat de ouders hun kinderen niet aanspoorden om de spelletjes te spelen, maar hen enkel de mogelijkheid gaven de spelletjes te spelen, bijvoorbeeld door te helpen met het bereiken van de spelletjeswebsite. Ook werd aangegeven dat het de bedoeling was dat de kinderen vóór het spelen het instructiefilmpje bekeken.

2.3.3 Controle op uitvoering van de interventie

Om te kunnen nagaan in hoeverre de interventie als beoogd werd uitgevoerd, vroegen we de leerkrachten tijdens de spelletjesperiodes een logboek bij te houden, waarin zij per week konden noteren of ze de verschillende onderdelen van de interventie wel en niet hadden uitgevoerd. Uit de logboekgegevens is gebleken dat niet op alle scholen de interventie zoals bedoeld is uitgevoerd. Op een aantal scholen zijn, door tijdgebrek of doordat de leerkracht het was vergeten, niet alle spelletjes behandeld (d.w.z. op school gespeeld (C en E1), aangekondigd (E2) of nabesproken (E3)). Om de effecten van de spelletjes zo zuiver mogelijk te kunnen meten, hebben we alleen die scholen waar meer dan de helft van de 16 spelletjes zijn behandeld meegenomen in de analyses (zie paragraaf 2.2).

In de DWO is voor de experimentele groepen bijgehouden hoeveel tijd de leerlingen met de spelletjes hebben gewerkt. In de E1-conditie (op school spelen) was dit gemiddeld 316 minuten per leerling voor de twee spelletjesperiodes van 10 weken samen ($SD = 124$, mediaan = 321), in de E2-conditie (thuis spelen) was dit gemiddeld 113 minuten ($SD = 209$, mediaan = 47), en in de E3-conditie (thuis spelen met nabespreking op school) gemiddeld 151 minuten ($SD = 255$, mediaan = 84). Opgemerkt moet worden dat de aan de spelletjes bestede tijd in de E1-conditie onderdeel was van het lesprogramma op school (het kwam in de plaats van andere les over multiplicatieve vaardigheden), terwijl in de E2- en E3-conditie de tijd die thuis aan de spelletjes is besteed een toevoeging was op de op school bestede tijd aan multiplicatieve vaardigheden.

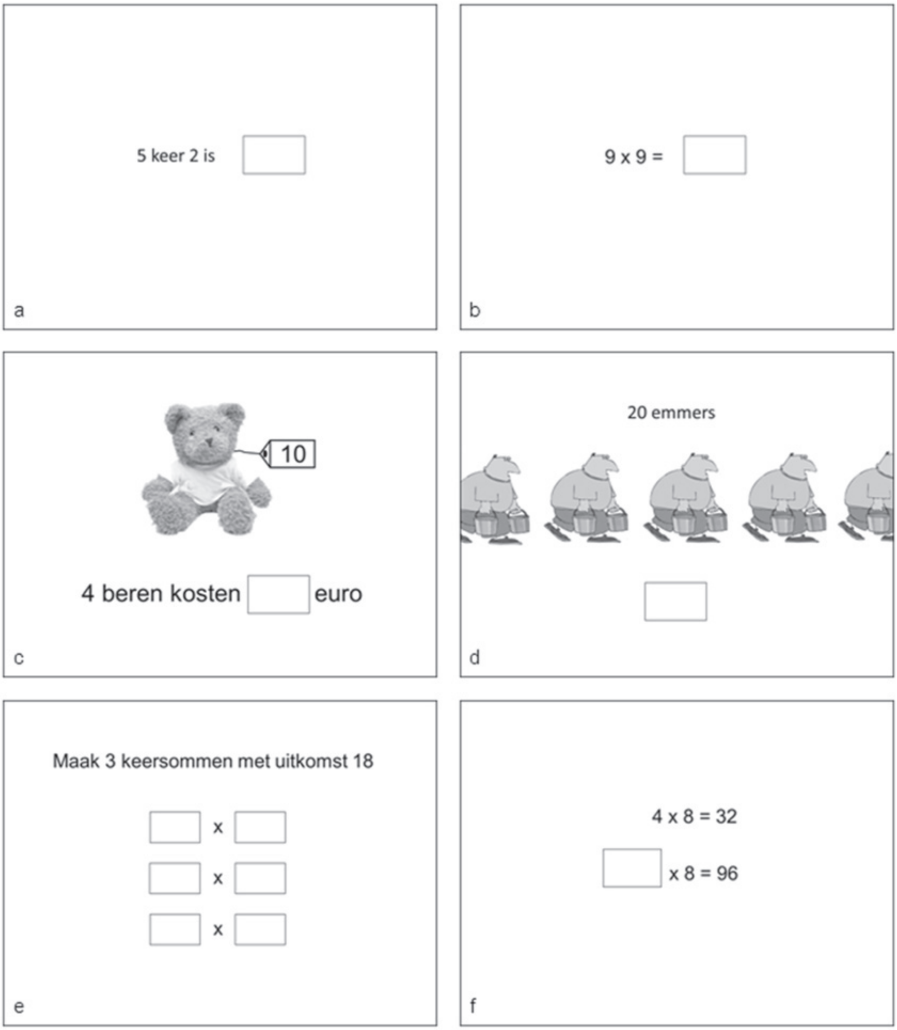
2.4 Toetsing van de multiplicatieve vaardigheden

2.4.1 Samenstelling van de toetsen

De multiplicatieve vaardigheden van de leerlingen werden gemeten met een voor- en een natoets die binnen het BRXXX-project zijn ontwikkeld. Voordat de toetsen op de onderzoeksscholen zijn afgenomen, zijn ze uitgeprobeerd op twee scholen die niet aan het onderzoek deelnamen.

Om de multiplicatieve vaardigheden van de leerlingen in brede zin te kunnen meten, bevatten de toetsen drie typen multiplicatieve opgaven: *kale sommen* om de tafelkennis te meten (bijv. Figuur 2a, 2b); *contextopgaven* om te meten in hoeverre deze kennis kan worden toegepast in een context (bijv. Figuur 2c, 2d); en *inzichtopgaven*, waarin de leerlingen hun tafelkennis op een hoger niveau moesten gebruiken (alleen in Toets 2; bijv.

Figuur 2e, 2f). In tegenstelling tot de kale sommen en contextopgaven waren de inzichtopgaven geen recht-toe-recht-aan opgaven; Bij deze opgaven was expliciet inzicht vereist in de relaties tussen getallen en de eigenschappen van operaties, zoals factoren van getallen en de commutatieve en distributieve eigenschap.



Figuur 2. Voorbeeld-items uit Toets 1 en 2. a. Kale som (Toets 1); "Vijf keer twee is… " b. Kale som (Toets 2); "Negen keer negen is… " c. Contextopgave (Toets 1 en 2); "Hoeveel kosten vier beren samen?" d. Contextopgave (Toets 1 en 2); "Twintig emmers. Hoeveel mannen zijn nodig om ze te dragen?" e. Inzichtopgave (Toets 2); "Maak drie verschillende keersommen met uitkomst 18. Je mag geen keersommen met één maken." f. Inzichtopgave (Toets 2); "Vier keer acht is 32. Hoeveel keer acht is 96?"

Naast multiplicatieve opgaven bevatten de toetsen ook zogenaamde ‘afleider-items’ gericht op ruimtelijk inzicht en optellen en aftrekken. Deze items waren bedoeld om voor de leerlingen en leerkrachten in de controlegroep de focus op multiplicatieve vaardigheden te verhullen. Toets 1 (de voortoets) bevatte 28 multiplicatieve items en 12 afleider-items. Toets 2 (de natoets) bevatte 50 multiplicatieve items – waarvan er 16 ook in Toets 1 zaten (anker-items) – en 16 afleider-items. In de analyses zijn alleen de multiplicatieve items meegenomen.

Om te corrigeren voor eventuele volgorde-effecten, hebben we beide toetsen in vier verschillende versies aangeboden. Hiertoe werden de items in clusters ingedeeld. Toets 1 bevatte vier clusters (A, B, C, en D) van elk 10 items, die in de verschillende toetsversies in verschillende volgordes werden aangeboden (resp. ABCD, CDAB, BADC, en DCBA). Bij Toets 2 hebben we, om een grotere verscheidenheid aan items te kunnen toetsen, gekozen voor zes clusters (A, B, C, D, E, en F) van elk 11 items. Elke toetsversie van Toets 2 bevatte vier van deze clusters (resp. BADE, ECFB, AFCD, en DEBA), dus 44 items. Door deze opzet kon later de totaalscore over alle 50 multiplicatieve items van Toets 2 worden berekend met behulp van een Rasch model (zie paragraaf 2.4.4).

2.4.2 Toets-procedure

De toetsen werden online via de DWO afgenomen en de leerlingen maakten de toetsen individueel. Door de online afname konden we het grote aantal deelnemers gemakkelijk bereiken en zorgden we voor een relatief formele, gestandaardiseerde toetssituatie. Alle items werden afzonderlijk op het scherm getoond, en de bijbehorende opgaven werden hardop voorgelezen door de computer. Beide toetsen duurden ongeveer 20 minuten per leerling.

Bij de afname van Toets 1 waren er op sommige scholen technische problemen door het gebruik van computers met een erg klein beeldscherm, waardoor de toets-items niet volledig zichtbaar waren. Hierdoor zijn op deze scholen veel items per ongeluk overgeslagen. Bij scholen waarvan de leerlingen gemiddeld 10 of meer ontbrekende antwoorden hadden, hebben we aangenomen dat er zich dergelijke technische problemen hebben voorgedaan en hebben we Toets 1 als ongeldig beschouwd (zie paragraaf 2.5). Dit was het geval voor 3 scholen (59 leerlingen).

2.4.3 Correctie van invoerfouten

Omdat de tekstvakken waarin de leerlingen hun antwoorden moesten typen allerlei soorten invoer accepteerden, waren niet alle antwoorden in de vorm van een getal. Invoerfouten waarbij duidelijk was welk getal bedoeld was, zoals “4’0” of “4o” in plaats van “40” of “vier” in plaats van “4”, werden gecorrigeerd. Bij Toets 1 leidde dit voor 0.60% van de itemresponsen ertoe dat een fout antwoord werd omgezet in een goed antwoord, bij Toets 2 was dat het geval voor 0.04% van de itemresponsen.

2.4.4 Schaling van toetsscores

De multiplicatieve items van Toets 1 (28 items) en Toets 2 (50 items) werden geschaald met een Rasch model, met behulp van de Conquest software (Wu, Adams, Wilson, & Haldane, 2007). Deze schaling resulteerde in schaalscores (*weighted likelihood estimates*, of WLE) voor de twee toetsen afzonderlijk. Om beide toetsscores vervolgens op eenzelfde schaal te kunnen zetten, zodat leerwinstscores berekend konden worden, hebben we gebruikgemaakt van *mean-mean linking* (Kolen & Brennan, 2004), met de aanname dat (bij gelijke leerlingvaardigheid) de item-moeilijkheden van de anker-items in beide toetsen gemiddeld genomen hetzelfde waren. Deze linking methode resulteerde in een verschuivingsconstante, die werd opgeteld bij de WLE scores van Toets 2. Alle analyses zijn gebaseerde op de uiteindelijke WLE scores.

2.4.5 Betrouwbaarheid

Voor Toets 1 vonden we een Cronbachs alfa van .88. Voor Toets 2 hebben we, omdat niet alle versies dezelfde items bevatten, de betrouwbaarheid van de vier verschillende versies apart berekend. Dit resulteerde in Cronbachs alfa's van, respectievelijk, .92, .90, .85 en .88. Deze Cronbachs alfa-waarden duiden op een voldoende betrouwbaarheid van de toetsen. Ook de zogenaamde WLE- betrouwbaarheid (Wu et al., 2007) van de schaalscores van Toets 1 en Toets 2, die op dezelfde manier geïnterpreteerd kan worden als een Cronbachs alfa, bleek voldoende (Toets 1: .84, Toets 2: .87). De lage percentages leerlingen (variërend van 0% tot 2%) met een minimale of maximale score op een toets en de geringe scheefheid (*skewness*) van de schaalscores (Toets 1: .28, Toets 2: .07) wijzen erop dat vloer- en plafondeffecten nauwelijks een rol speelden. Bovendien wordt door het gebruik van schaalscores de invloed van eventuele vloer- of plafondeffecten geminimaliseerd (zie Embretson, 2007).

2.5 Behandeling van ontbrekende gegevens

Voor leerlingen waarvan een van de toetsscores ontbrak of ongeldig was (Toets 1: 114 leerlingen; Toets 2: 89 leerlingen), hebben we door middel van meervoudige data-imputatie schattingen gemaakt voor de ontbrekende scores (zie Graham, 2009). Hiervoor hebben we een imputatiemodel gebruikt met 18 predictoren, waaronder leerlingkenmerken, schoolkenmerken, en toetsscores van leerlingen. Om recht te doen aan de geneste structuur van de data (leerlingen binnen scholen) hebben we gebruik gemaakt van meerniveau-imputatie in de "pan" software (Schafer, 2011). De imputatie resulteerde in

50 geïmputeerde datasets. In Mplus (Muthén & Muthén, 1998-2010) zijn met deze 50 datasets de analyses uitgevoerd, waarna de resultaten werden gecombineerd.

3 Resultaten

3.1 Beginverschillen tussen de groepen

Zoals beschreven in paragraaf 2.2, hebben we bij het indelen van de deelnemende scholen in de vier condities gebruik gemaakt van *blocking* en random toewijzing aan condities. Hoewel deze methode van toewijzing over het algemeen zorgt voor vergelijkbare groepen, kunnen er toevallige verschillen tussen de groepen zijn wat betreft de leerlingsamenstelling. Bovendien kunnen er door het uitvallen van scholen, en na het uitsluiten van scholen doordat zij maar de helft of minder van de interventie hadden uitgevoerd, verschillen zijn ontstaan tussen de vier groepen. Daarom hebben we, vóór de imputatie van ontbrekende toetsscores, de leerlingsamenstelling van de vier groepen vergeleken. Hierbij hebben we gekeken naar de leerlingkenmerken geslacht, leeftijd (leerlingen die een normale leeftijd hebben voor hun jaargroep vs. leerlingen die ouder en dus vertraagd zijn; vgl. Hop, 2012), leerlinggewicht en thuistaal, en naar de algemene rekenvaardigheid (gemeten met de Reken-Wiskunde toets E3 van het Cito leerling- en onderwijsvolgsysteem; Janssen, Scheltens, & Kraemer, 2005) en de scores op Toets 1. We vonden significante beginverschillen tussen de groepen voor leeftijd ($\chi^2(3) = 11.82$, $p < .01$, Cramers $V = .11$), thuistaal ($\chi^2(3) = 9.63$, $p = .02$, Cramers $V = .10$), en rekenvaardigheid ($F(3) = 3.12$, $p = .03$, $\omega^2 = .01$). Om te corrigeren voor deze beginverschillen, hebben we de variabelen *leeftijd*, *thuistaal*, en *rekenvaardigheid* als covariaten meegenomen in de hierna beschreven regressie-analyses.

3.2 Effecten van de spelletjes

Tabel 2 toont per conditie, en voor de drie experimentele condities samen, de gemiddelden en standaarddeviaties (na data-imputatie) van de schaalscores op Toets 1 en Toets 2, en de winstscore (Toets 2 – Toets 1). We zien dat de gemiddelde winstscore van de experimentele groepen samen (E: $M = 2.34$, $SD = 1.26$) wat hoger was dan die van de controlegroep (C: $M = 2.27$, $SD = 1.22$). Als we naar de verschillende experimentele groepen kijken, zien we dat de winstscore van E3 ($M = 2.54$, $SD = 1.15$) het hoogst was. Deze groep had van alle groepen de laagste score op Toets 1, maar de hoogste score op Toets 2.

Om de effecten van de spelletjes te toetsen hebben we lineaire regressie-analyses uitgevoerd met de winstscore als afhankelijke variabele, waarbij de controlegroep als referentiegroep diende. Om te corrigeren voor de geneste data (leerlingen genest binnen scholen), hebben we gebruik gemaakt van *cluster-robust standard errors* (Angrist & Pischke, 2009). Als covariaten werden *leeftijd*, *thuistaal*, en *rekenvaardigheid* meegenomen (zie paragraaf 3.1). Vanwege onze gerichte onderzoekshypotheses hebben we de verschillen van de experimentele groepen met de controlegroep eenzijdig getoetst (d.w.z., we hebben de tweezijdige *p*-waarden door 2 gedeeld).

Tabel 2
Schaalscores (WLE) van Toets 1 en Toets 2, en winstscores, per conditie

		Toets 1		Toets 2		Winstscore[a]	
Conditie	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
C	415	0.03	1.40	2.30	1.34	2.27	1.22
E totaal	590	-0.07	1.29	2.28	1.40	2.34	1.26
E1	176	-0.09	1.34	2.28	1.37	2.37	1.21
E2	254	-0.02	1.23	2.18	1.47	2.20	1.33
E3	160	-0.11	1.31	2.43	1.31	2.54	1.15
Totaal	1005	-0.03	1.34	2.29	1.38	2.31	1.24

[a]Toets 2 – Toets 1.

Model 1 in Tabel 3 toont de resultaten van de vergelijking van de drie experimentele groepen samen (E) met de controlegroep (onderzoeksvraag 1). Zoals we zagen in Tabel 2 hadden de leerlingen in de E-groep gemiddeld een hogere leerwinst dan de leerlingen in de controlegroep, maar uit de regressie-analyse bleek dat dit verschil niet significant was ($B = 0.11$, $p = .23$, $d = 0.08$). Wel zien we een significant effect van de covariaat leeftijd ($B = -0.67$, $p < .01$, $d = -0.50$): vertraagde leerlingen (zittenblijvers) hadden een significant lagere leerwinst dan niet-vertraagde leerlingen.

In Model 2 in Tabel 3 zijn de drie experimentele groepen E1, E2 en E3 afzonderlijk met de controlegroep vergeleken (onderzoeksvraag 2). Hier zien we dat de leerwinsten van de groepen E1 en E2 niet significant hoger waren dan die van de controlegroep (E1: $B = 0.10$, $p = .33$, $d = 0.08$; E2: $B = -0.01$, $p = .47$, $d = -0.01$). Ook voor de E3-groep was het verschil met de controlegroep niet significant op het $\alpha = .05$ niveau ($B = 0.31$, $p = .07$, $d = 0.23$). Echter, omdat de *p*-waarde kleiner is dan .10, zouden we dit verschil ‘marginaal significant’ kunnen noemen. In een analyse van contrasten vonden we ook een marginaal significant verschil tussen de E3- en de E2-conditie ($B = 0.32$, $p = .09$, eenzijdig, $d = 0.24$); andere verschillen tussen experimentele condities onderling waren niet significant ($p > .10$). Net als in Model 1 zien we ook in Model 2 een significant effect van leeftijd ($B = -0.67$, $p < .01$, $d = -0.50$).

Tabel 3

Regressie-analyse met winstscore (Toets 2 – Toets 1) als afhankelijke variabele

	Model 1				Model 2			
	B	*SE*	*p*	*d*[a]	*B*	*SE*	*p*	*d*[a]
Leeftijd[b]	-0.67	0.19	< .01	-0.50	-0.67	0.19	< .01	-0.50
Thuistaal[c]	-0.36	0.22	.10	-0.27	-0.34	0.22	.12	-0.25
Rekenvaardigheid	-0.00	0.00	.70	-0.00	-0.00	0.00	.67	-0.00
Conditie E	0.11	0.15	.23[d]	0.08				
Conditie E1					0.10	0.23	.33[d]	0.08
Conditie E2					-0.01	0.19	.47[d]	-0.01
Conditie E3					0.31	0.21	.07[d]	0.23

Noot. Model 1 vergelijkt de drie experimentele condities samen (E) met de controlegroep. Model 2 vergelijkt de condities E1, E2 en E3 afzonderlijk met de controlegroep.
[a]Effectgrootte: *B* gedeeld door de standaarddeviatie van de scores op Toets 1 ($SD = 1.34$). [b]Referentie-categorie: niet-vertraagde leerlingen. [c]Referentie-categorie: eentalig Nederlands. [d]Eenzijdig.

4 Discussie

4.1 Samenvatting en verklaring van de resultaten

De resultaten van ons onderzoek laten zien dat het inzetten van multiplicatieve mini-games in het lesprogramma multiplicatieve vaardigheden in groep 4, vergeleken met het reguliere lesprogramma voor deze vaardigheden zonder deze mini-games, niet noodzakelijkerwijs zorgt voor een grotere leerwinst op het gebied van multiplicatieve vaardigheden (onderzoeksvraag 1). Echter, wanneer de drie experimentele groepen afzonderlijk werden vergeleken met de controlegroep (onderzoeksvraag 2), vonden we een marginaal significant effect voor de E3-groep, waarin de spelletjes thuis werden gespeeld en op school werden nabesproken ($p = .07$, $d = 0.23$). De leerwinst van de andere twee experimentele groepen verschilde niet van die in de controlegroep ($p > .10$). Hieronder bespreken we voor elk van de drie experimentele condities de mogelijke verklaringen van onze bevindingen.

4.1.1 E1: op school spelen

Het ontbreken van een effect van de E1-conditie ten opzichte van de controlegroep suggereert dat het op school inzetten van de mini-games geen meerwaarde biedt aan het reguliere programma gericht op het leren van multiplicatieve vaardigheden. Een mogelijke verklaring hiervoor is dat als de spelletjes geïntegreerd in een les worden aangeboden, wat in de E1-conditie het geval was, het door de leerkracht gebruikte repertoire van didactische aanpak en uitleg mogelijk niet veel verschilt van wat leerkrachten doorgaans gebruiken bij het onderwijzen van vermenigvuldigen en delen. Verder heeft het spelen van de spelletjes door de E1-leerlingen wellicht slechts in beperkte mate geleid tot exploreren en experimenteren, omdat dit spelen op school en in beperkte tijd moest gebeuren.

4.1.2 E2: thuis spelen zonder aandacht op school

Dat het thuisspelen zonder aandacht op school (E2) geen positief effect heeft gehad op de leerwinst zou verklaard kunnen worden doordat de leerlingen mogelijk niet uit zichzelf hebben gereflecteerd op wat zij in de spelletjes geleerd hebben, waardoor er geen *transfer* van het geleerde heeft plaatsgevonden. Ook is de tijd die de leerlingen uit zichzelf aan de spelletjes hebben besteed mogelijk te kort geweest om tot een leerwinst te leiden.

4.1.3 E3: thuis spelen met nabespreking op school

In tegenstelling tot de E1- en de E2-conditie, wijzen de resultaten van de E3-conditie – het thuis spelen met een nabespreking op school – in de richting van een positief effect op de leerwinst ten opzichte van de controlegroep. Dit kan op verschillende manieren verklaard worden. Ten eerste was de tijd die leerlingen thuis aan de spelletjes besteedden een toevoeging op de rekentijd die op school beschikbaar was, waardoor de tijd die aan vermenigvuldigen en delen werd besteed groter was in de E3-conditie dan in de controlegroep (vgl. Kamil & Taitague, 2011). Mogelijk heeft deze extra bestede tijd zich in de E3-conditie vertaald naar een (marginaal significante) leerwinst doordat de leerlingen op het geleerde konden reflecteren door middel van de nabespreking op school (bijv. Egenfeldt-Nielsen, 2005; Garris et al., 2002), wat in de E2-conditie niet het geval was. Het is echter ook mogelijk dat de aandacht op school in de vorm van een nabespreking er simpelweg voor heeft gezorgd dat de leerlingen werden aangespoord om thuis regelmatig met de spelletjes aan de slag te gaan, iets wat ook wordt gesuggereerd door het feit dat de leerlingen in de E3-conditie meer tijd aan het spelen van de spelletjes hebben besteed dan in de E2-conditie (zie paragraaf 2.3.3). De precieze rol van de nabespreking in de E3-conditie (een reflecterende dan wel aansporende) kan op basis van onze resultaten niet worden vastgesteld.

Een andere mogelijke verklaring voor het marginaal significante effect in de E3-conditie ten opzichte van de controlegroep, is dat de spelletjes bij het thuis spelen meer *control* bieden (leerlingen mogen zelf kiezen wanneer, hoelang en welke spelletjes ze spelen) dan in schoolse situaties het geval is. Enerzijds kan deze grotere mate van *control* hebben geleid tot meer motivatie en daarmee tot een leereffect, anderzijds is het mogelijk dat leerlingen in de E3-conditie vrijer geëxperimenteerd hebben met multiplicatieve relaties dan ze in een schoolse situatie zouden doen, wat tot ertoe geleid kan hebben dat de nabespreking diepgaander was dan de reguliere lessen in vermenigvuldigen en delen.

Tenslotte is vermeldingswaardig dat, ondanks dat niet alle leerlingen in de E3-conditie de spelletjes thuis gespeeld hebben, er toch gemiddeld een positieve invloed lijkt te zijn van de interventie. Mogelijk hebben leerlingen die niet thuis gespeeld hebben toch geprofiteerd van de nabespreking in de klas.

4.2 Vergelijking met andere studies

In vergelijking met andere studies vallen de door ons gevonden resultaten van het op school spelen van reken-computerspelletjes (de E1-conditie) tegen. De niet-significante effectgrootte ($d = 0.08$) die wij vonden voor de E1-conditie is bijvoorbeeld kleiner dan de gemiddelde effectgroottes die gevonden werden in meta-analyses van effecten van ict in het reken-wiskundeonderwijs (Li & Ma, 2009: gemiddelde $d = 0.28$; Slavin & Lake, 2008: mediaan $d = 0.19$). Mogelijk speelt de lengte van onze interventie hierbij een rol. Li en Ma vonden namelijk dat bij langer durende interventies met ict (> een half jaar) over het algemeen minder grote effecten worden gevonden. Mogelijk wordt er in kortere interventies intensiever gewerkt met ict. Bovendien kunnen in een lange tijdsperiode veel andere gebeurtenissen buiten de interventie van invloed zijn op het leren.

Naar de effecten van het thuis spelen van educatieve computerspelletjes is, voor zover wij weten, nog weinig onderzoek gedaan. Kolovou e.a. (2013) onderzochten wel, vergelijkbaar met onze E3-conditie, de effecten van thuis spelen en daarna op school bespreken van een pre-algebra mini-game door leerlingen in groep 8. Zij vonden een effect van $d = 0.31$ ten opzichte van een controlegroep, maar deze controlegroep had geen regulier pre-algebra programma. In onze studie was dit anders. De leerlingen in de controlegroep kregen het reguliere programma voor vermenigvuldigen en delen. Rekening houdend hiermee is de effectgrootte die wij vonden in E3-conditie ($d = 0.23$; marginaal significant) redelijk te noemen. Deze effectgrootte is vergelijkbaar met het effect van $d = 0.16$ dat Kamil en Taitague (2011) vonden voor het buiten schooltijd, in een naschools programma, spelen van een computerspel gericht op het vergroten van de woordenschat. Dit werd door Kamil en Taitague beschouwd als een waardevol effect, gezien het feit dat het spelen van de spelletjes een ‘gratis’ toevoeging was op de leertijd op school, en dat er relatief weinig middelen (bijv. leerkracht-tijd) voor nodig waren.

4.3 Beperkingen van de studie

Opgemerkt moet worden dat de gevonden resultaten alleen iets zeggen over het op de onderzochte manieren inzetten van multiplicatieve mini-games in groep 4 van de basisschool. Generalisaties naar andere leerjaren, andere reken-wiskundeonderwerpen, andere onderwijscontexten, en andere rekencomputerspelletjes zijn in principe niet te maken.

Een verdere beperking is dat we, vanwege het grootschalige karakter van de studie, niet hebben kunnen observeren hoe de nabesprekingen en lessen op school daadwerkelijk zijn verlopen. Hoewel de leerkracht-logboeken een indicatie geven, weten we niet precies hoe nauwkeurig de leerkrachten onze instructies hebben opgevolgd. Precies gezegd geven onze bevindingen de effecten aan van de op basis van onze instructies door de leerkrachten gerealiseerde interventie. Dit sluit goed aan bij de onderwijspraktijk: ook leerkracht-instructies in onderwijsmethodes kunnen meer of minder nauwkeurig worden opgevolgd.

Verder hebben we ons in dit artikel alleen gericht op de algehele effecten van de verschillende interventies en hebben we niet gekeken naar de relatie tussen de tijd die individuele leerlingen aan de spelletjes hebben besteed en hun leerwinst. Vervolgonderzoek naar deze relatie zou meer inzicht kunnen geven in de rol van de spelletjes bij het leren van multiplicatieve vaardigheden. Ook is vervolgonderzoek gewenst naar de rol van een nabespreking op school in het geval spelletjes thuis gespeeld worden.

Een andere belangrijk punt is dat er, zoals vaak het geval is bij langdurige onderzoeken op scholen, gedurende het onderzoekstraject nogal wat scholen zijn afgehaakt (zie Tabel 1). De meest voorkomende redenen waren leerkrachtwisselingen en tijdgebrek. Omdat de scholen die het onderzoekstraject hebben voltooid mogelijk andere eigenschappen hebben dan de scholen die voortijdig zijn gestopt (bijvoorbeeld betere organisatie, betere ict-voorzieningen), kunnen de bevindingen uit dit onderzoek alleen gegeneraliseerd worden naar scholen die bereid zijn de benodigde tijd in een programma met computerspelletjes te steken. In verband hiermee is het interessant dat de interventie die de meeste potentie lijkt te hebben – thuis spelen met een nabespreking op school – relatief weinig vergt van de onderwijstijd en de ict-voorzieningen op school.

4.4 Conclusie

De resultaten van het onderzoek laten zien dat het inzetten van multiplicatieve mini-games in het reken-wiskundeonderwijs in groep 4 niet noodzakelijkerwijs leidt tot een hogere leerwinst bij de leerlingen. Hoewel harde bewijzen ontbreken (het resultaat is marginaal significant), lijkt het erop dat de mini-games wel een meerwaarde kunnen hebben wanneer ze thuis worden gespeeld en op school worden nabesproken. De bevindingen wijzen op de mogelijkheid om de leertijd uit te breiden met het thuis spelen van educatieve computerspelletjes.

Literatuur

Anghileri, J. (2006). *Teaching number sense* (2nd ed.). Londen: Continuum.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Ault, M., Adams, D., Rowland, A., & Tiemann, G. (2010, June). *Targeted educational games: Fun and so much more!* Paper gepresenteerd op de jaarlijkse conferentie van de International Society for Technology in Education, Denver, Colorado.

Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology, 43,* 993-1003.

CBS. (2012). *StatLine Basisonderwijs; leerlingen in het basis- en speciaal basisonderwijs*. Opgehaald op 24 maart 2012 van http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=37846sol&D1=a&D2= 1&D3=a&D4=18&HDR=G2,G3,G1&STB=T&VW=T

CFI. (2011). Gegevens Nederlandse basisscholen schooljaar 2009-2010 [Databestand]. Ontvangen per e-mail op 2 december 2011.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88,* 715-730.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment. Exploring the educational potential of computer games.* Dissertatie. IT-University of Copenhagen, Kopenhagen, Denemarken.

Embretson, S. E. (2007). lmpact of measurement scale in modeling developmental processes and ecological factors. In T.D. Little, J.A. Bovaird & N.A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 63-87). Mahwah, NJ: Lawrence Erlbaum.

Frazer, A., Argles, D., & Wills, G. (2007, September). *Assessing the usefulness of mini-games as educational resources*. Paper gepresenteerd op de jaarlijkse conferentie van de Association for Learning Technology, Nottingham, UK.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model, *Simulation & Gaming, 33,* 441-467.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549-576.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences, 20,* 169-206.

Hop, M. (red.). (2012). *Balans van het rekenwiskundeonderwijs halverwege de basisschool 5: Uitkomsten van de vijfde peiling in 2010.* Arnhem, Nederland: Cito.

Janssen, J., Scheltens, F., & Kraemer, J.-M. (2005). *Leerling- en onderwijsvolgsysteem. Rekenen-Wiskunde. Groep 3.* Arnhem, Nederland: Cito.

Jonker, V., Wijers, M., & Van Galen, F. (2009, October). *The motivational power of mini-games for the learning of mathematics.* Paper gepresenteerd op de European Conference on Game Based Learning, Graz, Oostenrijk.

Kamil, M. L., & Taitague, C. (2011). Developing an electronic game for vocabulary learning: A case study. In S. Tobias & J.D. Fletcher (Eds.), *Computer games and instruction* (pp. 331-351). Charlotte, NC: Information Age.

Kennisnet. (2009). *Vier in balans monitor 2009. Ict in het onderwijs: de stand van zaken.* Zoetermeer, Nederland: Stichting Kennisnet.

Kirriemuir, J. (2002). *The relevance of video games and gaming consoles to the higher and further education learning experience.* Techwatch Report. Opgehaald op 10 december 2012 van http://tecnologiaedu.us.es/nweb/htm/pdf/301.pdf

Klawe, M. M. (1998, June). *When does the use of computer games and other interactive multimedia software help students learn mathematics?* Paper gepresenteerd op de Technology and NCTM Standards 2000 Conference, Arlington, VA.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.

Kolovou, A., Van den Heuvel-Panhuizen, M., & Köller, O. (2013). An intervention including an online game to improve Grade 6 students' performance in early algebra. *Journal for Research in Mathematics Education, 44,* 510-549.

Leemkuil, H.,& De Jong, T. (2004). Games en gaming. In P. Kirschner (red.), *ICT in het onderwijs: The next generation. Katern bij onderwijskundig lexicon, uitbreiding editie III* (pp. 41-63). Alphen aan de Rijn, Nederland: Kluwer.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22,* 215-243.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: Vol. 3 Conative and affective process analyses* (pp. 223-253). Hillsdale, NJ: Lawrence Erlbaum.

Mullender-Wijnsma, M., & Harskamp, E. (2011). *Wat weten we over ... ict en rekenen in het basisonderwijs?* Zoetermeer, Nederland: Stichting Kennisnet.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles: auteur.

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology, 82*, 136-156.

Panagiotakopoulos, C. T. (2011). Applying a conceptual mini game for supporting simple mathematical calculation skills: Students' perceptions and considerations. *World Journal of Education, 1*(1), 3-14.

Parsons, H. M. (1974). What happened at Hawthorne? *Science, 183,* 922-932.

Prensky, M. (2001). *Digital Game-Based Learning.* New York: McGraw-Hill.

Schafer, J. L. (2011). pan: Multiple imputation for multivariate panel or clustered data. R package (Version 0.3) [Computer Software]. http://CRAN.R-project.org/package=pan.

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

Tobias, S. Fletcher, J.D., Dai, D.Y., & Wind, A.P. (2011). Review of research on computer games. In S. Tobias & J.D. Fletcher (Eds.), *Computer games and instruction* (pp. 127-222). Charlotte, NC: Information Age.

Van Borkulo, S., Van den Heuvel-Panhuizen, M., Bakker, M., & Loomans, H. (2012). One mini-game is not like the other: Different opportunities to learn multiplication tables. In S. De Wannemacker, S. Vandercruysse & G. Clarebout (Vol. Eds.), *Communications in computer and information science: Vol. 280. Serious games: The challenge* (pp. 61-64). Berlijn: Springer.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34,* 229-243.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). ACER ConQuest (Version 2.0) [Computer Software]. Camberwell, Australia: Australian Council for Educational Research.

Chapter 4

Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (submitted). Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability.

Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability

1 Introduction

1.1 Educational computer games

Computer games have often been suggested as promising educational tools (e.g., Egenfeldt-Nielsen, 2005; Malone, 1981; Prensky, 2001). The most commonly mentioned benefit of computer games for education is their motivational aspect (e.g., Garris, Ahlers, & Driskell, 2002; Malone, 1981; Malone & Lepper, 1987; Prensky, 2001). In addition, games are assumed to be beneficial for learning because they can provide immediate feedback. Players often instantly see the consequences of their actions in the game (e.g., Prensky, 2001). Moreover, games allow players to try, make mistakes, and then try again without losing face (e.g., Gee, 2005). Because of this risk-free environment and the immediate feedback provided by the computer, players are stimulated to explore and experiment, as was pointed out by Kirriemuir (2002). In other words, games can offer students opportunities for experiential learning (e.g., Egenfeldt-Nielsen, 2005; Garris et al., 2002), enabling them to discover new rules and strategies.

Because of these presumed advantages, computer games are more and more becoming part of primary school education (e.g., Williamson, 2009). In accordance with the expected educational benefits of computer games, a meta-analysis by Wouters, Van Nimwegen, Van Oostendorp, and Van der Spek (2013) reported an overall positive effect of educational computer games in comparison to conventional instruction. However, when only randomized studies were taken into account, they did not find a significant effect. Furthermore, other review studies revealed that there is still insufficient experimental evidence for the effectiveness of educational computer games in the school practice (Tobias, Fletcher, Dai, & Wind, 2011; Vogel et al., 2006; Young et al., 2012), and that large-scale in-class longitudinal studies are needed (Tobias et al., 2011; Young et al., 2012). Authors of review articles argued that studies on the effects of games and other educational software quite often suffer from methodological shortcomings, such as not using a control group (e.g., Vogel et al., 2006), not applying random assignment to conditions (e.g., Slavin & Lake, 2008), using a small sample (Bai, Pan, Hirumi, & Kebritchi, 2012), and not accounting for the nested data structure (Slavin & Lake, 2008).

Also in primary mathematics education, computer games and other educational software are often used (e.g., Mullis, Martin, Foy, & Arora, 2012). Yet, also for the domain of mathematics, evidence for the effects of educational computer games is still insufficient, as is apparent from Bai et al.'s (2012) literature overview. Meta-analyses by Li and Ma (2010) and Slavin and Lake (2008) did show that in general the use of ICT in mathematics education positively affects learning outcomes, but in these analyses games were not taken as a separate category.

To gain evidence about the effectiveness of deploying computer games in mathematics education, we conducted a large-scale randomized experiment, with a longitudinal design. The focus was on mini-games in the domain of multiplicative reasoning (multiplication and division) in the early grades of primary school, in which formal instruction of multiplicative reasoning commonly commences (e.g., Department for Education UK, 2011; NCTM, 2006; Van den Heuvel-Panhuizen, 2008).

1.2 Using computer games in mathematics education

1.2.1 Mini-games

A frequently used type of computer game in mathematics education is the so-called mini-game (e.g., Jonker, Wijers, & Van Galen, 2009; Panagiotakopoulos, 2011). Mini-games are short, focused games that are easy to learn (e.g., Frazer, Argles, & Wills, 2007; Jonker et al., 2009). They are often easily accessible (commonly free of charge), and usually have a flexible time duration; one game often takes only a few minutes and can be repeated endlessly (e.g., Jonker et al., 2009). Earlier studies have shown that mini-games have potential for mathematics education. In an evaluation study by Panagiotakopoulos, Sarris, and Koleza (2013), for example, positive learning outcomes were found in fifth-grade students who worked with a number mini-game. Furthermore, Miller and Robertson (2011) showed the effectiveness of handheld mathematics mini-games in improving 10- and 11-year-olds' mental computation skills.

1.2.2 Multiplicative number fact knowledge, skills, and insight

In learning multiplicative reasoning, it is important to develop ready knowledge of number facts (the multiplication tables), and skills in calculating multiplication and division operations. In addition, students need to develop insight in, or understanding of, multiplicative number relations (e.g., Anghileri, 2006; Nunes, Bryant, Barros, & Sylva, 2012). They should, for example, have insight into the factors of numbers and the properties of multiplication (see, e.g., Chang, Sung, Chen, & Huang, 2008), like the commutative property (e.g., $3 \times 7 = 7 \times 3$) and the distributive property (e.g., $6 \times 7 = 5 \times 7 + 1 \times 7$). These three aspects of multiplicative reasoning ability – number fact knowledge, operation skills, and insight – parallel the three types of knowledge often distinguished in mathematics education: declarative knowledge, procedural knowledge, and conceptual knowledge (see, e.g., Miller & Hudson, 2007).

Many of the computer games and other educational software currently used in primary school mathematics education focus on the first two aspects: number fact knowledge and operation skills (e.g., Mullis et al., 2012). However, also for developing mathematical insight, computer games can be employed (see, e.g., Van Borkulo, Van den Heuvel-Panhuizen, Bakker, & Loomans, 2012). Jonker et al. (2009), for example, described a mini-game for enhancing primary school students' understanding of divisibility, and two studies reported by Klawe (1998) showed the effectiveness of computer games in fostering fifth-graders' understanding of several mathematical concepts. In fact, Ke (2009), in her review article, noted that games seem more useful to promote higher-order thinking than factual knowledge acquisition. The instructional power of games that are focused on insight development is often related to experiential learning, as was, for example, the case for the mathematics game used by Kebritchi, Hirumi, and Bai (2010). In such games, students can learn new concepts and rules by experimenting with different mathematical strategies and discovering which strategies are convenient. To make this learning process happen, reflection is crucial, as is stated by, for example, Egenfeldt-Nielsen (2005) and Garris et al. (2002). By reflection students can generalize what they have learned, leading to transfer, by which what is learned can also be applied outside the game (see, e.g., Tobias et al., 2011). Many researchers argue that this reflection does not occur spontaneously in students (e.g., Leemkuil & De Jong, 2004). Rather, it is proposed that class discussion after playing a game is needed to encourage reflection (e.g., Egenfeldt-Nielsen, 2005; Garris et al., 2002; Klawe, 1998). In such a discussion – also called debriefing (e.g., Garris et al., 2002) – the learning points from the game are emphasized and different possible strategies are compared (e.g., Klawe, 1998). Indeed, Wouters et al. (2013), in their meta-analysis, found that interventions with computer games are more effective when the games are supplemented with other instructional methods, such as debriefing sessions, than when they are presented as a stand-alone activity. Also support before and during the game is assumed to foster learning (e.g., Leemkuil & De Jong, 2004).

1.2.3 Integration of mathematical content in games

Mathematics computer games may vary in the extent to which the learning content is integrated in the game. In a study by Habgood and Ainsworth (2011) it was found that a mathematics computer game had a larger effect on learning when the mathematical content was integrated in the main activity of the game (intrinsic integration) – and thus was really part of the game – than when the same mathematical content was presented in between the main gaming activities and thus was more separated from the game playing experience. Other researchers have also indicated the importance of integrating the learning content into the central game activity (e.g., Egenfeldt-Nielsen, 2005; Malone & Lepper, 1987).

1.3 Playing games at school versus at home

Mini-games can be played at school as well as at home. Because of the involvement of the teacher, playing at school has the advantage that all instructional aspects of the games can be exploited by discussing them in a lesson. Moreover, the teacher has control over whether the games are played. However, playing at home, which also occurs a lot (e.g., Ault, Adams, Rowland, & Tiemann, 2010; Jonker et al., 2009), has advantages as well. Jonker et al. (2009), for example, reported that the Dutch mathematics games website Rekenweb is visited mainly during after-school hours, which, for the students involved, implies an extension of the time that is spent on mathematics. According to Kamil and Taitague (2011) and Tobias et al. (2011), an important characteristic of educational computer games is that their motivational effect can cause students to be involved in a learning activity for a longer time period than is regularly the case. In a study by Sandberg, Maris, and De Geus (2011), for example, primary school students were found to voluntarily spend extra time on language learning when offered a mobile game, which led to increased learning.

Besides the advantage in the form of extra learning time, playing at home may imply that students have more control over the learning activity. This so-called learner control is often mentioned as an important motivating factor of educational computer games (e.g., Malone & Lepper, 1987), which can lead to improved learning. In a study by Cordova and Lepper (1996), for example, learner control in the form of choice of avatars and character names in a mathematics game resulted in enhanced learning outcomes. Freedom of choice concerning which game is played, and when and for how much time it is played, can also be considered an aspect of learner control (e.g., Wouters et al., 2013). When educational games are played in students' free time, this freedom of choice is larger than when they are played at school, which may lead to higher motivation in students, and consequently to higher learning outcomes.

A possible approach to combine the advantages of playing at school and those of playing at home is playing the games at home with afterwards a debriefing at school. In this way, students are stimulated to reflect upon their experiences in the games, as we mentioned in section 1.2.2. This manner of utilizing computer games in education was, for example, found effective in an experiment by Kolovou, Van den Heuvel-Panhuizen, and Köller (2013) focused on informal algebraic reasoning in primary school.

1.4 Gameplay behavior

When using games in education, the amount of time and effort students spend on the games may be an important predictor of their learning outcomes. Indeed, Jansen et al. (2013) found that students who had practiced more problems in a game environment on automatization of number facts, had higher gains in their number fact knowledge. However, for games meant to contribute to gaining mathematical insight, the relation might by less clear. Kolovou et al. (2013), for example, did not find a relation between students' online game involvement – measured as a composite variable consisting of logged-in time and online game actions – and their gain in understanding co-varying quantities. Although this finding might be explained by the class debriefing sessions in the experiment, in which students who had not played the game at home could have learned from the experiences of the students who had played the game, Kolovou et al. also argued that the absence of a relation between online involvement and learning outcomes might be explained by students requiring only a limited amount of experience to discover the concepts to be learned in the game, after which further game playing would not result in more learning.

1.5 Educational games and gender

In studies on computer games, the issue of gender has often been addressed. Although not evidenced in all studies (see, e.g., Volman, Van Eck, Heemskerk, & Kuiper, 2005), it has generally been found that computer games for entertainment are played more by boys than by girls (e.g., Lowrie & Jorgensen, 2011; Rideout, Foehr, & Roberts, 2010). This may lead boys to be more motivated to play educational computer games as well. Boys' larger gameplay experience may also lead them to more quickly learn how to use an educational game, as was found by Bourgonjon, Valcke, Soetaert, and Schellens (2010) for secondary school students. This may mean that boys have more room for learning of the game content. Related to this, De Jean, Upitis, Koch, and Young (1999), in a study with fourth to sixth graders, found that boys more easily recognized the mathematics concepts embedded in a mathematics game than did girls. Possibly, then, for boys, learning from a mathematics game is less dependent on teacher guidance than it is for girls.

From her literature review, Ke (2009) concluded that, while gender can influence gameplay and learning processes, it may less influence learning outcomes. Indeed, in most studies on the learning effects of educational games no influence of gender was found (e.g., Habgood & Ainsworth, 2011; Vogel et al., 2006). An exception is the study by Jansen et al. (2013), in which girls were found to profit more than boys from playing mathematics automatization games.

When educational games are played at home, gender can influence the amount of time and effort students choose to spend on the games. Based on boys' greater liking of playing games, one could expect boys to spend more time on the games. Yet, studies on homework have found that girls tend to devote more effort to homework (e.g., Trautwein, Ludtke, Schnyder, & Niggli, 2006), which could imply that girls might spend more effort on educational games as well. In accordance with the latter, Kolovou et al. (2013) found that girls showed more online game involvement than did boys; however, this did not lead to girls having higher learning outcomes.

1.6 Educational games and prior knowledge

Students' content domain knowledge prior to playing a game may also influence how much is learned from the game. In their review, Tobias et al. (2011) found that students with a lower initial performance tend to profit more from educational games than their higher performing peers. An explanation may be that the motivational aspects of games are especially helpful for lower achieving students, whereas higher achieving students may not need this extra motivation. However, for the domain of mathematics, findings are not clear. Habgood and Ainsworth (2011), and Kebritchi et al. (2010), for example, found no differences in learning gains between students with low and high initial mathematics ability. Another thing is that a certain level of prior knowledge may be needed to be able to discover new concepts and strategies in a game, which means that students with lower prior knowledge may need more teacher guidance in experiential learning games. This was, for example, suggested by Kebritchi et al.'s (2010) study, in which low prior knowledge students were found to need more assistance in playing a mathematics computer game.

Another way in which prior knowledge may influence the effect of playing educational games is through time and effort spent on the games. Although we did not find literature on this topic, it might be the case that students with higher prior knowledge choose to spend more time playing educational games, because they like the games more or because they understand them better. Especially when students play educational games at home and are more free in deciding how long they play, their prior knowledge can influence their gameplay behavior and, in consequence, their learning outcomes.

1.7 Our study

In the current study we investigated the effects of multiplicative mini-games on students' multiplicative reasoning ability in Grade 2 and Grade 3. We examined the effectiveness of three different ways of deploying the mini-games: playing at school, playing at home, and playing at home with debriefing at school. The mini-games used in the study focused both on automatizing multiplicative number facts and multiplicative operations (through practicing), and on developing insight in multiplicative number relations (through exploring and experimenting). The aim of our study was to investigate the effects of a mini-games intervention when implemented as part of the regular educational practice. As such, we studied the added value of the mini-games when employed as part of the regular multiplicative reasoning curriculum.

Earlier, we performed a preliminary analysis on the effects of the mini-games in Grade 2 (Bakker, Van den Heuvel-Panhuizen, Van Borkulo, & Robitzsch, 2013 [Chapter 3 of this thesis]), using a combined measure of multiplicative ability including multiplicative operation skills and insight. The analysis revealed no significant effects, but in the condition in which the games were played at home and afterwards debriefed at school, the effect was marginally significant ($p = .07$, $d = 0.23$). The current study covered a two-year intervention in Grade 2 and Grade 3 and investigated the effects of the games on three aspects of students' multiplicative reasoning ability: ready *knowledge* of multiplicative number facts, multiplicative operation *skills*, and *insight* in multiplicative number relations.

Furthermore, we examined the role of gameplay behavior, gender, and prior mathematics ability.

The following research questions were investigated:

1. *Does an intervention with multiplicative mini-games – either played at school, played at home, or played at home and afterwards debriefed at school – affect students' learning outcomes in multiplicative reasoning?*
2. *Does an intervention with multiplicative mini-games affect students' learning outcomes in all the three aspects of multiplicative reasoning: knowledge, skills, and insight?*
3. *In what setting – playing at school, playing at home, or playing at home with debriefing at school – are the multiplicative mini-games most effective?*
4. *Are students' learning outcomes related to their gameplay behavior?*
5. *Are students' learning outcomes related to gender?*
6. *Are students' learning outcomes related to their prior mathematics ability?*

Our hypothesis for Research question 1 was that, in each of the three game-playing settings, the intervention with multiplicative mini-games would positively affect the learning of multiplicative reasoning, in comparison to the regular mathematics curriculum without these mini-games. This hypothesis is based on the motivating environment and immediate feedback provided by educational games. Regarding Research question 2, we hypothesized that the mini-games would be effective in enhancing all three aspects of students' multiplicative reasoning ability. With respect to Research question 3, we hypothesized the mini-games to be most effective when played at home and afterwards debriefed at school. In this setting the advantage of playing at home (extra time-on-task, more learner control) is combined with the advantage of playing at school (debriefing). Furthermore, for Research question 4, our hypothesis was that students' gameplay behavior would be positively related to their learning outcomes with respect to number fact knowledge and skills. The relation with their insight learning outcomes may be less clear, but if there is a relation we expect it to be positive. Finally, regarding Research question 5 and 6, we did not specify a hypothesis, because findings from earlier studies on the effects of gender and prior mathematics ability on learning from games are inconclusive.

2 Method

2.1 Research design

To answer our research questions, we used a cluster randomized longitudinal experiment containing three experimental conditions (E1, E2, and E3) and a control condition (C):

E1 Playing multiplicative mini-games *at school*, integrated in a lesson.

E2 Playing multiplicative mini-games *at home*, with no attention at school.

E3 Playing multiplicative mini-games *at home*, with *debriefing at school*.

C Pseudo-intervention: playing at school mini-games on other mathematics domains, including spatial orientation, addition and subtraction.

In all conditions, the teachers were asked to keep the total in-class lesson time that was spent on each mathematics domain the same as would have been the case had the school not been participating in the study. In this way, we could compare the regular curriculum for multiplicative reasoning (in the control group) with a multiplicative reasoning curriculum including an intervention with mini-games (in the experimental groups). The pseudo-intervention in the control group prevented the effect of the mini-games from being obscured by the positive effect that participating in an experiment may have by itself (Hawthorne effect, see Parsons, 1974; Rosas et al., 2003).

Figure 1 shows the time schedule of the study. In Grade 2 as well as in Grade 3 there were two game periods, in which the mini-games were played according to one of the aforementioned conditions. To monitor students' learning of multiplicative reasoning, multiplicative ability tests were administered at three measurement points: at the end of Grade 1, at the end of Grade 2, and at the end of Grade 3.

	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Grade 1 (2009/2010)										*Measurement point 1:* Skills Test 1
Grade 2 (2010/2011)	Game period 1					Game period 2				*Measurement point 2:* Knowledge Test 2 Skills Test 2 Insight Test 2
Grade 3 (2011/2012)	Game period 3					Game period 4				*Measurement point 3:* Knowledge Test 3 Skills Test 3 Insight Test 3

Figure 1. Time schedule of the study. Skills Test = test of multiplicative operation skills; Knowledge Test = test of multiplicative fact knowledge; Insight Test = test of insight in multiplicative number relations.

2.2 Participants

When recruiting schools for our study, we aimed for a sample of schools that was representative for the primary schools in the Netherlands with respect to urbanization level, average level of parental education, and school size. When contacting schools by phone (response rate ca. 15%), e-mail (response rate ca. 2%), and an advertisement on a mathematics games website, we found 66 schools to be willing to participate. To evenly distribute the recruited schools over the research conditions, we used a method of blocking. Schools were matched in sets of four or five on the basis of similarity in school characteristics (urbanization level, average parental education, and school size), and after this, random assignment was used to assign from each set of schools one school to each of the experimental conditions, and one or two schools to the control condition. Table 1 shows how the 66 schools, with 81 classes and 1661 students, were distributed over the four conditions.

Table 1
Numbers of schools, classes, and students in the study

	Recruited sample		Sample that completed the study		Analysis sample	
Condition	Schools (classes)	Students	Schools (classes)	Students	Schools (classes)	Students
C	21 (25)	519	17 (19)	356	16 (18)	327
E1	15 (18)	381	8 (9)	168	6 (7)	112
E2	15 (19)	394	13 (16)	284	9 (11)	202
E3	15 (19)	367	9 (11)	185	4 (5)	78
Total	66 (81)	1661	47 (55)	993	35 (41)	719

Note. As some classes merged or split up in the course of the research project, the numbers of participating classes varied somewhat between grades. In the “Recruited sample” column the number of classes in Grade 1 (start of the study) is reported; in the other columns the number of classes in Grade 3 (end of the study) is reported.

For various reasons, such as changes in teachers, organizational problems, and problems with computers, some schools dropped out in the course of the research project. There were five schools that administered the first test in Grade 1 but did not continue the project in Grade 2. Furthermore, seven schools dropped out during the Grade 2 intervention, six schools dropped out after Grade 2, and one school dropped out during the Grade 3 intervention. This means that 47 schools stayed in the project till the end.

To measure the effects of the interventions in the different conditions as accurately as possible, we included in our analyses only those classes in which in both grades more than half of the games were treated (see section 2.3.3). Furthermore, we excluded one school in the E1 condition, because for this school the students' individual gameplay behavior could not be measured since the students played the games in dyads. Finally, we ended up with 35 schools, with 41 participating classes, which were used for the analyses. Of these classes, students who moved to another class or school during the experiment or did not complete any of the multiplicative ability tests were excluded, resulting in a sample of 719 students (see Table 1).

The initially recruited sample was found to be representative of the population of Dutch primary schools as well as of the population of Dutch primary school students with respect to the school characteristics urbanization level, average parental education, and school size, and the student characteristics gender and parental education. The analysis sample, however, differed from the population with respect to parental education. The students in this sample had parents with a higher level of education than had the students in the population (respectively 90.4% and 86.6% of the students had parents who completed at least secondary education, $\chi^2(1) = 8.84$, $p < .01$). Also with respect to the schools' average level of parental education, the analysis sample was not representative of the population ($t(34) = 3.88$, $p < .001$).

When checking for selective dropout, we found that the initially recruited students who were not included in the analysis sample had a significantly lower level of parental education than had the students who were included (respectively 81.7% and 90.4% of the students had parents with at least secondary education, $\chi^2(1) = 24.75$, $p < .001$). Moreover, not-included students had a lower Grade 1 score on general mathematics ability ($M = 39.9$, $SD = 16.6$, see section 2.4.3) than had included students ($M = 45.2$, $SD = 14.7$, $t(1474) = 6.51$, $p < .001$).

2.3 Intervention program

The intervention program included four game periods, each lasting 10 weeks (see Figure 1). In each game period eight different mini-games were offered; every week a new game, except for the fifth and tenth week, which were meant for repeating earlier presented games.

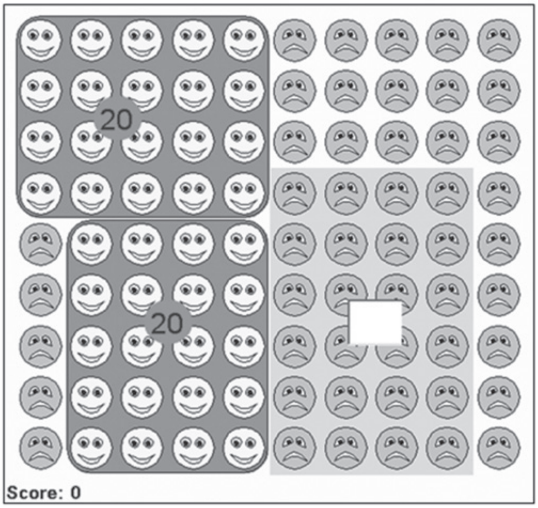
2.3.1 The mini-games

The mini-games that were used in the experimental conditions were mostly adapted versions of multiplicative mini-games selected from the Dutch mathematics games website Rekenweb (www.rekenweb.nl, English version: www.thinklets.nl). The adaptations concerned the inclusion of a scoring mechanism and some changes in the games' difficulty level to make them fit the students' stage in the learning trajectory. Moreover, we modified some games to create more learning opportunities, for example by emphasizing connections between different multiplication problems, and relations between representation and formal notation. The games we used in the control group were existing mini-games from Rekenweb about spatial orientation, addition, and subtraction. For both the experimental groups and the control group, the games were made available online at a games website created using the Digital Mathematics Environment (DME).[1]

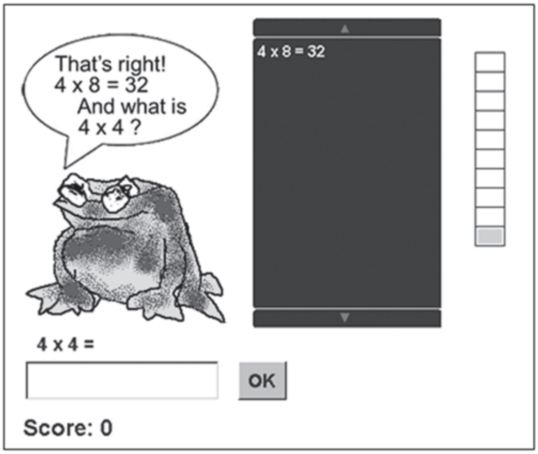
The games in the experimental conditions focused on automatizing multiplicative number facts and multiplicative operation skills (through practicing), and on developing insight in multiplicative number relations and properties of multiplicative operations (through exploration and experimentation). The properties embedded in the games were the principles of commutativity, distributivity, and associativity. Furthermore, the games promoted the use of derived fact strategies such as one more and one less, and doubling and halving. In addition, some games were meant to provide insight into the multiplication-related characteristics of numbers, such as factors of numbers and the divisibility of numbers. In most of the games, the mathematics content was intrinsically integrated into the main activity of the game (see Habgood & Ainsworth, 2011). In agreement with researchers' (e.g., Leemkuil & De Jong, 2004) suggestion that support provided before playing a game may stimulate learning, we added instruction videos to the games. In these videos someone plays the game while thinking aloud and thus introduces in a natural way how the game is played and which strategies can be used. The videos lasted about three minutes each.

A list of the mini-games that were used in the four game periods of the experimental groups intervention program is included in the Appendix of this thesis. As an example, two of the mini-games are shown in Figure 2. In the game "Making groups" (Figure 2a), the student had to make rectangular groups of smileys and then determine the number of smileys in the group. In this game, the student practiced solving multiplication problems (either as memorized multiplication facts or, for example, by repeated addition). Furthermore, the game could contribute to gaining insight into the relations between multiplication problems; for example, 5 rows of 4 is the same as 4 rows of 5 (commutative property), and if 5 rows of 4 is 20, then 6 rows is 4 more, resulting in 24 (derived fact strategy of one more, or distributive property). In the game "Frog" (Figure 2b) the student was asked to come up with their own multiplication problem, after which the frog asked for the answer to a related multiplication problem. Also in this game, the student practiced solving multiplication problems and could gain insight in the relations between multiplication problems.

[1] The DME has been developed by our colleague Peter Boon at the Freudenthal Institute of Utrecht University. See http://www.fi.uu.nl/wisweb/en/



a



b

Figure 2. Example games from the experimental groups intervention program. a. “Making groups”. b. “Frog”.

2.3.2 Instructions for the teachers

Before each game period, the teachers were given a manual in which, for each week, it was described which game was offered that week, and how it had to be treated in class. In summary, the manuals for the different research conditions gave the following instructions:

E1 The teacher introduces the new game in a whole-class lesson (20 minutes), using a worksheet. Afterwards, the students watch the instruction video and play the game.

After all students have played the game for approximately 10 minutes, the game is debriefed in a class discussion (15 minutes), using a digital blackboard or a class computer. In the manual it is indicated which topics should be treated in this discussion. The idea is that the class discusses which strategies are faster or more useful in the game. After this discussion, the students play the game for another 10 minutes, during which they can try the strategies that were discussed.

E2 The teacher announces that there is a new game on the games website and that the students can play this game at home. They can also play the earlier presented games. Apart from this announcement, no attention is paid to the game. The teacher does not check whether the students have played the game.

E3 At the beginning of the week the teacher announces that there is a new game on the games website and that the students can play this game at home. They can also play the earlier presented games. Furthermore, the teacher announces that the new game will be discussed in class at the end of the week. In the class discussion (ca. 15 minutes), for which the instructions in the teacher manual are the same as in the E1 condition, it is discussed what the students have discovered in the game and which strategies they find useful. Like in the E2 condition, the teacher does not check whether the students have played the game.

C The teacher introduces a game from the control group program in a whole-class lesson (10 minutes), using the digital blackboard or a computer. After this, the students play the game in one or two sessions of 10 minutes.

In addition to the teacher manual, in each grade, before the start of the intervention we organized a meeting to inform the teachers of the experimental groups about the intervention program. The teachers were told that there were different research conditions and that it was important to adhere to the instruction of their own condition, to make sure the different conditions could properly be compared. The control group teachers were informed through an extensive information letter sent by (e-)mail. These teachers were not told that other research conditions were included in the study. Moreover, they were not told that the study was about multiplicative reasoning; we only said that it was about computer games for promoting mathematics achievement. Furthermore, in all conditions a letter for the students' parents was handed out. In the E2 and E3 conditions, this letter explained the role of the parents in the playing at home. Parents were told not to urge their child to play the games; they should just give their child the opportunity to do so, for example by helping their child to get online. Also, it was indicated that the child needed to watch the instruction video before playing a game for the first time.

2.3.3 Intervention fidelity

To monitor the intervention fidelity we asked the teachers to keep a logbook in which they could note each week whether they had performed the intervention as described in the teacher manual. From the logbook data it appeared that in several classes not all games were dealt with, due to lack of time or because the teacher had forgotten it. A similar picture arose from the automatically logged gameplay data (see section 2.3.4). To be sure that the students had had sufficient experience with the games, we used an intervention fidelity criterion of more than half of the games having been treated (that is, played at school in E1 and C, debriefed at school in E3, and announced in E2) for deciding whether classes would be included in our analysis (see section 2.2). The decision whether a class met this fidelity criterion was primarily based on the teacher logbooks, as these provided information on teacher actions performed (e.g., debriefing sessions, announcements of new games) in addition to whether games were played. However, because of the possibility of unreliability of the logbook data (teachers may have exhibited socially desirable behavior in filling in the logbook, or may have filled in the logbook at a later time and not remembered exactly what they did), these data were verified using the logged gameplay data. In the case of missing logbook data (ca. five schools per game period), the number of games treated was estimated on the basis of the logged gameplay data together with information obtained through communications with the school.

In the analysis sample obtained using the mentioned intervention fidelity criterion, in Grade 2 on average 14.5 of the 16 games were treated. In Grade 3 this average was 14.0 (see Table 2).

Table 2
Number of games treated in each condition

	Grade 2			Grade 3		
Condition	*N*(classes)	*M*	*SD*	*N*(classes)	*M*	*SD*
C	16	14.8	2.0	17	14.2	2.4
E1	6	15.0	0.9	6	15.2	1.6
E2	9	14.2	2.5	10	13.5	2.0
E3	5	13.8	1.8	5	13.2	1.9
Total	36	14.5	1.9	38	14.0	2.1

Note. The higher number of classes in Grade 3 is because two of the Grade 2 classes were split into two classes when transferred to Grade 3.

2.3.4 Students' gameplay behavior

In the experimental conditions, the DME was used to log data on each student's gameplay behavior. To give an impression of the extent to which the games were played, Table 3 reports per condition the time students spent on the games and the number of different games they played, in Grade 2 (game periods 1 and 2) and in Grade 3 (game periods 3 and 4). In both grades the games were played most frequently in the E1 condition (playing at school) and least frequently in the E2 condition (playing at home without attention at school). All between-condition differences were significant ($p \leq .001$).[2] These differences between conditions correspond to the set-up of the conditions: In E1 there was the most teacher guidance, in E2 the least. Furthermore, in all conditions the games were played more in Grade 2 than in Grade 3 ($p < .001$).[3]

Table 3
Time students spent on games (in minutes) and number of different games they played in the three experimental conditions

	Total time spent on games					Number of different games played				
Condition	*M*	*SD*	*Mdn*	Min	Max	*M*	*SD*	*Mdn*	Min	Max
Grade 2										
E1	366	84	351	187	642	15.4	1.4	16	7	16
E2	120	228	43	0	1813	4.6	4.5	4	0	16
E3	139	130	120	0	569	8.1	5.2	8	0	16
Grade 3										
E1	299	97	275	98	493	14.3	1.5	15	10	16
E2	12	35	0	0	307	1.2	2.1	0	0	10
E3	60	133	0	0	860	3.2	3.9	1.5	0	16

Note. E1: $n = 112$; E2: $n = 202$; E3: $n = 78$. Min = minimum; Max = maximum.

[2] Because of the non-normal distribution of the playing time data, the non-parametric Mann-Whitney U test was used. Grade 2: E2-E1: $z = -12.36$, $r = -.70$; E3-E1: $z = -9.79$, $r = -.71$; E3-E2: $z = 3.26$, $r = .20$. Grade 3: E2-E1: $z = -15.08$, $r = -.85$; E3-E1: $z = -10.35$, $r = -.75$; E3-E2: $z = 4.68$, $r = .28$.
[3] Because of the non-normal distribution of the playing time data, the non-parametric Wilcoxon signed-rank test was used. E1: $z = 7.90$, $r = .75$. E2: $z = 10.03$, $r = .71$. E3: $z = 5.62$, $r = .64$.

The collected log data were used to compute four measures of gameplay behavior for each student: Time, Effort, Success, and NumberOfGames. The first three were computed per game as logarithmic transformations of, respectively, the time spent on the game in seconds, the number of attempted problems in the game, and the number of correct attempts. A logarithmic transformation ($f(x) = \log(x + 1)$) was employed to make the variables conform to a normal distribution and to diminish the impact of outliers. The transformed values were then z-standardized and, subsequently, for each student weighted sums of the Time, Effort and Success variables were computed over the intervention in Grade 2 and Grade 3 separately. The weights were based on the mean amount of time students spent on each game. The fourth measure, NumberOfGames, was computed for Grade 2 and Grade 3 as the number of different games the student played in these grades, ranging from 0 to 16. As our four measures of gameplay behavior were highly correlated for both Grade 2 and Grade 3 (correlations ranging from .76 to .96, $p < .001$), we computed summary measures of gameplay by taking the average of the four measures (after z-standardizing them). This resulted in the gameplay measures Gplay2 for Grade 2 and Gplay3 for Grade 3.

2.3.5 Multiplicative reasoning activities outside the intervention

In-class time spent on multiplicative reasoning. As we mentioned before, in all conditions, the teachers were asked to keep the total in-class lesson time that was spent on each mathematics domain the same as would have been the case if the school had not participated in the study. Thus, in the experimental conditions, the in-class parts of the mini-games intervention were scheduled as part of the time that was normally spent on the topic of multiplicative reasoning, whereas in the control group, the intervention was scheduled as part of the time normally spent on the topics of addition, subtraction, and spatial orientation. This means that the total in-class time spent on multiplicative reasoning was not influenced by the condition the school was in.

To get an idea of the in-class time that was spent on multiplicative reasoning in the different conditions, we asked the teachers to fill in an online questionnaire at the end of each game period. In this questionnaire, teachers were requested to estimate the average time per week that was spent in class on different mathematics topics, including the domain of multiplicative reasoning. Averaged over the four game-periods, we found roughly similar estimates for all conditions for the in-class time spent on multiplicative reasoning (E1: $M = 106$ minutes, $SD = 29$ minutes; E2: $M = 119$, $SD = 31$; E3: $M = 108$, $SD = 20$; C: $M = 103$, $SD = 24$; $F(3, 31) = 0.725$, $p > .10$).

Use of other educational software. Because we wanted to investigate the effects of embedding the mini-games in the real educational practice, no restrictions were placed on the contents of the multiplicative reasoning curriculum outside the mini-games intervention program. This means that teachers and students were not forbidden to work with other educational software as well, as this would also happen in normal school practice.[4] Thus, our study investigated the effectiveness of our mini-games intervention beyond the effects of possible other educational software used. To get an indication of the total amount of educational software for the multiplicative reasoning domain that was used in the different conditions, the abovementioned teacher questionnaire also contained a question on how much in-class time, on average per week, was spent on educational software/games in different mathematics domains, including multiplicative reasoning. Based on the setup of our study, we would expect the average amount of in-class time per week spent on multiplicative reasoning software to be highest in the E1 condition, in which the intervention consisted of playing multiplicative mini-games at school. The teacher estimates confirmed this: in the E1 condition, the estimated amount of time was significantly higher than in each of the other conditions (E1: $M = 21.0$ minutes, $SD = 4.0$ minutes; E2: $M = 9.7$, $SD = 4.0$; E3: $M = 8.8$, $SD = 6.0$; C: $M = 10.2$, $SD = 3.7$; t values ranging from 3.95 to 6.05, $p < .01$). However, also in the C condition some time was spent on educational software on multiplicative reasoning. This should be kept in mind when interpreting our results: we compare a curriculum including the mini-games intervention with a curriculum in which this intervention is not included, but which does include some working with other educational software on multiplicative reasoning.

2.4 Measurement instruments

In the current study, three dependent measures were used to assess the students' learning of multiplicative reasoning (see Table 4 for an overview): the Knowledge Test, measuring students' knowledge of multiplication number facts (declarative knowledge); the Skills Test, measuring students' multiplicative operation skills (procedural knowledge); and the Insight Test, measuring students' insight in, or understanding of, multiplicative number relations (conceptual knowledge). These tests were administered both at the end of Grade 2 and at the end of Grade 3, while the Skills Test was also administered as a pretest at the end of Grade 1 (see Figure 1). In addition to administering these three types of multiplicative ability tests, we measured students' general mathematics ability at the end of Grade 1 as a background variable.

[4] We note that the Rekenweb games on which most of the games in the experimental groups intervention program were based, are on a freely available website. This means that students in the control group could have played some of the original Rekenweb games. To check for this, in the teacher questionnaire we asked teachers which mathematics educational software or games were used in class outside the research project. For each game-period, only zero to three of the control group teachers mentioned Rekenweb as an answer to this question. Based on these data, we can assume that playing Rekenweb games by control group students occurred infrequently. Moreover, if some students in the control group played some of the original Rekenweb games, they did this in a different way than the students in the experimental groups (i.e., using the original, non-adapted games, without the accompanying lessons/discussions, and without instruction videos).

Table 4
Overview of the three multiplicative ability tests

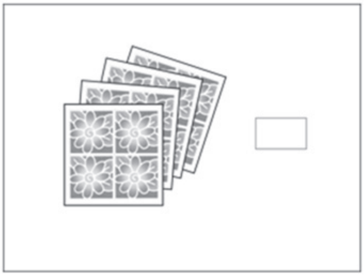
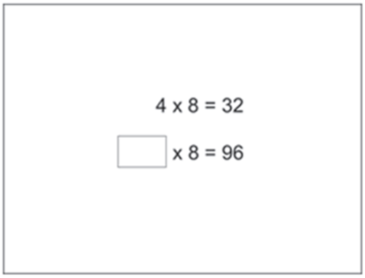
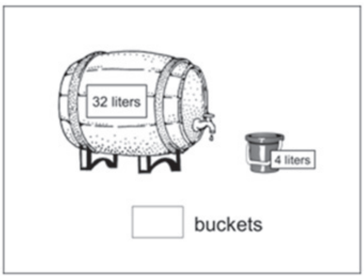
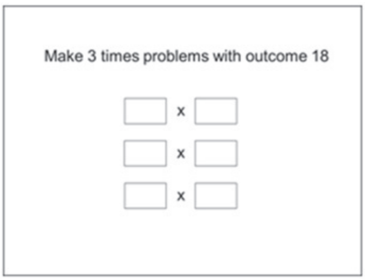
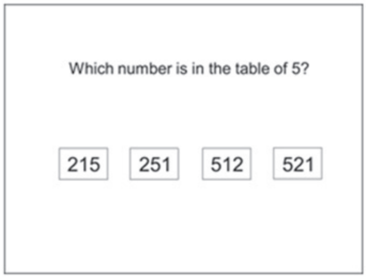
Knowledge Test	Skills Test	Insight Test
What is measured? Ready knowledge of multiplication number facts (declarative knowledge)	*What is measured?* Operation skills in multiplication and division (procedural knowledge)	*What is measured?* Insight in multiplicative number relations, and in properties of multiplicative operations (conceptual knowledge)
Test description Time-limited paper-and-pencil test with bare number multiplication problems.	*Test description* Part of online test. Multiplication and division problems presented with or without a context (no time limit).	*Test description* Part of online test. Non-straightforward problems requiring explicit insight in multiplicative number relations and properties of operations.
Scoring Number correct	*Scoring* Scale scores	*Scoring* Scale scores
Sample items $9 \times 2 = \ldots$ $6 \times 7 = \ldots$ $4 \times 8 = \ldots$	*Sample items*  "Four sheets with four stickers. How many stickers altogether?"	*Sample items*  "Four times eight is 32. How many times eight is 96?"
	 "Four liters of water go in one bucket. The barrel contains 32 liters of water. How many buckets can be filled?"	 "Make three times problems with outcome 18. You are not allowed to make times problems with the number one."

Table continues on next page.

"Two times four is …"



"Which number is in the table of five?"

2.4.1 Knowledge Test

To measure students' ready knowledge of multiplicative number facts, we used the multiplication subtest of the TempoTest Automatiseren (De Vos, 2010), which we refer to as the Knowledge Test. To conceal from the teachers and students in the control group the study's focus on multiplicative reasoning, we also administered the addition and subtraction subtests of this test, but these were not used in our analyses. The multiplication subtest is a time-limited paper-and-pencil test, consisting of a sheet of 50 bare number problems with the × symbol. Students get 2 minutes time to solve as many of the problems as possible; the score is the number of correct answers. The test has a split-half reliability of .96 (De Vos, 2010). As in the Netherlands the × symbol is commonly not introduced yet in Grade 1, the automaticity test was only administered at Measurement point 2 in Grade 2 (Knowledge Test 2) and at Measurement point 3 in Grade 3 (Knowledge Test 3).

2.4.2 Skills and Insight Tests

The Skills and Insight Tests were administered together as an online test. To match the developmental level of the students at the three different measurement points, at each measurement point a different online test was administered (Test 1, Test 2, and Test 3), of which the first test only consisted of a Skills Test. To be able to put the test scores at the different measurement points on a common scale, the tests were linked through anchor items. Each test was piloted at two schools that did not participate in the study.

Composition of the tests. The Skills Tests contained straightforward multiplicative problems, including both bare number problems and problems presented in a context. The Insight Tests consisted of problems in which students had to use their knowledge of multiplication and division at a higher comprehension level. These problems were non-straightforward problems, which required explicit insight in multiplicative relations between numbers (e.g., factors of numbers) and the properties of operations (e.g., the commutative and distributive property). For example, problems were included which were actually beyond the mathematics content taught to the students so far, which means that the students could only solve them by making use of their understanding of multiplicative number relations. Example items of the Skills Tests and the Insight Tests are presented in Table 4.

Besides the multiplicative problems of the Skills and Insight Tests, the online tests also contained some "distractor" items on spatial orientation, addition, and subtraction. These items, which were not used in our analyses, were meant to conceal from the students and teachers in the control group that the focus of the study was on multiplicative reasoning. Table 5 shows the numbers of items of different types in each test. Also the numbers of anchor items are given.

Table 5
Numbers of (anchor) items in the online tests at the three measurement points (Test 1, Test 2, and Test 3), and reliability estimates

	Number of (anchor) items			WLE-reliability		
Item type	Test 1	Test 2[a]	Test 3[b]	Test 1 (n = 689)	Test 2 (n = 665)	Test 3 (n = 694)
Skills items	28	29 (16)	31 (8, 12)	.84	.69	.71
Insight items	-	21 (0)	25 (0, 12)	-	.76	.77
Distractor items	12	16	16			
Total	40	66	72			

[a]Between parentheses is the number of items in Test 2 that were also in Test 1. [b]Between parentheses are, respectively, the number of items in Test 3 that were also in Test 1 and Test 2, and the number of items in Test 3 that were also in Test 2 but not in Test 1.

To control for order effects, for each measurement point four different versions of the online test were constructed. For this purpose the items of each test were organized into clusters. Test 1 contained four clusters of 10 items each, which were presented in different orders in the different test versions. In Test 2, to be able to assess a larger variety of items, including insight items, we decided to use six clusters of 11 items each. Each test version of Test 2 contained four of these clusters. With this design, we could later compute the total score over all 29 items of Skills Test 2 and all 21 items of Insight Test 2, using a Rasch model (see below). The same approach was used for Test 3, in which we used six clusters of 12 items each, with four clusters per test version. The different test versions were randomly assigned to the students.

Test procedure. The online tests were administered through the earlier mentioned Digital Mathematics Environment (DME). The use of an online test facilitated our large-scale data collection and ensured a relatively formal, standardized test setting. Each test item was individually displayed on the screen, and the accompanying question was read aloud by the computer. The tests were administered at school, with the administration organized by the class teacher. The duration of each online test was, on average, approximately 20-30 minutes.

Correction of input errors. Since the text boxes in which the students had to type their answers accepted all kinds of input, not all responses were in the form of a number. Input errors for which it was clear which number was meant, such as “4’0” or “4o” instead of “40”, or “vier” (Dutch for “four”) instead of “4”, were corrected. For Test 1, this resulted in a change to a correct answer for 0.60% of the item answers; for Test 2 and Test 3, this was the case for 0.08% and 0.05% of the item answers, respectively.

Scaling of test scores. Because different tests were administered at the different measurement points, and because at Measurement point 2 and 3 the different versions of the tests contained different subsets of the total set of Skills and Insight items, item response modeling was needed to put the Skills Test and Insight Test scores of the different measurement points and different test versions on a common scale. For the Skills Tests, the items of Skills Test 1, Skills Test 2, and Skills Test 3 were first separately scaled by a Rasch model, employing the Conquest software (Wu, Adams, Wilson, & Haldane, 2007). By this procedure, the students’ raw test scores were converted into scale scores (weighted likelihood estimates, or WLE) for each test. Subsequently, to put all three Skills Test scores on a common scale, we employed mean-mean linking (Kolen & Brennan, 2004), with the assumption that (for equal student ability) the item difficulties of the anchor items were equal on average in the different tests. The same procedure was employed for the two Insight Tests.

Reliability. Table 5 presents the WLE reliability estimates (Wu et al., 2007) of the scale scores of the Skills Tests and Insight Tests, which can be interpreted in the same way as Cronbach’s alpha. The Skills and Insight Tests can be considered adequately reliable, although the reliability of Skills Test 2 is just below the .70 boundary. Remaining unreliability was accounted for in our analyses (see section 2.7).

2.4.3 General mathematics ability test

Students’ general mathematics ability was measured at Measurement point 1 using the end Grade 1 mathematics test from the Cito student monitoring system (see Janssen, Verhelst, Engelen, & Scheltens, 2010). This test, which we refer to as GMath, was administered by the participating schools as part of their regular testing program, either as a paper-and-pencil test or as a digital test. The reliability of this test is .91 for the paper-and-pencil version, which was used in most schools, and .94 for the digital version (Janssen et al., 2010).

2.5 Treatment of missing data

As is inevitable in a large-scale longitudinal study carried out in real school practice, not all data were available for all students. Missing data were caused by students who missed a test. The percentage of missing scores ranged from 2.0% to 8.1% per test. To make estimates for the missing test scores, we employed multiple data imputation (see Graham, 2009). We specified an imputation model involving student background data, test scores, and, for the students in the experimental conditions, the gameplay data. Because the gameplay data can be expected to have a different relation with the learning outcomes in the different conditions, the imputation procedure was performed for each condition separately. To account for the clustered data structure (students nested within schools), we also included school mean test scores as predictors in the imputation model. The data imputation was run using the "mice" software (Van Buuren & Groothuis-Oudshoorn, 2011), and resulted in 50 imputed datasets. Statistical analyses were performed on these 50 datasets and results were combined using Rubin's rule (see Graham, 2009).

2.6 Initial differences in student characteristics between conditions

Although we employed blocking and random assignment to distribute the participating schools over the four research conditions, differences between groups may have arisen with respect to their student composition. Therefore, after data-imputation, we examined whether there were differences between the conditions with respect to students' gender, age (students with a grade-appropriate age vs. older, delayed students), parental education (higher vs. lower education[5]), and home language (monolingual Dutch vs. other), and their Grade 1 scores on general mathematics ability (GMath) and multiplicative reasoning ability (Skills Test 1). As is shown in Table 6, we found a significant difference between groups for gender and a marginally significant difference ($p = .088$) for age. In addition to gender and age, effect sizes were non-trivial ($\eta^2 \geq .01$) for parental education, home language, and GMath score. To be conservative, we decided in all analyses to control for gender (dummy variable Female), age (dummy variable AgeDelayed), parental education (dummy variable ParEdLow), home language (dummy variable NonDutch), and GMath score.

[5] Higher means that at least one parent has completed secondary education; lower means that none of the parents has completed secondary education.

Table 6
Initial differences in student characteristics between conditions

Condition	n	Gender % female	Age % not delayed	Parental education % secondary	Home language % Dutch	GMath score M (SD)	Skills Test 1 score M (SD)
C	327	50.2	93.9	93.0	94.5	46.8 (14.7)	0.09 (1.36)
E1	112	41.1	88.4	86.6	85.7	45.0 (16.6)	0.09 (1.33)
E2	202	50.5	88.1	87.1	93.6	42.9 (15.0)	-0.12 (1.14)
E3	78	35.9	80.8	93.6	89.7	43.5 (12.8)	-0.17 (1.47)
Total	719	47.3	90.0	90.4	92.4	45.1 (15.1)	0.00 (1.31)
Wald $\chi^2(3)$		*10.15**	*6.54*†	*3.08*	*1.45*	*1.03*	*1.44*
η^2		*.011*	*.019*	*.011*	*.014*	*.018*	*.007*

Note. Wald tests (comparable to one-way ANOVAs) were performed in Mplus, using cluster-robust standard errors (see section 2.7). The η^2 effect sizes were calculated based on regular ANOVA results.
† $p < .10$. * $p < .05$.

2.7 Data analysis

We analyzed our data using path analysis in Mplus (Muthén & Muthén, 1998-2010). Path analysis was used to be able to simultaneously study the effects of the intervention in Grade 2 and the intervention in Grade 3, as well as their combined effect.[6]

The path model we used in answering research questions 1 to 3 is displayed in Figure 3. This model can be interpreted as testing two ANCOVAs simultaneously, one with the Grade 2 score as the dependent variable, and one with the Grade 3 score as the dependent variable. As is shown in Figure 3, as predictors we used three dummy variables for the three experimental conditions (the control condition was modeled as the reference category). As covariates we used the pretest score (Skills Test 1) and the covariates related to initial differences between conditions (Female, AgeDelayed, ParEdLow, NonDutch, and GMath). The model was separately specified for the three aspects of multiplicative reasoning ability – knowledge, skills, and insight – measured by the Knowledge Tests, Skills Tests, and Insight Tests, respectively. In addition, we specified a joint model in

[6] As we had dependent variables at two time points (end Grade 2 and end Grade 3), a regular ANCOVA approach was not appropriate. Repeated measures ANOVA would have been a possibility, but we decided not to use such an analysis, as we had three measurement points for the Skills Test, but only two measurement points for the Knowledge Test and the Insight Test. Using repeated measures ANOVA, then, would have led to different, incomparable analyses for the different tests of multiplicative reasoning ability.

which the standardized paths for the three types of multiplicative ability tests were constrained equal, to test the effect of the mini-games interventions averaged over the three aspects of multiplicative reasoning ability. This joint model can be seen as testing the effect of the games on students' overall multiplicative reasoning ability. For answering research questions 4 to 6, the model in Figure 3 was extended by adding predictors and interactions, as we will explain later.

In the model, an arrow from a condition variable to Grade 2 score represents the direct effect of the particular condition on the Grade 2 score, that is, the effect of the Grade 2 intervention in that condition on the score at the end of Grade 2. Similarly, an arrow from a condition variable to Grade 3 score represents the direct effect of the Grade 3 intervention in this condition on the Grade 3 score. In addition to these direct effects, we also examined the indirect effect of the Grade 2 intervention, via Grade 2 score, on Grade 3 score (computed as the standardized product of the raw paths from condition variable to Grade 2 score and from Grade 2 to Grade 3 score), and the total effect of the interventions in Grade 2 and Grade 3 on Grade 3 score (computed as the standardized sum of the raw direct and indirect effect of condition on Grade 3 score). The indirect effect can be interpreted as

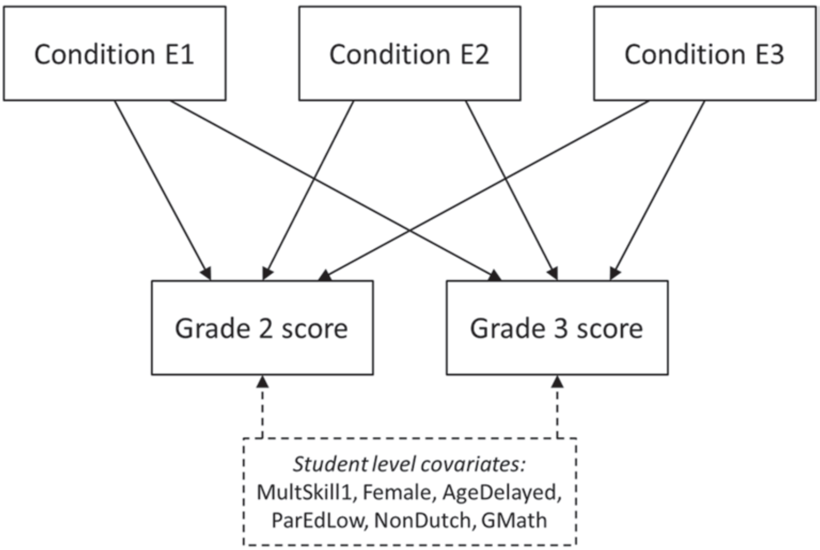


Figure 3. Path model used for comparing the three experimental conditions to the control group (reference category).

a long-term effect of the Grade 2 intervention; the total effect can be seen as the effect of the combined Grade 2-3 intervention.[7] [8]

To account for the clustered data structure, we employed cluster-robust standard errors in our analyses (see Angrist & Pischke, 2009), using the TYPE = COMPLEX option in Mplus. As the level of clustering we used the school, because random assignment to conditions was done at the school level and because participating classes within schools were sometimes merged or split up in the course of the research project. We do not expect large differences compared to an approach using class as the level of clustering, as in 80% of the schools only one class was participating. To control for unreliability in the Skills and Insight Test scores, these scores were modeled as latent variables, with their residual variance fixed at *[1 – reliability of test scores]*variance of test scores* (see Hayduk, 1987).

For all analyses, we report standardized or partially standardized path coefficients, which can be interpreted as effect size measures. For continuous predictors (e.g., prior mathematics ability), the standardized coefficient β represents the amount of change in standard deviation units in the dependent variable associated with a one standard deviation change in the predictor variable (STDYX in Mplus). This coefficient is practically equivalent to an *r* effect size, for which the values .10, .30, and .50 can be interpreted as a small, medium, and large effect, respectively (Cohen, 1988). For dummy (binary) predictors (e.g., the condition variables), we employed partially standardized coefficients β_{ps}, which represent the amount of change in standard deviation units in the dependent variable associated with a change in the dummy predictor from 0 to 1 (STDY in Mplus). β_{ps} is thus practically equivalent to a *d* effect size of the difference between the 0 and 1 category (interpretation guidelines: 0.20 (small effect), 0.50 (medium effect), 0.80 (large effect), see Cohen, 1988). For completeness, for the significant effects we also provide regular *r* or *d* values, which are approximately equal to the β and β_{ps} values, respectively.[9]

---

[7] Indirect and total effects were tested using the Delta method as implemented in Mplus (Muthén & Muthén, 1998-2010). Although we are aware that this method can be rather conservative, other methods like Bootstrap cannot easily be adapted for clustered samples.

[8] In accordance with the usual terminology for path analysis, the parameters of these models are called effects. This does not necessarily imply that these parameters can be interpreted as causal effects. Whether effects are causal depends on the research design: in our case, the effects of the condition variables can be seen as causal effects.

[9] The *d* values were calculated by dividing the raw coefficients by the pooled standard deviation of the dependent variable; the *r* values were calculated by multiplying the raw coefficient by the standard deviation of the predictor, and then dividing by the pooled standard deviation of the dependent variable. For Skills Test scores and Insight Test scores, in accordance with the abovementioned reliability correction (Hayduk, 1987), we used a standard deviation adjusted for unreliability: $SD = \sqrt{[variance*reliability]}$. Note: in cases were Mplus did not provide (partially) standardized coefficients (for indirect and total effects and for paired comparisons between coefficients), we computed them using the same formulas as we used for computing *d* and *r* values. In these cases, thus, the reported *d* and *r* values are by definition equal to the β_{ps} and β values, respectively.

Dependent of whether our hypotheses were directional or not, we use one-tailed or two-tailed significance tests. When multiple equalities were tested at once (e.g., in a Wald test), two-tailed tests were used.

3 Results

3.1 Effects of the interventions on overall multiplicative reasoning ability

The means and standard deviations of the scores on the multiplicative ability tests administered at the three measurement points are reported in Table 7 (correlations are included in Appendix A). To investigate the effects of the mini-games interventions in the three experimental conditions on students' multiplicative reasoning ability we used the path model displayed in Figure 3. We first examined the effects of the interventions on students' overall multiplicative reasoning ability, that is, the effects averaged over the three aspects of multiplicative ability. This was done by means of a joint model (as mentioned above) in which the paths for the three aspects of multiplicative ability were constrained equal. The direct, indirect, and total effects in this joint model are presented in Table 8 (first columns). For the E3 condition (playing at home with debriefing at school) we found a significant total effect of the Grade 2-3 intervention on overall multiplicative reasoning ability at the end of Grade 3 ($\beta_{ps} = 0.22$, $d = 0.22$), as compared to the control group (the reference category). For the E1 condition (playing at school) and the E2 condition (playing at home without attention at school), we did not find significant effects in this joint model ($p > .10$).

3.2 Effects of the interventions on the three aspects of multiplicative reasoning ability

In addition to the effects on overall multiplicative reasoning ability we investigated the separate effects of the interventions on each of the three aspects of multiplicative reasoning ability: knowledge, skills, and insight. The model results are presented in Table 8.

Regarding knowledge, we found no significant effects of the interventions ($p > .10$): the experimental group students did not differ from the control group students in their knowledge scores. With respect to skills, there was a significant total effect of the E3 intervention in Grade 2-3 on students' Grade 3 scores ($\beta_{ps} = 0.26$, $d = 0.26$). For the E1 and the E2 condition, no significant effects were found regarding skills ($p > .10$).

Also with respect to insight, the E3 intervention was found to be effective. First of all, the direct effect of the E3 intervention in Grade 2 on students' insight scores at the end of Grade 2 was significant ($\beta_{ps} = 0.32$, $d = 0.29$). Also the total effect of the Grade 2-3 intervention in E3 was found to be significant ($\beta_{ps} = 0.22$, $d = 0.22$). Besides effects of the E3 intervention, for multiplicative insight we also found an effect of the E1 intervention (playing at school): the E1 intervention in Grade 2 had a significant direct effect on insight

Table 7

Means (standard deviations) of all multiplicative ability test scores

		Grade 1	Grade 2			Grade 3		
Condition	*n*	Skills Test 1	Knowledge Test 2	Skills Test 2	Insight Test 2	Knowledge Test 3	Skills Test 3	Insight Test 3
C	327	0.09 (1.36)	20.50 (9.66)	2.40 (1.46)	-0.03 (1.48)	31.52 (12.08)	3.50 (1.32)	1.52 (1.38)
E1	112	0.09 (1.33)	20.15 (8.56)	2.41 (1.31)	0.31 (1.44)	28.60 (11.93)	3.50 (1.27)	1.60 (1.33)
E2	202	-0.12 (1.14)	18.22 (7.55)	2.12 (1.38)	-0.05 (1.51)	29.78 (12.02)	3.29 (1.48)	1.24 (1.50)
E3	78	-0.17 (1.47)	20.45 (7.75)	2.38 (1.30)	0.12 (1.37)	32.28 (10.48)	3.57 (1.25)	1.53 (1.28)
Total	719	0.00 (1.31)	19.80 (8.80)	2.32 (1.40)	0.03 (1.48)	30.66 (11.94)	3.45 (1.36)	1.45 (1.40)

Note. The scores on the Skills Tests and Insight Tests are scale scores; the scores on the Knowledge Tests are number correct scores (see section 2.4).

Table 8

Direct, indirect, and total effects of conditions E1, E2, and E3 on overall multiplicative reasoning ability and on the three aspects of multiplicative reasoning ability (as compared to the control group)

	Overall multiplicative reasoning ability[a]		Aspect of multiplicative reasoning ability[b]					
			Knowledge		Skills		Insight	
Effect	β_{ps}	*SE*	β_{ps}	*SE*	β_{ps}	*SE*	β_{ps}	*SE*
Condition E1								
Direct effect on Grade 2 score	0.18	0.19	0.01	0.24	0.10	0.24	0.39*	0.22
Direct effect on Grade 3 score	-0.10	0.11	-0.20	0.17	0.02	0.15	-0.07	0.14
Indirect effect	0.11	0.12	0.01	0.15	0.07	0.16	0.19†	0.12
Total effect	0.01	0.18	-0.20	0.23	0.09	0.16	0.13	0.17
Condition E2								
Direct effect on Grade 2 score	0.00	0.15	-0.16	0.23	-0.04	0.20	0.21†	0.15
Direct effect on Grade 3 score	0.01	0.08	0.05	0.11	0.06	0.15	-0.14	0.13
Indirect effect	0.00	0.09	-0.10	0.15	-0.02	0.13	0.10	0.08
Total effect	0.01	0.10	-0.05	0.16	0.03	0.13	-0.02	0.11
Condition E3								
Direct effect on Grade 2 score	0.19	0.17	0.08	0.26	0.21	0.20	0.32*	0.19
Direct effect on Grade 3 score	0.11	0.09	0.11	0.13	0.13	0.16	0.07	0.13
Indirect effect	0.12	0.11	0.05	0.16	0.14	0.14	0.16†	0.10
Total effect	0.22**	0.09	0.16	0.13	0.26*	0.15	0.22*	0.11

Note. SkillsTest1, Female, AgeDelayed, ParEdLow, NonDutch, and GMath were included as covariates (see Figure 3). β_{ps} = partially standardized coefficient.
[a]These are the effects averaged over the three aspects of multiplicative reasoning ability (standardized paths of the path models of the three aspects constrained to be equal). [b]The model was separately specified for each of the three aspects of multiplicative reasoning ability.
† $p < .10$. * $p < .05$. ** $p < .01$. One-tailed.

in Grade 2 ($\beta = 0.39$, $d = 0.35$). For the E2 intervention (playing at home without attention at school) we only found a marginally significant effect.

The above results suggest that the games were most effective in enhancing students' multiplicative insight. To test whether there was indeed a difference between the three aspects of multiplicative reasoning ability in the extent to which they were affected by the mini-games interventions, we simultaneously ran the path models of the three aspects of multiplicative ability, and tested whether partially standardized path coefficients differed between the three aspects. For the E1 and the E2 condition some significant differences between multiplicative ability aspects were found. Both for E1 and E2, the intervention in Grade 2 was more effective in enhancing insight than in enhancing skills (E1: $\Delta\beta_{ps} = 0.26$, $p < .05$, $d = 0.26$; E2: $\Delta\beta_{ps} = 0.22$, $p < .05$, $d = 0.22$).[10] Furthermore, in the E1 condition, the combined Grade 2-3 intervention was more effective in enhancing insight than number fact knowledge ($\Delta\beta_{ps} = 0.33$, $p < .05$, $d = 0.33$). For the E3 condition, there were no significant differences in effectiveness for different aspects of multiplicative ability.

3.3 Comparisons between the three game-playing settings

To statistically test the difference between the three experimental interventions in their effectiveness as compared to the control group, we compared the path coefficients of the three condition variables in the model of overall multiplicative reasoning ability (effects averaged over the three aspects of multiplicative ability). We used Wald χ^2 tests (comparable to one-way ANOVAs), and pair-wise comparisons between the path coefficients. None of the Wald test results were significant ($p > .10$), but some of the paired comparisons were, as is shown in Table 9. We found a significant difference between E3 and E1, in favor of E3, for the direct effect of the Grade 3 intervention ($\Delta\beta_{ps} = 0.21$, $d = 0.21$). Furthermore, we found that the total effect of the Grade 2-3 intervention was significantly higher in the E3 condition than in the E2 condition ($\Delta\beta_{ps} = 0.21$, $d = 0.21$). Also when we looked at the three aspects of multiplicative ability separately (see Appendix B), we found several significant differences indicating that the E3 intervention was more effective than the E1 and the E2 intervention, while there were no differences between the E1 and the E2 intervention.

3.4 Relations between gameplay behavior and learning outcomes

To investigate the role of students' gameplay behaviour in the effectiveness of the mini-games interventions, we added to the path model in Figure 3 interactions between the condition variables and the variables Gplay2 and Gplay3 (as defined in section 2.3.4), as

[10] $\Delta\beta_{ps}$ denotes the difference in partially standardized coefficients.

Table 9

Paired comparisons between the E1, E2, and E3 condition of direct, indirect, and total effects on overall multiplicative reasoning ability[a]

	Comparison					
	E1 – E2		E3 – E2		E3 – E1	
Effect	$\Delta\beta_{ps}$	*SE*	$\Delta\beta_{ps}$	*SE*	$\Delta\beta_{ps}$	*SE*
Direct effect on Grade 2 score	0.18	0.17	0.19	0.15	0.01	0.19
Direct effect on Grade 3 score	-0.11	0.11	0.10	0.09	0.21*	0.11
Indirect effect	0.11	0.11	0.12	0.09	0.01	0.12
Total effect	0.00	0.17	0.21**	0.09	0.22	0.17

Note. E1: $n = 112$; E2: $n = 202$; E3: $n = 78$. SkillsTest1, Female, AgeDelayed, ParEdLow, NonDutch, and GMath were included as covariates (see Figure 3). $\Delta\beta_{ps}$ = difference in partially standardized coefficients.

[a]Averaged over the three aspects of multiplicative reasoning ability (standardized paths of the path models of the three aspects constrained to be equal).

* $p < .05$. ** $p < .01$. Two-tailed for the E1-E2 comparison, one-tailed for the E3-E2 and the E3-E1 comparison (because of our directional hypothesis).

predictors of Grade 2 score and Grade 3 score.[11] In this way we could measure for each of the experimental conditions the influence of gameplay behavior (time and effort spent on the games) on the learning effects of the mini-games. The paths from the Gplay*Condition interactions to the test scores for the three aspects of multiplicative reasoning ability are presented in Table 10 (correlations between all variables in the model are included in Appendix A).

Significant influences of gameplay behavior on test scores were found for the E1 and the E3 condition; for the E2 condition only some marginally significant influences were found. In the E1 condition, gameplay in Grade 2 was a significant predictor of the knowledge scores ($r = .25$) and the insight scores ($r = .17$). In the E3 condition, gameplay in Grade 2 was a significant predictor of the skills scores ($r = .21$), whereas gameplay in Grade 3 significantly predicted insight scores ($r = .12$).

[11] The possible influence of Gplay2 on Grade 3 test scores was controlled for in the model.

Table 10

Interactions of gameplay with condition variables predicting Grade 2 and Grade 3 test scores on the three aspects of multiplicative reasoning ability

	Aspect of multiplicative reasoning ability[a]								
	Knowlegde			Skills			Insight		
Effect	β	*SE*	*r*	β	*SE*	*r*	β	*SE*	*r*
Condition E1									
Effect of Gplay2 on Grade 2 score	.39*	.20	.25	.20	.18	.10	.35*	.22	.17
Effect of Gplay3 on Grade 3 score	.03	.14	.02	.07	.10	.04	-.06	.16	-.03
Condition E2									
Effect of Gplay2 on Grade 2 score	.01	.07	.01	.09†	.05	.10	.07†	.05	.08
Effect of Gplay3 on Grade 3 score	.02	.05	.01	-.06	.08	-.03	-.06	.07	-.03
Condition E3									
Effect of Gplay2 on Grade 2 score	.03	.04	.09	.09***	.03	.21	.01	.02	.02
Effect of Gplay3 on Grade 3 score	.05†	.04	.15	.01	.05	.02	.06*	.03	.12

Note. $N = 719$. SkillsTest1, Female, AgeDelayed, ParEdLow, NonDutch, and GMath were included as covariates (see Figure 3). Because of the large differences in gameplay behavior between conditions, r effect sizes were computed using per-condition standard deviations.
[a]The model was separately specified for each of the three aspects of multiplicative reasoning ability.
† $p < .10$. * $p < .05$. *** $p < .001$. One-tailed.

3.5 Influence of gender and prior mathematics ability on the effects of the interventions

To test whether the effects of the interventions in the different experimental conditions were moderated by gender and prior mathematics ability, we added to the model in Figure 3 interactions of Female and GMath with the condition variables. Using these interactions, we could examine the influence of gender and prior mathematics ability on the earlier reported direct, indirect, and total effects of the interventions. Table 11 displays the results for the model of overall multiplicative reasoning ability (effects averaged over the three tests).

Table 11

Interactions of gender and prior mathematics ability with condition variables predicting direct, indirect, and total effects on overall multiplicative reasoning ability[a]

	Interaction with			
	Gender (Female)		Prior mathematics ability (GMath)	
Effect	β_{ps}	*SE*	β	*SE*
Condition E1				
Direct effect on Grade 2 score	-0.28†	0.17	-.02	.10
Direct effect on Grade 3 score	0.06	0.13	-.03	.07
Indirect effect	-0.17†	0.10	-.01	.06
Total effect	-0.11	0.14	-.05	.09
Condition E2				
Direct effect on Grade 2 score	-0.20†	0.12	.07	.07
Direct effect on Grade 3 score	0.07	0.10	.13*	.06
Indirect effect	-0.12†	0.07	.05	.04
Total effect	-0.06	0.11	.17*	.07
Condition E3				
Direct effect on Grade 2 score	-0.15	0.18	-.06	.09
Direct effect on Grade 3 score	0.06	0.13	-.01	.11
Indirect effect	-0.09	0.11	-.04	.06
Total effect	-0.04	0.12	-.05	.15

Note. $N = 719$. SkillsTest1, Female, AgeDelayed, ParEdLow, NonDutch, and GMath were included as covariates. β_{ps} = partially standardized coefficient (a positive value signifies a female advantage).
[a]Averaged over the three aspects of multiplicative reasoning ability (standardized paths of the path models of the three aspects constrained to be equal).
† $p < .10$. * $p < .05$. Two-tailed.

Regarding gender, for overall multiplicative reasoning ability no significant moderating effects were found. We did, however, find some marginally significant gender effects for the E1 and E2 interventions in Grade 2, all in favor of boys. The results for the three different aspects of multiplicative reasoning ability (see Appendix C) revealed several significant gender differences in favor of boys for the effectiveness of the Grade 2 intervention. The E1 and E2 intervention in Grade 2 were more effective for boys than for girls in enhancing multiplicative insight (direct effect E1: $\beta_{ps} = -0.57$, $p < .05$, $d = -0.45$; indirect effect E1: $\beta_{ps} = -0.28$, $p < .05$, $d = -0.28$; direct effect E2: $\beta_{ps} = -0.37$, $p < .05$, $d = -0.29$; indirect effect E2: $\beta_{ps} = -0.18$ $p < .05$, $d = -0.18$), and the E3 intervention in Grade 2 was more effective for boys than for girls in enhancing multiplicative fact knowledge (direct effect: $\beta_{ps} = -0.47$, $p < .01$, $d = -0.48$; indirect effect: $\beta_{ps} = -0.30$ $p < .01$, $d = -0.30$). In contrast, in Grade 3 the E1 intervention was more effective for girls than for boys in promoting insight ($\beta_{ps} = 0.39$, $p < .05$, $d = 0.30$). Overall, then, the mini-games interventions in Grade 2 were found to be more effective for boys than for girls, whereas this difference disappeared, and occasionally reversed, for the Grade 3 interventions.

When investigating the moderating effect of prior mathematics ability, we only observed some small effects for the E2 condition (playing at home without attention at school). For overall multiplicative reasoning ability (see Table 11), we found that the E2 intervention in Grade 3 was more effective for students with a higher prior mathematics ability ($\beta = .13$, $r = .13$), and this also was the case for the combined Grade 2-3 intervention ($\beta = .17$, $r = .17$). Also for the different aspects of multiplicative reasoning ability (see Appendix C) we only found significant effects of prior mathematics ability for the E2 condition (total effect on skills: $\beta = .22$, $p < .05$, $r = .22$; direct effect on insight in Grade 2: $\beta = .28$, $p < .05$, $r = .16$; indirect effect on insight: $\beta = .10$, $p < .05$, $r = .10$). In sum, the E2 intervention was found to be more effective for students with higher prior mathematics ability. In fact, when we reran the path analyses on the effectiveness of the games with only the students with above-average GMath scores ($n = 341$), we did find a significant effect of the E2 intervention in Grade 2 on insight (direct effect: $\beta_{ps} = 0.36$, $p < .05$, $d = 0.35$; indirect effect: $\beta_{ps} = 0.19$, $p < .05$, $d = 0.19$), and a significant effect of the E2 intervention in Grade 3 on multiplicative fact knowledge ($\beta_{ps} = 0.27$, $p < .05$, $d = 0.27$) and on overall multiplicative reasoning ability ($\beta_{ps} = 0.19$, $p < .05$, $d = 0.19$).

3.6 Influence of gender and prior mathematics ability on gameplay behavior

Possibly, some of the moderating effects we found of the student characteristics gender and prior mathematics ability on the effects of the interventions can be explained by differences in gameplay behavior between students with different characteristics. To test whether gender and prior mathematics ability predicted gameplay behavior in the three experimental conditions, we performed, for the students in the experimental conditions ($n = 392$), linear regression analyses with the gameplay variables Gplay2 and Gplay3 as the dependent variables and Female*Condition and GMath*Condition interactions as the predictors.

AgeDelayed, ParEdLow, and NonDutch were included as covariates. The results of these regression analyses, displayed in Table 12, showed that in the E1 condition (playing at school), gender and prior mathematics ability did not influence gameplay behavior ($p > .10$), which is probably due to the fact that in the E1 condition the teachers told the students how long to play the games. In the home-playing conditions E2 and E3 we did find relations between student characteristics and gameplay behavior. Regarding gender, girls were found to play the games more than did boys in both E2 and E3 and in both grades (*d*s ranging from 0.39 to 0.72). With respect to prior mathematics ability, in Grade 2 we found that students with higher prior mathematics ability played the games more, both in E2 ($r = .16$) and in E3 ($r = .19$). In Grade 3, no such relation was found.

Table 12.
Interactions of gender and prior mathematics ability with condition variables predicting gameplay

	Interaction with					
	Gender (Female)			Prior mathematics ability (GMath)		
Effect	β_{ps}	*SE*	*d*	β	*SE*	*r*
Condition E1						
Effect on Gplay2	0.02	0.06	0.09	.05	.03	.14
Effect on Gplay3	-0.08	0.05	-0.22	.06	.06	.12
Condition E2						
Effect on Gplay2	0.28**	0.09	0.39	.19*	.08	.16
Effect on Gplay3	0.13***	0.03	0.50	-.01	.04	-.02
Condition E3						
Effect on Gplay2	0.57**	0.20	0.72	.22*	.11	.19
Effect on Gplay3	0.28*	0.12	0.54	.03	.03	.04

Note. $n = 392$. AgeDelayed, ParEdLow, and NonDutch were included as covariates. β_{ps} = partially standardized coefficient (a positive value signifies a female advantage). Because of the large differences in gameplay behavior between conditions, *d* and *r* effect sizes were computed using per-condition standard deviations.
* $p < .05$. ** $p < .01$. *** $p < .001$. Two-tailed.

4 Discussion

This study aimed at investigating the effects of three different ways of deploying multiplicative mini-games on second- and third-graders' multiplicative reasoning ability.

We examined the effects of the games on students' overall multiplicative reasoning ability, and on three different aspects of this multiplicative reasoning ability. Furthermore, we examined the role of gameplay behavior, gender, and prior mathematics ability in the effectiveness of the mini-games. In the following sections, we address each of our six research questions. After that, we discuss the generalizability of our findings, we mention some limitations of our study, and we present our main conclusions.

4.1 Effects of multiplicative mini-games on students' multiplicative reasoning ability

Our first research question was whether an intervention with multiplicative mini-games positively affects students' overall multiplicative reasoning ability. We found that when mini-games were played at home and afterwards debriefed at school (E3), they were effective in enhancing students' overall multiplicative reasoning ability. Specifically, we found a small positive effect of the combined Grade 2-3 intervention ($d = 0.22$). In other words, averaged over the three aspects of multiplicative reasoning ability measured in our study, students in the E3 condition outperformed the control group students at the end of Grade 3 (controlling for the covariates). This finding contributes to the still relatively sparse knowledge on the educational effectiveness of (mathematics) computer games (e.g., Bai et al., 2012; Wouters et al., 2013).

Our second research question focused on the effects on the three different aspects of multiplicative reasoning ability – number fact *knowledge*, operation *skills*, and *insight* in multiplicative number relations. Here we found that the games were effective in enhancing skills and insight, but not knowledge. In particular, when the games were played at home and debriefed at school (E3), they affected both skills and insight, and both the Grade 2 intervention and the combined Grade 2-3 intervention were found to be effective (significant *d*s ranging from 0.22 to 0.29). When the games were played at school (E1), they only affected insight, and only the Grade 2 intervention was effective ($d = 0.35$). No significant effects were found when the games were played at home with no attention at school (E2). When we compared the effectiveness of the games on the three aspects of multiplicative ability, we found some evidence that in the E1 and E2 condition the games were more effective in enhancing insight than in enhancing skills and knowledge. This is in line with the finding from Ke's (2009) review that games seem to promote higher-order thinking more than factual knowledge acquisition. Computer games may be especially useful for the teaching of insight when they allow for free exploration and experimentation, as was the case in our study. Such an approach differs from games that contain adaptive features for matching mathematics problems to individual students' ability levels, in which case students cannot control which problems to solve. However, these latter games might be better for automatizing fact knowledge (see, e.g., Jansen et al., 2013).

A remarkable finding of our study was that effects were primarily found for the Grade 2 intervention (in E3 effects were also found for the combined Grade 2-3 intervention, but we did not find effects of the Grade 3 intervention alone). This may be explained by the stage of the students' learning process. In Grade 2, students are at the beginning of learning multiplicative reasoning, which may imply that there is more room for improvement than in Grade 3. Another possible explanation is the occurrence of a novelty effect (e.g., Li & Ma, 2010). Students, as well as teachers, may be more motivated to put attention into the games when they are new for them. This explanation is supported by our finding that the games were played more in Grade 2 than in Grade 3.

4.2 Effects of multiplicative mini-games in different settings

As an answer to our third research question, we found evidence that the mini-games were more effective when they were played at home and debriefed at school (E3) than when they were played at school (E1) or played at home with no attention at school (E2) (significant *d*s of 0.21). This finding was as expected and can be explained by the E3 setting having the advantage of playing at home (extended learning time, more learner control) as well as the advantage of an in-class intervention (the debriefing sessions). Another explanation of the higher effectiveness of the E3 intervention as compared to the E2 intervention may lie in the amount of time spent on the mini-games, which was higher in condition E3 than in E2. This means that, apart from having a reflective role, the debriefing sessions in the E3 condition may also have functioned as an encouragement for the students to play the games at home.

4.3 The role of gameplay behavior

Our fourth research question focused on the relation between students' gameplay behavior (time and effort put in the games) and the effects of the interventions. Regarding multiplicative fact knowledge and skills, we found an influence of gameplay in E1 for multiplicative fact knowledge ($r = .25$), and in E3 for multiplicative skills ($r = .21$). These findings are as expected and are in line with Jansen et al.'s (2013) finding that more practice leads to more automatization. However, it is unclear why the other relations between gameplay and learning effects on number fact knowledge and skills, for example in Grade 3, were not significant.

Regarding multiplicative insight, we did not necessarily expect a relation between gameplay behavior and learning effects, as once the learning concepts in a game are discovered, more gameplay may not lead to more learning (Kolovou et al., 2013). Yet, in contrast with the finding of Kolovou et al. (2013), we did find a positive influence of gameplay behavior on the learning effect on multiplicative insight in the E1 condition in Grade 2 ($r = .17$), and in the E3 condition in Grade 3 ($r = .12$). This may be explained by the fact that in our study more gameplay not only meant that more time and effort was spent on particular games, but also that more different games were played, which differs from the Kolovou et al. study, in which there was only one mini-game. As in different games different concepts were embedded, playing more different games may have led to more development of insight.

Finally, the fact that in the E2 condition no significant relations were found between gameplay and learning effects suggests that playing the games alone was not sufficient for promoting learning, but reflection, for example in the form of a debriefing session, was necessary. This corresponds to the importance of debriefing as indicated by several researchers (e.g., Egenfeldt-Nielsen, 2005; Garris et al., 2002; Klawe, 1998).

4.4 The role of gender

Regarding our fifth research question, about the role of gender in the effects of the mini-games interventions, we found evidence that in Grade 2 the mini-games interventions were somewhat more effective for boys than for girls. In this grade, for overall multiplicative reasoning ability there were only marginally significant differences between boys and girls, while significant differences were found for number fact knowledge (in E3), and especially for multiplicative insight (in E1 and E2) (*d*s ranging from -0.18 to -0.45). In Grade 3 the interventions were generally equally effective for boys and girls, but an occasional advantage for girls was found (for the effect of the E1 intervention on insight).

A further analysis showed that the gender difference in effectiveness of the games in Grade 2 cannot be explained by gameplay behavior. For example, in the E2 and E3 interventions in Grade 2, whereas girls appeared to have profited less from the games, they played the games more than did boys.

Our general finding that the games were more effective for boys in Grade 2, while there was generally no difference in Grade 3, may be related to boys' possible higher initial experience with and motivation for ICT and computer games (e.g., Bourgonjon et al., 2010; Lowrie & Jorgensen, 2011). Girls may need more time to get used to working with mathematics computer games before they profit from them to the extent boys do. Our finding relates to the finding by De Jean et al. (1999) that many girls did not spontaneously see the mathematics embedded in a game. One may argue that the mathematics content related to the insight domain, in which we found most gender differences, is most hidden in the games and is thus, possibly, less well found by girls than by boys.

4.5 The role of prior mathematics ability

Our sixth research question was about the role played by students' prior mathematics ability in the effectiveness of the interventions. We found that in the E1 and E3 condition the effect of the intervention was not influenced by students' initial mathematics ability. This result corresponds to findings of earlier studies on the effects of mathematics computer games (e.g., Habgood & Ainsworth, 2011; Kebritchi et al., 2010). However, we did find some evidence for the E2 intervention – playing at home without attention at school – being more effective for students with higher prior mathematics ability (*r*s ranging from .10 to .22). This may partly be related to our finding that, in Grade 2, in the E2 condition (as well as in the E3 condition) students with higher prior mathematics ability played the games more, possibly because they were more motivated or because they understood the games better. Moreover, higher ability students may less require teacher guidance for learning from their game playing (Kebritchi et al., 2010), and thus may profit more from an intervention without teacher attention. In fact, we found evidence that, although not effective in general, the E2 intervention was effective for students with above-average prior mathematics ability: for these students the E2 intervention in Grade 2 positively affected insight ($d = 0.35$), while the E2 intervention in Grade 3 positively affected multiplicative fact knowledge ($d = 0.27$).

4.6 Generalizability of our findings

It should be noted that our findings apply only to the use of multiplicative mini-games in Grade 2 and Grade 3 of primary school, in three specific instructional settings. Results can, in principle, not be generalized to other grade levels, other mathematics domains, other instructional settings, other games, or other countries. Another issue regarding generalizability is the fact that our analysis sample was not fully representative of the Dutch population of primary schools and students. The selective dropout of schools and students (as mentioned in section 2.2) caused the analysis sample to contain students with more favorable characteristics (higher average level of parental education and higher average mathematics ability) than was the case for the representative sample that was initially recruited. This means that our results can, essentially, only be generalized to (schools with) students with similarly favorable characteristics. Furthermore, the fact itself that several schools dropped out, or had to be excluded from the analysis because they did not meet the intervention fidelity criterion, indicates that our findings are generalizable only to schools or classes of which teachers are willing and able to actually execute the intervention.

4.7 Limitations of the study

In addition to the above generalizability issues, some further limitations of our study should be noted. Most of these limitations are a natural consequence of the fact that we performed a large-scale experiment in the real school practice. First of all, as is common in such an experiment, in our study the interventions were conducted by the regular class teachers. The teachers might have interpreted our instructions in their own way; as is generally the case when teachers use instructional materials. Although the teacher logbooks and gameplay log data informed us on how many of the games were treated by the teachers, and despite the fact that we took several measures to prevent the intervention from being implemented other than intended (e.g. providing precise guidelines and organizing information meetings), we cannot be sure about the actual in-class activities that have contributed to the effectiveness of the games. In fact, the micro-level of instruction needs further research.

Another issue related to doing research in the school practice is that, beyond the mini-games in our intervention, other educational software or games could have been used. As we wanted to examine the effects of our mini-games in an educational situation as realistic as possible, we did not forbid teachers or students to work with other educational software (as explained in section 2.3.5). This means that our results should be interpreted as the effects of implementing the mini-games as part of the regular curriculum for multiplication and division, as compared to such curriculum without these mini-games but with possibly some other educational software related to multiplicative reasoning.

A further point is that the intervention that was found to be most effective – playing the games at home with debriefing at school (E3) – seemed to be hard to maintain for some of the participating teachers (several of the E3 classes did not meet the intervention fidelity criterion). Possibly, this had something to do with decreasing enthusiasm of the students for playing the games at home, for example, due to a decreasing novelty effect. Teachers may have skipped debriefing sessions when they noticed that only a few students had played the games. This means that, possibly, the effect we found of this intervention primarily counts for classes in which students are sufficiently motivated to keep playing the games at home, or, alternatively, for classes in which teachers are willing to hold debriefing sessions regardless of whether students have played the games. The specific requirements for successfully implementing an intervention including playing at home with afterwards debriefing at school should be further investigated.

4.8 Conclusions

Our findings give evidence for the possibility of increasing primary school students' multiplicative reasoning ability through an intervention with multiplicative mini-games. The mini-games were found to be most effective when played at home and afterwards debriefed at school. When utilized in this way, mini-games were found to promote students' multiplicative operation skills (procedural knowledge) as well as their insight in multiplicative number relations (conceptual knowledge), and both an intervention in Grade 2 and a combined Grade 2-3 intervention were effective. The mini-games were also found effective when played at school, but only for enhancing multiplicative insight and only in Grade 2. When the mini-games were played at home without attention at school, they were only effective for students with above-average prior mathematics ability, for enhancing insight (Grade 2) and multiplicative fact knowledge (Grade 3). Our findings further show that more gameplay was in some cases related to more learning, but this relation was not always present, indicating that there was not always a one-to-one relation between learning time and learning outcomes. Regarding gender, we found that in Grade 2 the games were more effective for boys than for girls, whereas this difference disappeared (and occasionally reversed) in Grade 3.

In the course of our research project, it appeared that a large-scale study situated in school practice is hard to carry out. Because of teachers' busy schedules it was hard to find them willing to participate in a long-term study, and to motivate teachers in subsequent grades to continue the study. However, we think that conducting this research in real school settings to collect evidence for the effectiveness of mathematics games in primary education was worth the effort. It provided us with knowledge of when and for what students mathematics mini-games are useful. Moreover, as the interventions were delivered by the teachers themselves, our results are directly applicable to the school practice.

References

Anghileri, J. (2006). *Teaching number sense* (2nd ed.). London: Continuum.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Ault, M., Adams, D., Rowland, A., & Tiemann, G. (2010, June). *Targeted educational games: Fun and so much more!* Paper presented at the annual conference of the International Society for Technology in Education, Denver, Colorado.

Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology, 43,* 993-1003.

Bakker, M., Van den Heuvel-Panhuizen, M., Van Borkulo, S., & Robitzsch, A. (2013). Effecten van online mini-games op multiplicatieve vaardigheden van leerlingen in groep 4 [Effects of online mini-games on second-graders' multiplicative reasoning ability]. *Pedagogische Studiën, 90*(3), 21-36.

Bourgonjon, J., Valcke, M., Soetaert, R., & Schellens, T. (2010). Students' perceptions about the use of video games in the classroom. *Computers & Education, 54,* 1145-1156.

Chang, K.-E., Sung, Y.-T. Chen, Y.-L., & Huang, L.-H. (2008). Learning multiplication through computer-assisted learning activities. *Computers in Human Behavior, 24,* 2904-2916.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88,* 715-730.

De Jean, J., Upitis, R., Koch, C., & Young, J. (1999). The story of Phoenix Quest: How girls respond to a prototype language and mathematics computer game. *Gender and Education, 11,* 207-223.

Department for Education UK. (2011). *Primary national curriculum until 2014. Mathematics: Key stage 1*. Retrieved May 27, 2013, from http://www.education.gov.uk/schools/teachingandlearning/curriculum/primary/b00199044/mathematics/ks1.

De Vos, T. (2010). *TempoTest Automatiseren* [Tempotest automatization]. Amsterdam: Boom test uitgevers.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment. Exploring the educational potential of computer games* (Doctoral dissertation). IT-University of Copenhagen, Denmark.

Frazer, A., Argles, D., & Wills, G. (2007, September). *Assessing the usefulness of mini-games as educational resources.* Paper presented at the annual conference of the Association for Learning Technology, Nottingham, UK.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33,* 441-467.

Gee, J. P. (2005). Good video games and good learning. *Phi Kappa Phi Forum, 85*(2), 33-37.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549-576.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences, 20,* 169-206.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore, MD: Johns Hopkins University Press.

Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24,* 190-197.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8* [Scientific justification of the LOVS mathematics tests for Grade 1 to 6]. Arnhem, Netherlands: Cito.

Jonker, V., Wijers, M., & Van Galen, F. (2009, October). *The motivational power of mini-games for the learning of mathematics.* Paper presented at the European Conference on Game Based Learning, Graz, Austria.

Kamil, M. L., & Taitague, C. (2011). Developing an electronic game for vocabulary learning: A case study. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 331-351). Charlotte, NC: Information Age.

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 1-32). Hershey, PA: IGI Global.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55,* 427-443.

Kirriemuir, J. (2002). *The relevance of video games and gaming consoles to the higher and further education learning experience.* Techwatch Report. Retrieved December 10, 2012, from http://tecnologiaedu.us.es/nweb/htm/pdf/301.pdf.

Klawe, M. M. (1998, June). *When does the use of computer games and other interactive multimedia software help students learn mathematics?* Paper presented at the Technology and NCTM Standards 2000 Conference, Arlington, VA.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.

Kolovou, A., Van den Heuvel-Panhuizen, M., & Köller, O. (2013). An intervention including an online game to improve Grade 6 students' performance in early algebra. *Journal for Research in Mathematics Education, 44,* 510-549.

Leemkuil, H., & De Jong, T. (2004). Games en gaming [Games and gaming]. In P. Kirschner (Ed.), *ICT in het onderwijs: The next generation. Katern bij onderwijskundig lexicon, uitbreiding editie III* (pp. 41-63). Alphen aan de Rijn, Netherlands: Kluwer.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22,* 215-243.

Lowrie, T., & Jorgensen, R. (2011). Gender differences in students' mathematics game playing. *Computers & Education, 57*(4), 2244-2248.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4,* 333-369.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: Vol. 3 Conative and affective process analyses* (pp. 223-253). Hillsdale, NJ: Lawrence Erlbaum.

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research & Practice, 22,* 47-57.

Miller, D. J., & Robertson, D. P. (2011). Educational benefits of using game consoles in a primary classroom: A randomised controlled trial. *British Journal of Educational Technology, 42*, 850-864.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles: Author.

NCTM (National Council of Teachers of Mathematics). (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence.* Reston, VA: Author.

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology, 82,* 136-156.

Panagiotakopoulos, C. T. (2011). Applying a conceptual mini game for supporting simple mathematical calculation skills: Students' perceptions and considerations. *World Journal of Education, 1*(1), 3-14.

Panagiotakopoulos, C. T., Sarris, M. E., & Koleza, E. G. (2013). Playing with numbers: Development issues and evaluation results of a computer game for primary school students. In T. Sobh & K. Elleithy (Eds.), *Emerging trends in computing, informatics, systems sciences, and engineering. Lecture notes in electrical engineering. Volume 151* (pp. 263-275).

Parsons, H. M. (1974). What happened at Hawthorne? *Science, 183,* 922-932.

Prensky, M. (2001). *Digital Game-Based Learning.* New York: McGraw-Hill.

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M^2: Media in the lives of 8-to 18-year-olds.* Menlo Park, CA: Henry J. Kaiser Family Foundation.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., ... & Salinas, M. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education, 40,* 71-94.

Sandberg, J., Maris, M., & De Geus, K. (2011). Mobile English learning: An evidence-based study with fifth graders. *Computers & Education, 47,* 1334-1347.

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

Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2011). Review of research on computer games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 127-222). Charlotte, NC: Information Age.

Trautwein, U., Ludtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology, 98,* 438-456.

Van Borkulo, S., Van den Heuvel-Panhuizen, M., Bakker, M., & Loomans, H. (2012). One mini-game is not like the other: Different opportunities to learn multiplication tables. In S. De Wannemacker, S. Vandercruysse & G. Clarebout (Vol. Eds.), *Communications in computer and information science: Vol. 280. Serious games: The challenge* (pp. 61-64). Berlin: Springer.

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1-67.

Van den Heuvel-Panhuizen, M. (Ed.). (2008). *Children learn mathematics: A learning-teaching trajectory with intermediate attainment targets for calculation with whole numbers in primary school.* Rotterdam, Netherlands: Sense.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34,* 229-243.

Volman, M., Van Eck, E., Heemskerk, I., & Kuiper, E. (2005). New technologies, new differences. Gender and ethnic differences in pupils' use of ICT in primary and secondary education. *Computers & Education, 45,* 35-55.

Williamson, B. (2009). *Computer games, schools, and young people: A report for educators on using games for learning.* Bristol, UK: Futurelab.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105,* 249-265.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). ACER ConQuest (Version 2.0) [Computer Software]. Camberwell, Australia: Australian Council for Educational Research.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., ... & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61-89.

Appendix A

Correlations between variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. SkillsTest1	–																			
2. SkillsTest2	**.55**	–																		
3. SkillsTest3	**.50**	**.60**	–																	
4. InsightTest2	**.52**	**.73**	**.62**	–																
5. InsightTest3	**.55**	**.61**	**.77**	**.61**	–															
6. KnowledgeTest2	**.12**	**.38**	**.38**	**.39**	**.36**	–														
7. KnowledgeTest3	**.11**	**.30**	**.39**	**.30**	**.40**	**.66**	–													
8. GMath	**.63**	**.55**	**.57**	**.55**	**.60**	**.27**	**.28**	–												
9. E1	.03	.03	.02	.08	.05	.02	-.07	.00	–											
10. E2	-.06	-.09	-.07	-.04	**-.10**	-.11	-.05	-.09	**-.27**	–										
11. E3	-.05	.01	.03	.02	.02	.03	.05	-.04	**-.15**	**-.22**	–									
12. Gplay2*E1	.04	.04	.02	**.10**	.05	.04	-.06	.01	**.97**	**-.26**	**-.15**	–								
13. Gplay2*E2	.08	**.13**	**.10**	.09	**.12**	.09	.08	**.12**	**.16**	**-.59**	**.13**	**.15**	–							
14. Gplay2*E3	.05	**.08**	.05	.02	**.05**	.02	.01	.04	.02	.03	-.13	.02	-.02	–						
15. Gplay3*E1	.03	.04	.03	**.11**	.05	.06	-.05	.02	**.96**	**-.26**	**-.14**	**.97**	**.15**	.02	–					
16. Gplay3*E2	.06	.09	.06	.03	**.09**	.10	.05	.08	**.25**	**-.91**	**.20**	**.24**	**.67**	-.03	**.24**	–				
17. Gplay3*E3	.04	.05	.01	.01	.03	.01	.01	.02	**.09**	**.13**	**-.58**	**.09**	**-.08**	**.50**	**.08**	**-.12**	–			
18. Female	.00	-.07	**-.11**	**-.16**	**-.08**	-.06	-.01	**-.11**	**-.05**	.04	-.08	-.05	.06	.11	**-.06**	.02	**.11**	–		
29. AgeDelayed	-.06	**-.17**	**-.16**	**-.17**	**-.19**	**-.15**	**-.14**	**-.11**	.02	.04	.11	.04	-.05	-.02	.03	-.03	**-.09**	-.02	–	
20. ParEdLow	**-.15**	**-.20**	**-.17**	**-.19**	**-.15**	-.09	-.06	**-.14**	.06	.07	-.04	.06	-.08	.02	.05	-.08	.01	-.01	.08	–
21. NonDutch	-.09	-.08	-.09	-.10	-.01	.01	.04	-.10	.11	-.03	.03	.13	-.02	-.02	.13	.03	-.04	-.02	.10	**.23**

Note. $N = 719$. E1, E2 and E3 are the condition dummy variables, Gplay2*E1 to Gplay3*E3 are interactions between condition dummy variables and gameplay variables. Bolded correlations are significant at the $\alpha = .05$ level (two-tailed).

Appendix B

Paired comparisons between the E1, E2, and E3 condition of direct, indirect, and total effects on the three aspects of multiplicative ability (knowledge, skills, and insight)

	Aspect of multiplicative reasoning ability[a]					
	Knowledge		Skills		Insight	
Effect	$\Delta\beta_{ps}$	*SE*	$\Delta\beta_{ps}$	*SE*	$\Delta\beta_{ps}$	*SE*
Comparison E1 – E2						
Direct effect on Grade 2 score	0.18	0.21	0.10	0.15	0.14	0.15
Direct effect on Grade 3 score	-0.25	0.17	-0.03	0.11	0.06	0.09
Indirect effect	0.11	0.13	0.08	0.11	0.08	0.08
Total effect	-0.15	0.23	0.05	0.13	0.13	0.15
Comparison E3 – E2						
Direct effect on Grade 2 score	0.24	0.24	0.18†	0.11	0.09	0.12
Direct effect on Grade 3 score	0.06	0.13	0.06	0.11	0.16*	0.09
Indirect effect	0.15	0.15	0.14†	0.09	0.05	0.07
Total effect	0.21†	0.14	0.19*	0.11	0.21*	0.09
Comparison E3 – E1						
Direct effect on Grade 2 score	0.06	0.24	0.08	0.15	-0.05	0.18
Direct effect on Grade 3 score	0.31*	0.18	0.08	0.10	0.10	0.09
Indirect effect	0.04	0.15	0.06	0.11	-0.03	0.10
Total effect	0.35*	0.21	0.14	0.13	0.08	0.14

Note. E1: $n = 112$; E2: $n = 202$; E3: $n = 78$. SkillsTest1, Female, AgeDelayed, ParEdLow, NonDutch, and GMath were included as covariates (see Figure 3). $\Delta\beta_{ps}$ = difference in partially standardized coefficients.
[a]The model was separately specified for each of the three aspects of multiplicative reasoning ability.
† $p < .10$. * $p < .05$. Two-tailed for the E1-E2 comparison, one-tailed for the E3-E2 and the E3-E1 comparison (because of our directional hypothesis).

Appendix C

Interactions of gender and prior mathematics ability with condition variables predicting direct, indirect, and total effects on the three aspects of multiplicative reasoning ability (knowledge, skills, and insight)[a]

	Interaction with gender (Female)						Interaction with prior mathematics ability (GMath)					
	Knowledge		Skills		Insight		Knowledge		Skills		Insight	
Effect	β_{ps}	*SE*	β_{ps}	*SE*	β_{ps}	*SE*	β	*SE*	β	*SE*	β	*SE*
Condition E1												
Direct effect on Grade 2 score	-0.21	0.17	-0.10	0.18	-0.57*	0.22	-.07	.17	-.11	.12	.08	.15
Direct effect on Grade 3 score	-0.13	0.19	0.05	0.17	0.39*	0.19	-.02	.09	.05	.14	-.13	.13
Indirect effect	-0.14	0.11	-0.07	0.12	-0.28*	0.13	-.04	.09	-.06	.07	.03	.06
Total effect	-0.27	0.23	-0.02	0.14	0.06	0.16	-.06	.13	-.02	.12	-.07	.10
Condition E2												
Direct effect on Grade 2 score	-0.20	0.14	-0.06	0.18	-0.37*	0.18	-.14	.11	.28†	.15	.28*	.14
Direct effect on Grade 3 score	-0.05	0.13	0.03	0.20	0.27	0.19	.18†	.10	.14	.16	.09	.15
Indirect effect	-0.12	0.09	-0.04	0.12	-0.18*	0.09	-.06	.05	.13†	.07	.10*	.05
Total effect	-0.17	0.17	-0.01	0.14	0.06	0.16	.07	.09	.22*	.11	.16†	.09
Condition E3												
Direct effect on Grade 2 score	-0.47**	0.15	0.07	0.33	0.01	0.25	-.17	.12	-.10	.16	.09	.09
Direct effect on Grade 3 score	0.00	0.20	-0.01	0.33	0.03	0.18	.03	.14	.05	.10	-.14	.13
Indirect effect	-0.30**	0.10	0.05	0.22	0.00	0.12	-.11	.08	-.07	.11	.05	.05
Total effect	-0.30†	0.17	0.04	0.18	0.03	0.14	-.08	.21	-.02	.16	-.09	.12

Note. $N = 719$. SkillsTest1, Female, AgeDelayed, ParEdLow, NonDutch, and GMath were included as covariates. β_{ps} = partially standardized coefficient (a positive value signifies a female advantage).
[a]The model was separately specified for each of the three aspects of multiplicative reasoning ability.
† $p < .10$. * $p < .05$. ** $p < .05$. Two-tailed.

Chapter 5

Effects of mathematics computer games on special education students' multiplicative reasoning ability

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (submitted). Effects of mathematics computer games on special education students' multiplicative reasoning ability.

Effects of mathematics computer games on special education students' multiplicative reasoning ability

1 Introduction

Students in special education are often considerably behind in their mathematics ability, as compared to their same-aged peers in general education (e.g., Cawley, Parmar, Foley, Salmon, & Roy, 2001). In the Netherlands, for example, it has been found that 12-year-old students in special primary education perform, on average, at a level of mathematics achievement similar to that of 9-year-old students in regular primary education (Kraemer, Van der Schoot, & Van Rijn, 2009). Therefore, much effort is put in developing effective instructional methods to achieve better learning outcomes in special education. One method that has been found promising is computer-assisted instruction (CAI; e.g., Bouck & Flanagan, 2009). Some advantages of CAI are that it can offer students immediate feedback (e.g., Seo & Woo, 2010; Woodward & Rieth, 1997), it often has positive motivational effects (e.g., Okolo, 1992), and it can accommodate for individual differences (e.g., Woodward & Rieth, 1997). Indeed, meta-analytic studies have indicated the effectiveness of CAI in enhancing mathematics learning outcomes in students with special educational needs (Li & Ma, 2010; Xin & Jitendra, 1999). However, in meta-analyses by Kroesbergen and Van Luit (2003) and Seo and Bryant (2009) results were less conclusive and sometimes favored other instructional methods. One of the possible explanations given by Seo and Bryant is that in many of the reviewed CAI studies the intervention had a rather short duration (1-2 weeks), which might have been too short for special education students to improve. Furthermore, Kroesbergen and Van Luit stressed that CAI cannot replace the teacher, corresponding to Woodward and Rieth's (1997) conclusion that often CAI alone is not sufficient to establish gains in performance.

One way of employing CAI in mathematics education is through the use of mathematics computer games. Games are considered particularly motivating for children (e.g., Garris, Ahlers, & Driskell, 2002; Malone, 1981). As far as we know, the existing knowledge base on effects of mathematics computer games in special education is rather limited. Yet, a few studies did find evidence for mathematics computer games to enhance special needs students' mathematics performance (e.g., Brown, Ley, Evett, & Standen, 2011; Okolo, 1992). In the present research we examined the effects of a long-term teacher-delivered intervention with computer games in the field of multiplicative reasoning (multiplication and division), which is a mathematics domain in which students in special primary education are often particularly delayed (Kraemer et al., 2009).

2 Background

2.1 Computer games for special education students

To have potential for special education students, instructional computer games need to meet certain requirements. For example, the games should be simple, meaning that they should include few distracting features (Christensen & Gerber, 1990; Ke & Abras, 2013; Seo & Woo, 2010), be easy to learn, and require few reading (e.g., Ke & Abras, 2013). Furthermore, to fully engage students in the learning content of a game, the learning content should be integrated in the main gameplay activity (e.g., Habgood & Ainsworth, 2011; Ke & Abras, 2013). A type of game that often meets these requirements is the so-called mini-game, which is a short, focused game that is easy to learn (e.g., Jonker, Wijers, & Van Galen, 2009). Additionally, it appears important for special education students that games provide continuous rewards (Ke & Abras, 2013) and that they allow for different difficulty levels, for example, by providing choices in the types of problems to be solved or making available supportive features (e.g., Brown et al., 2011; Ke & Abras, 2013).

2.2 Computer games for enhancing multiplicative reasoning

The mathematics domain of multiplicative reasoning, like other mathematics domains, entails different types of knowledge (e.g., Goldman & Hasselbring, 1997; Miller & Hudson, 2007): declarative knowledge (knowledge of multiplicative number facts), procedural knowledge (knowledge of how to calculate multiplicative problems), and conceptual knowledge (conceptual understanding of the multiplication and division operation and relations between multiplicative problems). Mathematics games and other CAI programs used in special education are often aimed at the development of declarative and/or procedural knowledge (e.g., Okolo, 1992; Seo & Bryant, 2009). However, computer games can also address conceptual knowledge (e.g., Jonker et al., 2009; Klawe, 1998). Through offering opportunities for exploration and experimentation, enabling experiential learning (e.g., Garris et al., 2002; Kebritchi, Hirumi, & Bai, 2010), games can help students develop conceptual understanding of, for example, relations between mathematics problems. For the domain of multiplicative reasoning this entails, among others, the principles of commutativity (e.g., $3 \times 7 = 7 \times 3$) and distributivity (e.g., $7 \times 8 = 5 \times 8 + 2 \times 8$), and derived fact strategies such as one more/one less and doubling and halving. Furthermore, the use in games of visual mathematical representations can contribute to the development of conceptual knowledge (e.g., Seo & Woo, 2010); for multiplicative reasoning a useful representation is, for example, the rectangular array (e.g., Barmby, Harries, Higgins, & Suggate, 2009). For the learning from games based on exploration and experimentation, reflection is crucial (e.g., Garris et al., 2002), as it leads students to generalize what they have learned, such that they can also apply it outside the game (transfer). Because this reflection often does not spontaneously occur in students (Garris et al., 2002), especially not in special education students (e.g., Ke & Abras, 2013), it should be elicited. Reflection may, for example, be encouraged when, after playing, a game is discussed in class or in small groups (e.g., Klawe, 1998), often called debriefing (e.g., Garris et al., 2002). Similarly, support before and during the game may foster learning (e.g., Ke & Abras, 2013).

2.3 Our study

In the current study we investigated the effects of an intervention with multiplicative mini-games on special education students' multiplicative reasoning ability. The mini-games used in the study focused on developing declarative, procedural, as well as conceptual knowledge of multiplicative reasoning. The games were accompanied by lessons and class discussions. We aimed to examine the effects of the intervention as implemented in a real special education setting, with students' own teachers delivering the intervention.

Our research question was:

Does an intervention with multiplicative mini-games affect special education students' learning outcomes in multiplicative reasoning?

We hypothesized that multiplicative mini-games, in comparison to the regular mathematics curriculum without these mini-games, would positively affect the learning of multiplicative reasoning, because they provide an engaging environment in which students are motivated to practice basic multiplicative number facts and operation skills (declarative and procedural knowledge) and can develop conceptual knowledge of multiplicative reasoning through exploration and experimentation.

3 Method

3.1 Research design

To answer our research question, we employed a pretest-posttest control-group design. In the experimental group (E), multiplicative mini-games were played,[1] while in the control group (C) there was a pseudo-intervention with mini-games on other mathematics domains (spatial orientation, addition and subtraction). The pseudo-intervention was meant to control for the positive effect that participating in an experiment may have by itself (Hawthorne effect, see Parsons, 1974). In both conditions, teachers were asked to keep the total lesson time spent on each of the mathematics domains included in the curriculum the same as would have been the case had the school not been participating in the study. In this

[1] The E condition in this chapter is equivalent to the E1 condition in Chapter 3 and 4.

way, we could compare the regular curriculum for multiplicative reasoning (in C) with a multiplicative reasoning curriculum including mini-games (in E).

As is shown in Figure 1, the study lasted one year. In this year, there were two game periods in which mini-games were played. Students' progress in multiplicative reasoning was measured using a pretest (Mult1) and a posttest (Mult2) of multiplicative reasoning ability, including items focused on procedural and conceptual knowledge. At posttest, also a test of automaticity of multiplication facts was administered (MAut2), measuring students' declarative knowledge. Furthermore, we collected background data on students' gender, age and home language and their general mathematics ability.

	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
School year 2009/2010										Mult1
School year 2010/2011	Game period 1					Game period 2				Mult2 MAut2

Figure 1. Time schedule of the study. Note: Mult = Multiplicative ability test (procedural and conceptual knowledge); MAut = Automaticity test of multiplication facts (declarative knowledge).

3.2 Participants

Our study was carried out in schools for special primary education in the Netherlands. These schools are meant for students with learning difficulties, students with mild mental retardation, and students with mild to moderate behavioral or developmental problems. Children with severe mental retardation, severe behavioral or psychiatric problems, or physical disabilities are not included in these schools. In special primary education schools the mathematics curriculum is in principle the same as in general primary education, but children are given more support and can progress at their own pace.

Schools were recruited by contacting them by phone (response rate ca. 24%) or e-mail (response rate ca. 3%). We found 11 schools to be willing to participate. These schools were blocked on school characteristics and then randomly divided over the two conditions (our initial research design included two additional experimental conditions – playing at home and playing at home with debriefing at school – but because the schools in these extra conditions did not manage to conduct the research project as intended, these conditions could not be included in our analysis[2]). Unfortunately, for various reasons such as organizational problems and problems with computers, only three experimental schools and two control schools managed to complete the intervention and administer the posttest.

As we aimed to investigate the effects of mini-games in the first year of learning multiplicative reasoning, we asked the schools to include in the study those students that were expected to have, at the beginning of the 2010/2011 school year, a mathematics ability level corresponding to the beginning of Grade 2, where, in the Netherlands, formal instruction of multiplicative reasoning commonly commences (see Van den Heuvel-Panhuizen, 2008). After excluding students that left the school in the course of the research project (three students), did not complete any of the Mult tests (three students), or received less than half of the E intervention (four students), our sample consisted of 97 students. Furthermore, from the C condition, 16 students had to be excluded because their teacher did not administer them the MAut test for the (assumed) reason that they were expected not yet to be able to take this test, which suggests that these students might not have received a multiplicative reasoning curriculum comparable to that of the E condition. This led to a final analysis sample of 81 students, of which 40 were in the E condition and 41 were in the C condition. Because in the Netherlands, special primary education students are often grouped in classes on the basis of their reading level rather than their mathematics level, the participating students were in many different mathematics classes. A total of 17 classes was involved (E: 5; C: 12), with 1 to 17 participating students in a class.

3.3 Intervention

The intervention consisted of two game periods, each lasting 10 weeks (see Figure 1). In each game period there were eight different mini-games; every week a new game, except for the fifth and tenth week, in which earlier games were repeated.

The mini-games used in the experimental condition were adapted versions of multiplicative mini-games from the Dutch mathematics games website Rekenweb (www.rekenweb.nl, English version: www.thinklets.nl; see Jonker et al., 2009). The adaptations involved both the difficulty level of the underlying multiplicative problems and the learning opportunities of the games. In the adapted games there were, for example, more and clearer connections between multiplicative problems and representations (e.g., formal notation and rectangular array representation). Also, we added to the games a scoring mechanism, with a score that increased as children more often successfully finished the game. In the control condition

[2] We started the project with 19 schools of special primary education (seven in the control group, and four in each of the three experimental groups). The schools in the home-playing conditions could not be included in our analysis because in all but one of these schools less than half of the intervention was carried out. In one of the schools in the condition in which the games were played at home and debriefed at school the intervention was sufficiently carried out, but this school could not be included in our analysis because of invalid Mult2 scores, due to technical problems in the test administration.

existing mini-games from Rekenweb, about spatial orientation, addition, and subtraction, were used. In both conditions the mini-games were made available online, through the Digital Mathematics Environment (DME; see http://www.fi.uu.nl/wisweb/en/).

The mini-games used in the experimental condition focused on practicing multiplicative number facts and operation skills (declarative and procedural knowledge), as well as on developing conceptual understanding of multiplicative number relations and properties of multiplicative operations. The games covered the principles of commutativity, distributivity, and associativity, and derived fact strategies such as one more and one less, and doubling and halving. In accordance with the needs of special education students, the games were easy to learn, contained simple graphics and minimal sound effects, provided rewards (e.g., in the form of an increasing score), and the mathematics content was integrated into the main activity of the games. Furthermore, many of the games allowed for different difficulty levels.

Figure 2 shows two sample games from the intervention in the experimental group.[3] The game "Making groups" (Figure 2a) involved a rectangular array representation in which the student had to make rectangular groups of smileys and then determine the number of smileys in the group. In this way, the student practiced calculating multiplication problems – by using memorized multiplication facts or, for example, by repeated addition – and could gain conceptual understanding of the relations between multiplication problems; for example, 3 rows of 5 is the same as 5 rows of 3 (commutative property), and if 5 rows of 3 is 15, then 6 rows is 3 more, resulting in 18 (derived fact strategy of one more, or distributive property). The difficulty level could be determined by the student by choosing which arrays to make. In the game "Frog" (Figure 2b), the student had to come up with their own multiplication problem, after which the answer to a related multiplication problem was asked. Again, the student practiced the calculation of multiplication problems and could gain insight in the relations between multiplication problems.

Before each game period, the participating teachers were given a manual in which, for each week, it was described which game was offered that week, and how it had to be treated in class. In the E condition the teachers were asked to introduce each new game in a whole-class lesson (20 minutes), using a worksheet. Then the students had to watch a short instruction video introducing the game, and play the game for approximately 10 minutes. Afterwards, the game was debriefed in a teacher-led discussion (15 minutes), using a digital blackboard or a class computer. Guided by questions posed by the teacher, which were given in the teacher manual, the students discussed which strategies were faster or more useful in the game. Finally, the students played the game for another 10 minutes, during which they could try the discussed strategies. In the C condition, the teacher was asked to

[3] The Appendix of this thesis contains a desciption of all games in the experimental group intervention (Game period 1 and Game period 2).

introduce each new C game in a whole-class lesson (10 minutes), after which the students played the game in one or two sessions of 10 minutes.

The teachers were asked to keep a logbook, in which they could note each week whether the different parts of the intervention were executed. From these logbooks it appeared that in all three E schools, all 16 games were played.

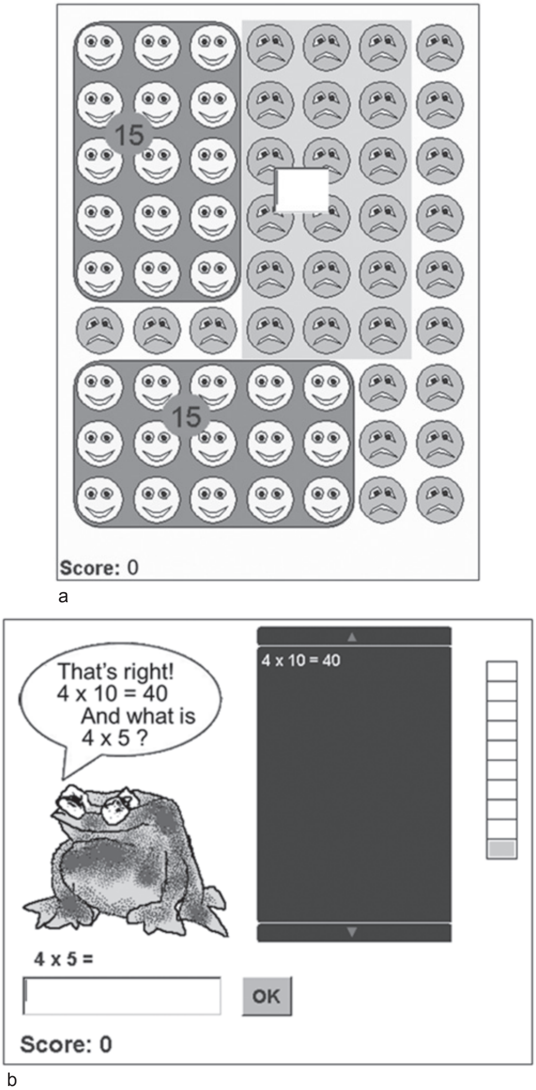


Figure 2. Sample games from the intervention in the experimental group. a. “Making groups”. b. “Frog”.

3.4 Measurement instruments

3.4.1 Multiplicative ability tests

The multiplicative ability tests Mult1 and Mult2 were specially constructed for our study, and were meant to measure students' procedural and conceptual knowledge of multiplicative reasoning. The tests contained bare number problems (see Figure 3a) and context problems (see Figure 3b). In addition Mult2 included "insight problems" (see Figure 3c), in which students had to use their knowledge of multiplication and division at a higher comprehension level. Besides the multiplicative items, both tests also included some "distractor" items (not used in our analyses) on spatial orientation, addition, and subtraction, which were meant to conceal from the students and teachers in the control group that the focus of the study was on multiplicative reasoning. Mult1 contained 22 multiplicative items and 8 distractor items; Mult2 contained 26 multiplicative items – of which 10 were also in Mult1 (anchor items) – and 7 distractor items.[4]

The Mult tests were administered online, using the DME. The digital administration allowed for a relatively standardized test setting in which students could yet work at their own pace. The question accompanying each test item was read aloud by the computer. To control for item-order effects, both tests were administered in four, differently ordered, versions, which were randomly assigned to the students. The duration of each Mult test was, on average, approximately 15 minutes.

On average, the students correctly answered 47.4% of the items in Mult1 and 51.3% of the items in Mult2, indicating that the tests were neither too easy nor too difficult. The Mult items were scaled using a Rasch model in the Conquest software (Wu, Adams, Wilson, & Haldane, 2007), resulting in scale scores (weighted likelihood estimates, or WLE) for both Mult tests separately. To subsequently put the two tests on a common scale, we employed mean-mean linking (Kolen & Brennan, 2004), assuming equal item difficulties, on average, of the anchor items in the two tests (for equal student ability). This method resulted in scale scores, or WLE scores (Wu et al., 2007), for the tests. The WLE reliability of these scores, which can be interpreted in the same way as a Cronbach's alpha, was .80 for Mult1 and .79 for Mult2. This means that the tests can be considered sufficiently reliable.

[4] Mult1 and Mult2 contained a subset of the items of Test 1 and Test 2 administered to the regular education students in our project (Chapter 3 and 4). The tests were made shorter by omitting some of the more difficult items, in order to match the ability level of the special education students (in line with pilot test results). The number of insight items in Mult2 (8 items) was not enough to construct a reliable insight scale; therefore, the test score was treated as a one-dimensional score (as was also done in Chapter 3).

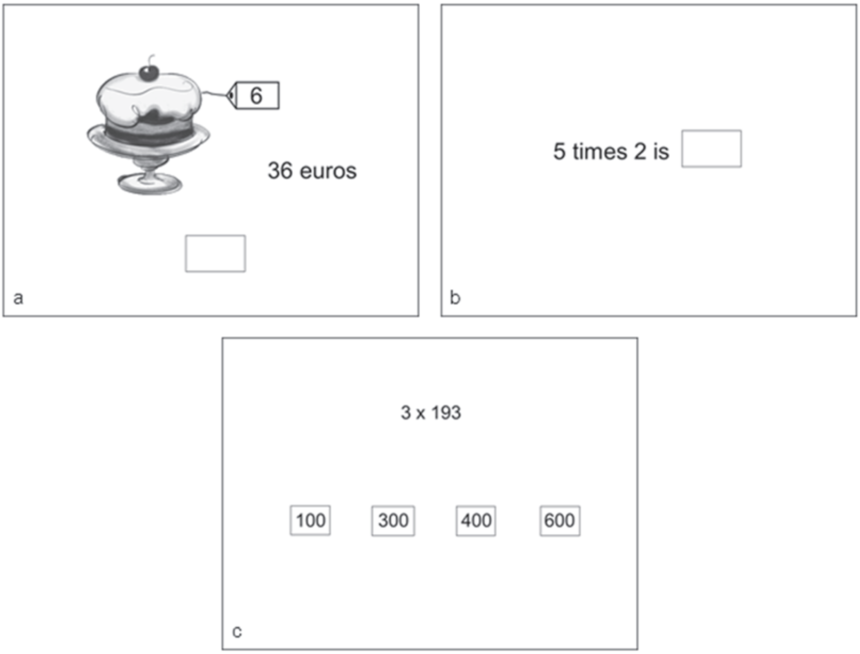


Figure 3. Sample Mult items. a. “Five times two is…” (Mult1 and Mult2) b. “How many pies can you buy for 36 euros?” (Mult2) c. “Three times 193 is closest to… Click the correct number” (Mult2)

3.4.2 Automaticity test

In addition to the Mult tests, we administered a standardized test of declarative knowledge (automaticity) of multiplication facts (MAut): the multiplication subtest of the TempoTest Automatiseren (De Vos, 2010).[5] To conceal from the teachers and students in the control group the focus on multiplicative reasoning, also the addition and subtraction subtests were administered, but these were not included in our analyses. The multiplication subtest consists of a sheet of 50 bare number × problems. Students get 2 minutes time to solve as many of these problems as possible; the test score is the number of correct answers. The split-half reliability of the test is .96 (De Vos, 2010). The MAut test was only administered as a posttest (MAut2), because at pretest the students were not yet familiar with the × symbol.

[5] This is the same test as the Knowledge Test in Chapter 4.

3.4.3 General mathematics ability test

As a background variable, students' initial general mathematics ability was measured using a standardized test from the Cito student monitoring system (see Janssen, Verhelst, Engelen, & Scheltens, 2010). This test, which we refer to as GMath, was administered as part of the schools' regular testing program. The reliability coefficients of the different versions of this test range from .91 to .97 (Janssen et al., 2010). The students in our sample scored on average 28.1 (SD = 10.5) on the GMath test, which is considerably less than the average of 34.8 (SD = 14.6) for general primary education students just before the start of Grade 2 (Janssen et al., 2010).

3.5 Treatment of missing data

As is inevitable in a long-term study carried out in school practice, some students missed one of the tests (Mult1: 10 students; Mult2: 15 students; MAut2: 14 students; GMath: 4 students). We employed multiple data imputation to make estimates for the missing test scores (see Graham, 2009). Our imputation model included student background data, test scores, a condition dummy variable, and school and class mean test scores to account for the clustered data structure. The data imputation, which was performed using the mice software (Van Buuren & Groothuis-Oudshoorn, 2011), resulted in 50 imputed datasets. Statistical analyses were executed on these 50 datasets. The results were combined by using Rubin's rule (see Graham, 2009).

3.6 Data analysis

Our analyses were performed using Mplus (Muthén & Muthén, 1998-2010). To control for unreliability in the Mult1 and Mult2 scores, we modeled these scores as latent variables, with their residual variance fixed at *[1 – reliability of test scores]*variance of test scores* (see Hayduk, 1987). In our analyses we did not use a correction for the clustered data structure, as such corrections are unreliable when only few clusters are involved (see Angrist & Pischke, 2009).

4 Results

4.1 Initial differences between groups

Before data analysis, we checked for differences between the experimental group and the control group with respect to their student composition. We looked at students' gender, age, and home language (monolingual Dutch vs. other), their initial general mathematics ability (GMath score), and their initial multiplicative reasoning ability (Mult1 scale score). As is shown in Table 1, we did not find significant differences between the groups for any of these variables ($p > .05$). However, for age and GMath score, d was larger than 0.2, which is a non-negligible effect size. Therefore, to be conservative, we decided in our analyses to control for these variables.

Table 1
Initial differences between conditions

Condition	n	Gender % female	Age M (SD)	Home language % Dutch	GMath score M (SD)	Mult1 score M (SD)
E	40	30.0	9.32 (1.33)	82.4	26.1 (8.8)	-0.02 (1.14)
C	41	31.7	9.06 (1.04)	87.8	30.1 (11.6)	0.00 (1.24)
Total	81	30.9	9.19 (1.18)	85.2	28.1 (10.5)	-0.01 (1.19)
E vs. C: d[a]		*-0.04*	*0.21*	*-0.15*	*-0.38*	*-0.02*

[a]No between-condition differences were significant.

4.2 Effect of the intervention

Table 2 presents descriptives of the scale scores of Mult1 and Mult2, and the MAut2 test scores. In both conditions the Mult2 scale scores were significantly higher than the Mult1 scale scores (E: $t = 4.66$, $p < .001$, $d = 0.97$; C: $t = 4.73$, $p < .001$, $d = 1.05$), indicating that in both conditions students' multiplicative ability improved from pretest to posttest.

To examine the effect of the intervention in the experimental group as compared to the control group, we performed linear regression analyses, with posttest scores as the dependent variable and the condition dummy variable CondE as the predictor. Additional predictors were the pretest score on Mult1 and the previously mentioned covariates Age and GMath score. The standardized regression coefficients are displayed in Table 3. Because of our directional hypothesis, the regression coefficients of CondE were tested one-tailed; the others were tested two-tailed.

From the regression results it is clear that the condition a student was in did not significantly influence their Mult2 score ($d = -0.02$, n.s.). For MAut2, however, the effect of condition was significantly positive ($d = 0.39$, $p = .047$).[6] Thus, students in the E

[6] d effect sizes were calculated by dividing the raw regression coefficients of the CondE dummy variable by the pooled standard deviation of the dependent variable.

condition had higher learning outcomes than the C students on declarative knowledge of multiplication facts.

Table 2
Descriptives of Mult scale scores and MAut scores

		Mult1		Mult2		MAut2	
Condition	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
E	40	-0.02	1.14	1.08	1.16	9.68	7.31
C	41	0.00	1.24	1.31	1.04	8.15	5.70
Total	81	-0.01	1.19	1.20	1.11	8.90	6.60

Table 3
Standardized regression coefficients of condition and covariates predicting posttest scores

	Dependent variable					
	Mult2 score			MAut2 score		
Predictor	β	*SE*	*p*	β	*SE*	*p*
CondE	-0.01	0.15	.944[a]	0.19	0.11	.047[b]
Mult1 score	0.06	0.17	.737	-0.10	0.17	.559
GMath score	0.58	0.11	<.001	0.49	0.11	<.001
Age	0.03	0.13	.791	0.12	0.11	.287

[a]Tested two-tailed because the coefficient was not in the direction of our hypothesis. [b]One-tailed.

5 Discussion

5.1 Interpretation of our findings

We found that special education students who received an intervention with multiplicative mini-games as part of their multiplicative reasoning curriculum (the experimental group) had higher learning outcomes on declarative knowledge of multiplication facts than special education students who received the regular curriculum without these mini-games (the control group). This points out the usefulness of mini-games for enhancing special needs students' mathematics fact knowledge. For procedural and conceptual knowledge, learning outcomes were the same as those obtained with the regular curriculum in the control group. For these types of multiplicative knowledge, thus, there was no added value of the mini-games. Yet, for procedural and conceptual knowledge an intervention with mini-games can still be seen as a “safe approach” to be employed as part of the multiplicative reasoning curriculum in special education, as learning outcomes were not different from those obtained in the control group.

An explanation for only finding an effect of the mini-games for declarative knowledge, as compared to the control group, may be that this knowledge was more easily transferred from the games. In fact, multiplication facts practiced in the mini-games often had the same format, with the × symbol, as the multiplication problems occurring in the MAut2 test, whereas the problems included in the Mult2 test did not have the same format as occurred in the mini-games. This transfer issue relates to earlier findings that transfer is hard for students with special educational needs (e.g., Shiah, Mastropieri, Scruggs, & Fulk, 1994). Another possible explanation is that special education teachers may not be used to a teaching method focusing on experiential or discovery-based learning (e.g., Woodward, 2004), which may have caused that in the debriefing sessions teachers were not so focused on discussing the strategies and concepts students discovered in the games. Teachers’ unfamiliarity with this kind of teaching may have led them to not optimally promote transfer in their students, especially for procedural and conceptual knowledge. Furthermore, it might be the case that declarative knowledge is more easily taught to special education students than are more complex skills (e.g., Kroesbergen & Van Luit, 2003).

5.2 Carrying out field experiments in the special education school practice

Our experience with this study was that a long-term field experiment consisting of a teacher-delivered intervention is hard to carry out in the special education school practice. It appeared to be very difficult to keep the participating teachers sufficiently involved in the research project, which caused the sample that could be included in our analysis to be much smaller than we intended. The difficulties teachers had in executing the intervention and administering the tests were on the one hand related to their high work load and on the other hand to the class organization in the schools. Students with the mathematics ability level focused on in our study were often divided over a large number of classes, which made performing the project activities rather cumbersome for the teachers. Even in schools where there were special mathematics classes, the situation often became more difficult in the course of the research project, as participating students had different rates of development and therefore became spread over a larger number of classes during the study.

5.3 Limitations and further research

In addition to the relatively small sample size, some other limitations of our study should be noted. First of all, although from the logbook data we know that all the games in the experimental intervention were treated, we do not know exactly how accurately the instructions in the teacher manuals were adhered to. Therefore, our results should be interpreted as the effect of the intervention as it was implemented by the teachers on the basis of our instructions. This is inherent to studying a teacher-delivered intervention and doing research in the school practice. In the regular school practice, it is also the case that teacher guidelines in textbooks can be followed more or less accurately. The previously mentioned possibility that teachers in special education are not so familiar with instructional methods based on experiential learning might imply, however, that more instruction or training for the teachers could have been helpful.

A further limitation is that we do not know what other mathematics instruction, apart from our intervention, was given in the participating schools. We did instruct teachers to spend the same amount of time on each mathematics domain as they would normally do, but we cannot be sure whether they indeed did this, or whether this normal situation was the same for the classes in the control group as for the classes in the experimental group (the pacing of the curriculum is usually not fixed in special education schools).[7] This implies that our results should be taken with caution. A study involving a closer monitoring of instructional activities outside the intervention would enable stronger conclusions.

Another limitation concerns the employed Mult tests for measuring multiplicative reasoning ability. Because we did not want to overburden the students by giving them too many test items, the number of items in the Mult tests was too small to have separate scales for procedural and conceptual knowledge. Further research could examine whether an intervention with mathematics games such as the ones we employed differentially influences students' procedural and conceptual knowledge.

Finally, we note that the mini-games we employed, although they did meet several requirements argued to be important for special needs students (see Ke & Abras, 2013; Seo & Woo, 2010), were not specially developed for use in special education. Moreover, the sequence of games used in the intervention was not designed to form a continuing learning trajectory. Possibly, to make the mini-games intervention add to the effectiveness of the multiplicative reasoning curriculum for special education students' procedural and conceptual knowledge, specific adaptations of the games may be necessary. Further research is required here.

[7] We did ask teachers to estimate, in teacher questionnaires, the average amount of time per week spent on the different mathematics domains, including multiplicative reasoning, just as we did with the regular education teachers (see Chapter 4). However, these questionnaires were only filled in for half of the classes (even after multiple reminders), so they did not provide us with sufficient information.

5.4 Conclusion

Our results indicate that multiplicative mini-games can effectively be applied in special education to enhance students' declarative knowledge of multiplication facts beyond what is achieved with the regular multiplicative reasoning curriculum without these games. With respect to procedural and conceptual knowledge of multiplicative reasoning, the learning outcomes of the students receiving the mini-games intervention did not differ from the learning outcomes obtained with the regular curriculum. Regarding these latter types of knowledge, further research may reveal ways to make the use of mini-games more beneficial for special education students.

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Barmby, P., Harries, T., Higgins, S., & Suggate, J. (2009). The array representation and primary children's understanding and reasoning in multiplication. *Educational Studies in Mathematics, 70,* 217-241.

Bouck, E. C., & Flanagan, S. (2009). Assistive technology and mathematics: What is there and where can we go in special education. *Journal of Special Education Technology, 24*(2), 17-29.

Brown, D. J., Ley, J., Evett, L., & Standen, P. (2011, November). *Can participating in games based learning improve mathematic skills in students with intellectual disabilities?* Paper presented at IEEE 1st International Conference on Serious Games and Applications for Health, Braga, Portugal.

Cawley, J., Parmar, R., Foley, T. E., Salmon, S., & Roy, S. (2001). Arithmetic performance of students: Implications for standards and programming. *Exceptional Children, 67,* 311-328.

Christensen, C. A., & Gerber, M. M. (1990). Effectiveness of computerized drill and practice games in teaching basic math facts. *Exceptionality: A Special Education Journal, 1,* 149-165.

De Vos, T. (2010). *TempoTest Automatiseren* [Tempotest automatization]. Amsterdam: Boom test uitgevers.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33,* 441-467.

Goldman, S. R., & Hasselbring, T. S. (1997). Achieving meaningful mathematics literacy for students with learning disabilities. *Journal of Learning Disabilities, 30,* 198-208.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549-576.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences, 20,* 169-206.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore, MD: Johns Hopkins University Press.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8* [Scientific justification of the LOVS mathematics tests for Grade 1 to 6]. Arnhem, Netherlands: Cito.

Jonker, V., Wijers, M., & Van Galen, F. (2009, October). *The motivational power of mini-games for the learning of mathematics*. Paper presented at the European Conference on Game Based Learning, Graz, Austria.

Ke, F., & Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology, 44,* 225-242.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55,* 427-443.

Klawe, M. M. (1998, June). *When does the use of computer games and other interactive multimedia software help students learn mathematics?* Paper presented at the Technology and NCTM Standards 2000 Conference, Arlington, VA.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.

Kraemer, J.-M., Van der Schoot, F., & Van Rijn, P. (2009). *Balans van het reken-wiskundeonderwijs in het speciaal basisonderwijs: Uitkomsten van de derde peiling in 2006* [Monitor of the mathematics education in special primary education: Outcomes of the third audit in 2006]. Arnhem, Netherlands: Cito.

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24,* 97-114.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22,* 215-243.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4,* 333-369.

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research & Practice, 22,* 47-57.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles: Author.

Okolo, C. M. (1992). The effect of computer-assisted instruction format and initial attitude on the arithmetic facts proficiency and continuing motivation of students with learning disabilities. *Exceptionality: A Special Education Journal, 3,* 195-211.

Parsons, H. M. (1974). What happened at Hawthorne? *Science, 183,* 922-932.

Seo, Y.-J., & Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education, 53,* 913-928.

Seo, Y.-J., & Woo, H. (2010). The identification, implementation, and evaluation of critical user interface design features of computer-assisted instruction programs in mathematics for students with learning disabilities. *Computers & Education, 55,* 363-377.

Shiah, R.-L., Mastropieri, M. A., Scruggs, T. E., & Fulk, B. J. M. (1994). The effects of computer-assisted instruction on the mathematical problem solving of students with learning disabilities. *Exceptionality: A Special Education Journal, 5,* 131-161.

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1-67.

Van den Heuvel-Panhuizen, M. (Ed.). (2008). *Children learn mathematics: A learning-teaching trajectory with intermediate attainment targets for calculation with whole numbers in primary school.* Rotterdam, Netherlands: Sense.

Woodward, J. (2004). Mathematics education in the United States: Past to present. *Journal of Learning Disabilities, 37,* 16-31.

Woodward, J., & Rieth, H. (1997). A historical review of technology research in special education. *Review of Educational Research, 67,* 503-536.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). ACER ConQuest (Version 2.0) [Computer Software]. Camberwell, Australia: Australian Council for Educational Research.

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *The Journal of Special Education, 32,* 207-225.

Chapter 6

The development of primary school students’ attitude towards mathematics and the relation with mathematics achievement, gender, and playing mathematics computer games

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. The development of primary school students’ attitude towards mathematics and the relation with mathematics achievement, gender, and playing mathematics computer games.

The development of primary school students' attitude towards mathematics and the relation with mathematics achievement, gender, and playing mathematics computer games

1 Introduction

Many mathematics educators have noted the relevance of students' attitude towards mathematics as a factor in mathematics education (e.g., Leder & Forgasz, 2006; McLeod, 1992, 1994). Interest in and enjoyment of mathematics can facilitate mathematics learning and thereby enhance mathematics achievement (e.g., Aunola, Leskinen, & Nurmi, 2006; Viljaranta, Lerkkanen, Poikkeus, Aunola, & Nurmi, 2009), and can lead students to choosing more or higher level mathematics courses later in their school career (e.g., Gottfried, Marcoulides, Gottfried, & Oliver, 2013; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Watt, 2006). Gottfried et al. (2013) even found that students' attitude towards mathematics is related to the general educational level they eventually attain.

Apart from the possible influence of mathematics attitude on mathematics achievement, there may also be an influence of students' mathematics achievement on their mathematics attitude (e.g., Neale, 1969). This means that achievement and attitude may develop cyclically, with each influencing each other (e.g., Aunola et al., 2006; Fisher, Dobbs-Oates, Doctoroff, & Arnold, 2012; Viljaranta et al., 2009). To make this cycle a positive one, a goal of mathematics education should be not only to enhance students' mathematics achievement, but also to foster students' attitude towards mathematics. Researchers have argued that interventions to promote students' attitude can best be implemented early in the educational career, such that possible negative cycles of attitude and achievement can be remedied at an early stage (e.g., Gottfried, 1990; Gottfried et al., 2013). One possible means of promoting students' attitude is having them play educational computer games, as games are known for their motivating characteristics (e.g. Malone, 1981; Prensky, 2001).

In the present study, we investigated the development of Dutch early primary school students' attitude towards mathematics, its relation to mathematics achievement, and its relation to playing mathematics computer games. Furthermore, we looked at the role of gender, as gender has been an important variable in much previous work on attitude towards mathematics (see, e.g., Hyde, Fennema, Ryan, Frost, & Hopp, 1990).

1.1 Attitude towards mathematics

Many different terms and definitions have been used to describe attitude or motivation related to mathematics (see, e.g., Di Martino & Zan, 2010; Leder & Forgasz, 2006; McLeod, 1992). The term *attitude towards mathematics* has often been used as an umbrella term denoting several aspects of a person's view of mathematics, such as liking mathematics, beliefs of one's ability in mathematics, beliefs of the importance and usefulness of mathematics, and mathematics anxiety (e.g., Fennema & Sherman, 1976; McLeod, 1994; Neale, 1969). This has led to a rather broad use of the term, which, in several cases, leaves unclear which aspects of attitude are referred to. To clarify the term, McLeod (1992), in an influential overview article, proposed to view attitude towards mathematics as denoting relatively stable positive or negative affective responses towards mathematics, and to distinguish it from beliefs, which he argues to be more of a cognitive nature. As examples of attitudes, McLeod mentions liking, disliking, being curious, and being bored about particular mathematics topics. In the present study, we adopted McLeod's definition of mathematics attitude, which coincides with the emotional dimension of attitude as distinguished by Di Martino and Zan (2010). Specifically, in this study we viewed attitude towards mathematics as the enjoyment, or liking, of the subject of mathematics. This conceptualization is similar to the *intrinsic or interest value* component of task value in the Expectancy-Value model by Eccles et al. (1983), and the concept of *intrinsic motivation*, defined as engaging in an activity for its own sake (e.g., Ryan & Deci, 2000). Enjoyment of mathematics has also been referred to with terms like *task motivation* (Aunola et al., 2006; Nurmi & Aunola, 2005; Viljaranta et al., 2009) and *interest* (e.g., Fisher et al., 2012; Frenzel, Goetz, Pekrun, & Watt, 2010; Lerkkanen et al., 2012). A related but slightly different construct is Gottfried's (e.g., 1990) *academic intrinsic motivation*, which can be seen as the enjoyment of the learning of mathematics.

In the following, we use the terms *attitude*, *liking*, *enjoyment*, and *interest* interchangeably. When we refer to literature in which another conceptualization of attitude was used (e.g., including beliefs), we specifically indicate this.

1.2 Development of mathematics attitude over time

Previous research has shown that primary school students' attitude towards mathematics is generally quite positive (e.g., Dowker, Bennett, & Smith, 2012; Mullis, Martin, Foy, & Arora, 2012). However, this finding does not apply for all ages. From studies involving students of different ages, it is known that mathematics attitude is generally less positive in older students than in younger students (e.g., Eccles et al., 1983; Mata, Monteiro, & Peixoto, 2012; Mullis et al., 2012). Longitudinal studies have shown that students gradually lose their enjoyment of (learning) mathematics over time (e.g., Ahmed, Van der Werf, Kuyper, & Minnaert, 2013; Fredricks & Eccles, 2002; Frenzel et al., 2010; Gottfried, Fleming, & Gottfried, 2001; Ma & Cartwright, 2003). This decreasing pattern of attitude towards mathematics has been found to start already in the early grades of primary school (e.g., Fredricks & Eccles, 2002; Krinzinger, Kaufmann, & Willmes, 2009; Nurmi & Aunola, 2005), and to continue over the school career (e.g., Fredricks & Eccles, 2002; Gottfried et al., 2001; see also Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). Such a declining trend has also been found for attitudes towards other school subjects and towards learning and school in general (e.g., Dotterer, McHale, & Crouter, 2009; Gottfried et al., 2001; Hornstra, Van der Veen, Peetsma, & Volman, 2013). However, attitude towards mathematics seems to decrease faster than other school-related attitudes (Gottfried et al., 2001).

1.3 Mathematics attitude and mathematics achievement

The relationship between attitude towards mathematics and achievement in mathematics should be viewed as a potentially bidirectional or reciprocal one (e.g., McLeod, 1992; Neale, 1969). On the one hand, when a student is more positive about mathematics, he/she will probably put more effort and concentration in mathematics activities, which may enhance the student's learning of mathematics and thus their achievement. On the other hand, students' achievement in mathematics may influence their attitude. Students who perform better generally have more positive emotional experiences during mathematics activities – for example, they experience less frustration and have more feelings of competence (e.g., Eccles et al., 1983; Ryan & Deci, 2000) – which may lead to a more positive attitude (e.g., Eccles et al., 1983; McLeod, 1992).

Relations between attitude towards mathematics and mathematics achievement have frequently been found, although generally these relations are quite weak (see, e.g., Neale, 1969). In a meta-analysis, Ma and Kishor (1997) found a significant positive overall effect size of $r = .12$. For primary school students, the effect size was only $r = .03$, but still statistically significant. The studies included in this meta-analysis, however, differed widely in their conceptualization of attitude towards mathematics, which makes it hard to interpret the findings. Yet, also in studies in which the enjoyment of mathematics was examined as a separate construct, a positive relation was found with achievement in mathematics (e.g., Grootenboer & Hemmings, 2007; Mata et al., 2012; Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Mullis et al., 2012). Consistent with Ma and Kishor's (1997) finding that the relation between mathematics attitude and mathematics achievement was stronger in older students, Mortimore et al. (1988) found the relation between enjoyment and achievement in mathematics to increase slightly with age (from age 8 to age 10).

Many of the studies investigating the relationship between attitude and achievement in mathematics just examined concurrent correlations – that is, correlations of attitude and achievement measured at the same time point – and did not study whether it is attitude that influences achievement or achievement that influences attitude, or both. To do this, cross-lagged analyses of longitudinal data can be employed, in which influences of earlier achievement on later attitude are tested while controlling for earlier achievement and influences of earlier attitude on later achievement are tested while controlling for earlier attitude (see, e.g., Ma & Xu, 2004). Using such analyses, some studies in the early primary school years (Gottfried, 1990) and in kindergarten (Aunola et al. 2006; Viljaranta et al., 2009) found evidence for prior enjoyment of (learning) mathematics predicting subsequent mathematics achievement, as well as prior mathematics achievement predicting subsequent enjoyment. Gottfried (1990), however, argued that achievement might be a more consistent predictor of enjoyment than the other way around. This idea was further supported in a longitudinal study by Gottfried, Marcoulides, Gottfried, Oliver, and Guerin (2007), in which mathematics achievement at age 9 was found to predict later enjoyment of learning mathematics, whereas enjoyment of learning mathematics did not predict later achievement. Also in Krinzinger et al.'s (2009) study with early primary school students, mathematics achievement more clearly predicted attitude towards mathematics than the other way around (in this study attitude was measured as a combination of liking mathematics and competence beliefs). In contrast, Nurmi and Aunola's (2005) study with first- and second-graders provided some evidence for enjoyment of mathematics predicting subsequent mathematics performance, whereas no evidence was found for mathematics performance predicting subsequent enjoyment.

1.4 Mathematics attitude and gender

Much research on attitude towards mathematics has focused on gender differences. As Hyde et al. (1990) pointed out, underlying this research is the idea that gender differences in mathematics attitude can explain later gender differences in mathematics performance and in the selection of mathematics courses and mathematics-related occupations. In their meta-analytic study, Hyde et al. indeed found gender differences in attitude towards mathematics, with boys having a slightly more positive attitude than girls. This gender difference tended to be somewhat larger in older students. However, Hyde et al. used a broad definition of attitude, and the meta-analysis included very few studies that separately investigated enjoyment of mathematics. In a cross-national meta-analysis of PISA data (15-year-old students; Else-Quest, Hyde, & Linn, 2010), enjoyment towards mathematics was studied as a separate construct, and also here a significant gender difference was found in favor of boys ($d = 0.20$). Similar results were reported by, for example, Frenzel et al. (2010), Skaalvik and Skaalvik (2004), and Watt (2006), but in some other studies no differences between secondary school boys and girls were found in the liking of mathematics (e.g., Mata et al., 2012; Eccles et al., 1983). Also in studies in primary school and kindergarten, results varied somewhat. In several studies, boys and girls did not differ in their mathematics enjoyment (e.g., Lerkkanen et al., 2012; Meelissen et al., 2012; Mortimore et al., 1988), but also here, when differences were found they tended to be in favor of boys (e.g., Bouffard, Marcoux, Vezeau, & Bordeleau, 2003; Meelissen & Luyten, 2008; Nurmi & Aunola, 2005). In contrast to the findings for attitude towards mathematics, for the case of attitude towards school in general, or towards the subject of reading, gender differences are generally found in favor of girls (e.g., Mortimore et al., 1988; West, Hailes, & Sammons, 1997).

Apart from gender differences in attitude towards mathematics at particular time points, boys and girls may also differ in their developmental pattern of mathematics attitude. Gender socialization theories hypothesize that gender differences in attitudes increase with age as a result of increasing experiences with stereotypic gender roles (e.g., Eccles, 1987).

This hypothesis was supported by Bouffard et al.'s (2003) study, in which gender differences in mathematics attitude increased from first to third grade, due to a decrease in attitude over time in girls but not in boys. Other studies, however, failed to provide such support: Fredricks and Eccles (2002) found a marginally significantly larger decrease in mathematics interest for boys than for girls (from Grade 1 to Grade 12), whereas Frenzel et al. (2010) found a similar decrease for both genders (from Grade 5 to Grade 9).

1.5 Mathematics attitude and mathematics computer games

As attitude towards mathematics has been found to predict students' later mathematics achievement (e.g., Aunola et al., 2006; Viljaranta et al., 2009) and their selection of mathematics courses and mathematics-related careers (e.g., Gottfried et al., 2013; Nagy et al., 2006; Watt, 2006), it is worthwhile to try to improve students' mathematics attitude. According to McLeod (1992), when students experience positive emotions related to the subject of mathematics, this can positively influence their attitude towards mathematics. Similarly, according to Hidi and Renninger's (2006) model of interest development, a person's interest towards a particular topic can change in response to experiences with the topic.

One of the educational activities that have been suggested to positively influence students' attitudes is the use of computers (e.g., McLeod, 1992; West et al., 1997), and, specifically, educational computer games. Computer games are often praised for their potential to motivate students (e.g., Malone, 1981; Prensky, 2001). Some reasons are that computer games can provide immediate feedback and explicit goals, can invoke curiosity, and can give the students a sense of control (Malone & Lepper, 1987). Although the motivational benefit of games is generally seen as a game-specific motivation (students like to engage in learning with the game), this momentary, situational motivation can, according to the theories by McLeod (1992) and Hidi and Renninger (2006), lead to an increase in students' more general motivation or attitude towards the subject, through the positive feelings with the subject that are experienced while working with the game. For this to occur, an important prerequisite is that the learning content of the game is intrinsically integrated in the main gameplay activity (e.g., Habgood & Ainsworth, 2011; see also Malone & Lepper, 1987), because then the positive emotions experienced while playing the game are most clearly linked to the learning content and thus to the subject.

In two recent meta-analyses of the effectiveness of educational computer games (Vogel et al., 2006; Wouters, Van Nimwegen, Van Oostendorp, & Van der Spek, 2013), motivational effects were examined. Vogel et al. (2006) reported a significant positive effect of computer games on students' attitudes, and in Wouters et al.'s (2013) meta-analysis the overall effect of educational games on students' motivation, although non-significant, was of non-trivial size ($d = 0.26$). In both these meta-analyses, however, no distinctions were made between game-specific motivation, attitudes towards technology, and attitudes towards the subject.

From Wouters et al.'s descriptions of incorporated studies it appears that most of them actually measured game-specific motivation. Evidence on the influence of games on students' subject-related attitudes seems to be rather scarce.

For the subject of mathematics, there have been a few studies that showed a positive influence of playing educational computer games on students' attitude towards mathematics. Afari, Aldridge, Fraser, and Khine (2013) found such an influence for college students, whereas Ke (2008), Ke and Grabowski (2007) and Pilli and Aksu (2013) reported positive attitudinal effects of working with (a digital learning environment including) mathematics computer games in fourth-grade students. Van Eck (2006), in a study with secondary school students, also found some evidence for the positive influence of playing mathematics computer games on attitude towards mathematics, but only regarding mathematics anxiety and only in one of several gameplay conditions. It should be noted that in most of these studies mathematics attitude was conceptualized as a combination of different aspects, including enjoyment and beliefs. Only Afari et al. and Van Eck separately examined the effect on students' enjoyment of mathematics, which was found to be significant in Afari et al.'s study.

We know of no studies examining the influence of playing mathematics computer games on students' mathematics attitude in the early years of primary education. Furthermore, in the aforementioned studies, students were only followed for rather short periods of time (ranging from 3 days to 3 months).

1.6 Our study

In the current study, we examined the development of primary school students' attitude towards mathematics from Grade 1 to Grade 4, using a longitudinal design. We tested a possible linear trend in this development, we compared the developmental pattern for mathematics attitude with the development of attitudes towards other domains, we looked at gender differences, and we investigated the cross-lagged relations between mathematics attitude and mathematics achievement. Finally, we studied the influence of playing mathematics computer games on students' attitude towards mathematics.

Our study aimed to extend the still quite limited research base on the development of students' attitude towards mathematics in primary school with a longitudinal study in the Netherlands, where, to our knowledge, such a study has not been carried out before. A rather novel aspect of the present study is our investigation of the longitudinal relations between playing mathematics computer games and students' mathematics attitude.

Our research questions were as follows:

1. *How does students' attitude towards mathematics develop in the early grades of primary school?*

2. *How is this development related to the development of other school-related attitudes?*

3. *How is this development related to students' gender?*

4. *How is students' attitude towards mathematics related to their achievement in mathematics?*

5. *Does the extent of playing mathematics computer games influence students' attitude towards mathematics?*

With respect to Research question 1, our hypothesis was that the students' attitude towards mathematics would, on average, be positive, but would decrease over the grades. Regarding Research question 2, we hypothesized attitude towards mathematics to decrease faster than other school-related attitudes. For the case of Research question 3, we did not specify a hypothesis, because of the mixed findings in earlier studies. Furthermore, regarding Research question 4, we hypothesized that mathematics achievement and attitude at the same time points would be correlated to each other, and also that prior mathematics achievement would predict subsequent mathematics attitude and that prior mathematics attitude would predict subsequent mathematics achievement. Finally, for Research question 5 our hypothesis was that the extent of playing mathematics computer games would positively influence students' later attitude towards mathematics.

2 Method

2.1 Context and design of the study

The data for the current study were collected as part of a research project on the effects of mathematics computer games in the domain of multiplicative reasoning (multiplication and division) (see Bakker, Van den Heuvel-Panhuizen, & Robitzsch, 2014 [Chapter 4 of this thesis]; Bakker, Van den Heuvel-Panhuizen, Van Borkulo, & Robitzsch, 2013 [Chapter 3 of this thesis]). In this project, a large sample of students was followed from the end of Grade 1 to the end of Grade 4. During these three years, students' attitude towards mathematics and their mathematics achievement were assessed at six time points: at the end of Grade 1 (T1), in the middle and at the end of Grade 2 (T2 and T3), in the middle and at the end of Grade 3 (T4 and T5), and at the end of Grade 4 (T6). Thus, the time distance between time points was approximately half a year for the period from T1 to T5, and approximately one year from T5 to T6. In Grade 2 and Grade 3, the students played mathematics computer games, either on the domain of multiplicative reasoning or on other mathematics topics.

2.2 Participants

In recruiting schools for our study, we contacted Dutch primary schools by phone (response rate ca. 15%), e-mail (response rate ca. 2%), and an advertisement on a mathematics games website. We found 66 schools to be willing to participate, with a total of 1661 Grade 1 students. Of these schools, only those 45 schools that kept participating in the project at least until the end of the computer game intervention period were included in the currently presented analysis. The participating schools were situated in both urban and more rural regions of the Netherlands and were located in a range of socioeconomic areas. In most of the schools, there was one class that participated in the study; nine schools had two participating classes. We included in our analysis only those students who were in a participating class during the whole game intervention period (students who entered the class later or left the class earlier were excluded). Of the resulting 935 students, three students were excluded because they completed less than half of the six attitude measures and/or less than half of the six achievement measures. Our analysis sample thus included 932 students (482 boys, 450 girls). Their mean age was 7.2 years ($SD = 0.4$) at the end of Grade 1.

In answering Research question 5, we only included the students who were offered computer games on the topic of multiplicative reasoning, because only for these students data were collected on the extent to which they played the games. Therefore, for this research question, the sample was smaller and consisted of 606 students (320 boys, 286 girls; $M = 7.2$ years, $SD = 0.4$) from 29 schools.

2.3 Attitude questionnaire

2.3.1 Questionnaire format and procedure

Students' attitude toward mathematics and other domains was measured using an attitude questionnaire that was specifically developed for this study. The same questionnaire was used at all six time points. To be able to easily collect data on many students, the questionnaire was administered online, through the Digital Mathematics Environment (DME).[1]

The questionnaire consisted of 40 items, each involving a five-point rating scale ranging from *I hate this* (1) to *I like this very much* (5). The scale categories were visualized by smileys (see Figure 1 for a sample item), to make them easily understandable for young children. The questionnaire mainly focused on mathematics attitude but also contained items for measuring attitude towards reading and towards school in general, as well as some

[1] The DME has been developed by our colleague Peter Boon at the Freudenthal Institute of Utrecht University.

items on various school subjects other than mathematics and reading, and some items on out-of-school activities.

Each item was represented by a picture, with an accompanying instruction that was read aloud by the computer. This was done to avoid any influence of reading difficulties. The spoken text could be repeated by clicking on a loudspeaker button. The questionnaire started with a practice item ("Going to the dentist. Do you like this or not?") to explain the meaning of the five smileys. Here it was also explained that the student should listen carefully to each question and not base their answers solely on the pictures.

The administration of the test was organized by the classroom teachers. Students completed the questionnaires individually, using headphones. The duration of the questionnaire was, on average, approximately 10 minutes.

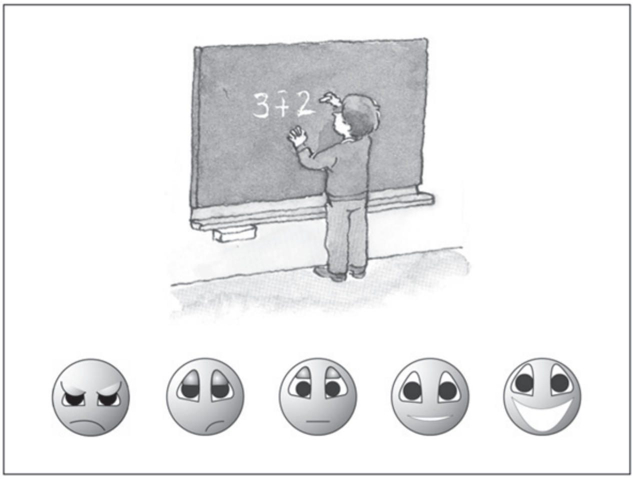


Figure 1. Sample item of the attitude questionnaire. Spoken instruction: "Solving mathematics problems yourself on the blackboard. Do you like this or not? Click on the smiley that shows how much you like this."

2.3.2 Attitude scales

To identify attitude scales in our attitude questionnaire, we first performed an exploratory factor analysis for each time point, including all 40 questionnaire items. For attitude towards mathematics, two or three different factors were found, depending on the time point. However, as the main interest of our study was to investigate a general attitude towards mathematics, we decided to use in our analysis a single mathematics attitude scale containing all 18 items loading on one or more of these mathematics factors. This is in agreement with Reise, Moore and Haviland's (2010) suggestion that it is possible to use a unidimensional scoring even when a factor analysis reveals multiple dimensions. According to Reise et al., this is allowed when the general factor emerging in fitting a bifactor model is of sufficient importance compared to all further uncorrelated factors, which was the case for our mathematics attitude scale (as indicated by the omega hierarchical which ranged from .47 to .65 for the six time points; $M = .56$).

Next to the factors related to mathematics attitude, at all time points three additional factors emerged: a reading factor (a scale of three items), a school factor (a scale of two items), and a holiday factor (a scale of two items). As the holiday scale had a rather low reliability (Cronbach's alpha range for the six time points .15–.50, $M = .31$), we did not use it in our analysis. The reliability coefficients of the mathematics, reading, and school scale were sufficiently high (mathematics scale: Cronbach's alpha range .85–.88, $M = .86$; reading scale: .64–.73, $M = .68$; school scale: .73–.81, $M = .77$). Table 1 lists the items of these three scales, including the mean scores on the items, and the factor loadings (averaged over the six time points).

In addition to the two holiday items, 15 other items of the questionnaire were not used in our analysis, because they either did not clearly load on one of the factors or the allocation to factors differed between time points.

2.4 Mathematics achievement tests

Students' mathematics achievement was measured using the standardized mathematics tests from the Cito student monitoring system (see Janssen, Verhelst, Engelen, & Scheltens, 2010). These tests belonged to the schools' regular assessment program and were administered by the teachers in (approximately) the same months as the attitude questionnaire. The tests contain mathematics problems presented in a context as well as bare number problems, and cover various mathematics domains, such as addition and subtraction, multiplication and division, and measurement. Each test from the Cito monitoring system is different, but their scores are scaled on a common ability scale (Janssen et al., 2010). The tests can be administered either as a paper-and-pencil test or as a digital test, of which the scores are comparable. The reliability coefficients of the tests range from .91 to .97 (Janssen et al., 2010).

Table 1
Content description of the items of the mathematics, reading, and school attitude scale, including item mean scores and factor loadings

Item: Liking of …	Item mean score *M* (*SD*)	Factor loading *M* (*SD*)
Mathematics attitude scale		
1. Having a mathematics lesson	3.44 (0.13)	0.83 (0.05)
2. Doing mathematics problems at home	2.62 (0.33)	0.76 (0.11)
3. Doing mathematics problems yourself on the blackboard	3.73 (0.26)	0.74 (0.06)
4. Doing mathematics problems in your notebook	3.31 (0.24)	0.93 (0.06)
5. Drawing shapes in your notebook	3.52 (0.18)	0.50 (0.08)
6. Doing mathematics problems in a row	3.38 (0.17)	0.97 (0.03)
7. Doing mathematics problems with the number line	2.83 (0.38)	0.75 (0.08)
8. Doing mathematics problems with rods and blocks	3.39 (0.07)	0.59 (0.04)
9. Doing mathematics problems about money	3.74 (0.16)	0.71 (0.06)
10. Doing mathematics problems with the arithmetic rack	2.72 (0.27)	0.61 (0.08)
11. Doing mathematics problems about the clock	3.28 (0.26)	0.73 (0.06)
12. Doing problems about measurement	3.34 (0.15)	0.69 (0.04)
13. Doing addition problems	3.85 (0.19)	0.89 (0.06)
14. Doing subtraction problems	2.95 (0.13)	0.88 (0.05)
15. Doing multiplication problems	3.70 (0.22)	0.83 (0.05)
16. Doing mathematics problems on the computer	4.19 (0.17)	0.56 (0.06)
17. Doing mathematics problems with a calculator	3.99 (0.05)	0.41 (0.08)
18. Drawing shapes on the computer	4.01 (0.09)	0.34 (0.04)
Average	*3.44 (0.19)*	*0.71 (0.06)*
Reading attitude scale		
1. Having a reading lesson	3.27 (0.11)	0.72 (0.04)
2. Reading a book yourself in class	3.71 (0.05)	1.04 (0.03)
3. Reading at home	3.76 (0.10)	0.87 (0.04)
Average	*3.58 (0.09)*	*0.87 (0.03)*
School attitude scale		
1. Being at school	3.40 (0.10)	-[a]
2. Being in the classroom	3.34 (0.08)	-[a]
Average	*3.37 (0.09)*	-

Note. Item mean scores and factor loadings were computed – after employing multiple imputation (see section 2.6) – for the six time points; the mean and standard deviation over these six time points is presented. Factor loadings were obtained using one-dimensional confirmatory factor analyses.
[a]Because the scale has only two items, factor loadings cannot be computed.

2.5 Mathematics games

2.5.1 Games and procedure

The mathematics games were played in four game periods, which were in the first and second half of Grade 2 and in the first and second half of Grade 3 (see Figure 2). Each game period lasted 10 weeks, in which eight different games were offered to the students. The 29 schools included in the analysis on the relation between the extent of playing mathematics games and students' attitude towards mathematics, were divided over three conditions: in E1 (8 schools) the games were played in class, integrated in a lesson; in E2 (12 schools) the games were played at home without attention in class; and in E3 (9 schools) the games were played at home followed by a discussion in class. The games were offered online, through the Digital Mathematics Environment (DME).

	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Grade 1										T1
Grade 2	Game period 1				T2	Game period 2				T3
Grade 3	Game period 3				T4	Game period 4				T5
Grade 4										T6

Figure 2. Time schedule of the research project.

The games provided to the students were mini-games, which are short, focused games that are easy to learn (e.g., Jonker, Wijers, & Van Galen, 2009). These mini-games focused on the mathematics domain of multiplicative reasoning (multiplication and division) and aimed both at practicing multiplicative facts and operations, and at discovering new concepts and strategies related to multiplication and division, through experiential learning. The games contained several motivating features, such as immediate feedback, opportunities for choice, and curiosity-provoking aspects (see Malone & Lepper, 1987). Furthermore, the mathematics content was intrinsically integrated in the games (see Habgood & Ainsworth, 2011).

2.5.2 Students' gameplay behavior

Through the DME we monitored the students' online gameplay behavior, that is, the extent to which they played the games. As an indication of this extent of gameplay, Table 2 reports the time students spent on the games, for each condition and each game period. As is shown by the wide ranges of values and the relatively large standard deviations, the extent of gameplay varied widely between students. In the E1 condition this was because schools implemented the game intervention to various degrees, while in the E2 and E3 conditions this was because the students could decide for themselves whether and for how much time they played the games. These varieties in gameplay behavior could be reflected in students' subsequent attitude towards mathematics. Of course we acknowledge the possibility of a reversed relation: mathematics attitude could also have influenced students' gameplay, with students with higher attitudes probably having spent more time on the games, at least in the home-playing conditions. This relation was controlled for in our analyses.

Table 2
Time students spent on the games (in minutes) in the three conditions

	Condition E1 (n = 168)				Condition E2 (n = 253)				Condition E3 (n = 185)			
Game period	*M*	*SD*	*Mdn*	Range	*M*	*SD*	*Mdn*	Range	*M*	*SD*	*Mdn*	Range
1	147	65	149	0–503	93	183	32	0–1633	98	160	55	0–1105
2	170	73	180	0–455	21	59	0	0–671	40	94	6	0–825
3	141	62	136	6–255	7	28	0	0–277	26	61	0	0–420
4	98	72	94	0–275	3	14	0	0–128	16	67	0	0–567

To have a multi-facetted measure of the students' gameplay behavior, we used a combination of four types of values derived from the collected log data: the time spent on a game in seconds (Time), the number of attempted mathematics problems in a game (Effort), the number of correct attempts (Success), and the number of games played in a game period (NumberOfGames). To diminish the impact of some extreme outliers in the Time, Effort, and Success variables, logarithmic transformations ($f(x) = \log(x + 1)$) were employed. The transformed values were then z-standardized and, subsequently, for each student we computed weighted sums of these variables for each game period, with the weights based on the mean amount of time students spent on each game. The NumberOfGames variable was computed for each game period as the number of different games the student played, ranging from 0 to 8. Because our four measures of gameplay behavior were highly correlated in each game period (correlations ranging from .79 to .97, $p < .001$), we computed for each game period a composite measure of gameplay by z-standardizing and then averaging the four measures. As these composite measures were rather skewed (skewnesses ranging from -0.23 to 1.36), for each measure students were ranked after which their rank scores were z-standardized to obtain normally distributed measures. This resulted in the gameplay variables Gplay1 to Gplay4, referring to the four game periods.

2.6 Missing data

As is insurmountable in a large-scale longitudinal study carried out in the school practice, not all data were available for all students. Some students missed one or more attitude questionnaires, and for some students we did not receive all mathematics achievement test scores. The percentage of missing attitude questionnaires ranged from 1.2% (T2) to 14.3% (T6); the percentage of missing mathematics achievement test scores ranged from 0.1% (T2) to 20.6% (T5). Furthermore, within an attitude questionnaire, students sometimes skipped some items. The average percentage of missing items in questionnaires that were not entirely missing ranged from 1.4% (T5) to 4.6% (T1).

We employed multiple data imputation to make estimates for the missing values (see Graham, 2009). We specified an imputation model containing the item answers of the attitude scales of all six attitude questionnaires, the six mathematics achievement test scores, some student background variables, and, for the students included in the gameplay analysis, the Gplay variables. Because the gameplay data could have a different relation to attitude in the different gameplay conditions, the imputation procedure was performed for each condition separately. To account for the clustered data structure (students nested within schools), also school mean scores on the mathematics achievement tests were included in the imputation model. The data imputation was performed using the “mice” software (Van Buuren & Groothuis-Oudshoorn, 2011), and resulted in 20 imputed datasets. After data imputation, for each time point the scores on the mathematics, reading, and school attitude scale were computed by averaging the item scores of the items belonging to these scales. Statistical analyses were run on the 20 datasets and results were combined using Rubin’s rule (see Graham, 2009).

2.7 Data analysis

All analyses were performed in Mplus (Muthén & Muthén, 1998-2010), using maximum likelihood estimation. To account for the clustered data structure (students within schools), we employed cluster-robust standard errors (see Angrist & Pischke, 2009).

For answering Research question 1, 2, and 3, we examined the mean developmental trends of attitudes over the six time points. To investigate decreases or increases over time, these developmental trends were modeled as a linear function.[2]

For answering Research question 4, we employed a cross-lagged path model (see, e.g., Bollen & Curran, 2006), in which influences of prior mathematics achievement on subsequent mathematics attitude, as well as of prior mathematics attitude on subsequent achievement, could be tested controlling for the influence of prior attitude on subsequent attitude and of prior achievement on subsequent achievement. For Research question 5, this cross-lagged model was extended by including the variables Gplay1 to Gplay4, such that the influence of gameplay behavior on subsequent attitude could be tested controlling for earlier achievement and attitude, and for the influence of earlier attitude and achievement on gameplay behavior. For both Research question 4 and Research question 5, first an unrestricted model was specified, after which some paths were constrained to be equal to examine average influences over all time points (Research question 4 and 5) or over the all three gameplay conditions (Research question 5). For all path models we present standardized path coefficients β. These β values are practically equivalent to *r* effect sizes, for which the values .10, .30, and .50 can be interpreted as a small, medium, and large effect, respectively (Cohen, 1988).

To evaluate the fit of the models, we computed the χ^2 goodness-of-fit statistic, using Allison's (2001) combination rules for multiply imputed data sets. The χ^2 statistic signifies a good model fit when it is non-significant. One should note, however, that with large sample sizes a significant value is easily found, even when there is only a slight discrepancy between the model and the data. Unfortunately, for multiple imputation data, there are no established methods for computing other fit statistics such as the root mean square error of approximation (RMSEA) and the comparative fit index (CFI) (see, e.g., Enders, 2010), which are generally more informative for large samples. Yet, to give an indication of model fit beyond the χ^2 statistic, we used an ad hoc procedure to obtain values for RMSEA, CFI, and the χ^2/df ratio, by directly computing them from the χ^2 values of the tested models (and baseline models). We note that the RMSEA, CFI, and χ^2/df values we report should be taken with caution. For RMSEA a good fit is denoted by a value smaller than .05, whereas values between .05 and .08 indicate reasonable fit (Browne & Cudeck, 1993). For CFI, values larger than .95 indicate a good fit (Hu & Bentler, 1999), while values between .90 and .95 are often regarded as denoting reasonable fit (e.g., Marsh, Hau, & Wen, 2004). Finally, according to Schermelleh-Engel, Moosbrugger, and Müller (2003),

[2] As Wu, West, and Taylor (2009) pointed out, when looking at mean trends it is important to disentangle the mean structure and covariance structure of the variables. In our analyses we put parametric restrictions (i.e., a linear trend) on means, with all covariances between variables being freely estimated. We did not employ ordinary latent growth curve models, because in such models the same functional specification (e.g., a linear function) is posed for both the mean and covariance structure.

a χ^2/df ratio smaller than 2 or 3 can be interpreted as denoting good or acceptable fit, respectively.

Depending on whether our hypothesis was directional or not, we used one- or two-tailed significance tests. In cases where no specific hypotheses were specified, or when multiple equalities were tested at once using a Wald chi-square test, two-tailed tests were employed.

3 Results

3.1 Development of students' attitude towards mathematics

Table 3 presents the means and standard deviations of the scores on the mathematics attitude scale at each of the six time point. At all six time points, the mean attitude towards mathematics was higher than 3, the midpoint of the scale, but lower than 4. This indicates that the students had, on average, a moderately positive attitude towards mathematics. The thick line graph in Figure 3 displays the mean development of students' attitude towards mathematics over the time points. With the exception of a small increase from T1 (End Grade 1) to T2 (Mid Grade 2), we see an overall decreasing pattern.

Table 3
Means and standard deviations of the mathematics, reading, and school attitude scale at the six time points (N = 932)

	Attitude scale					
	Mathematics		Reading		School	
Time point	*M*	*SD*	*M*	*SD*	*M*	*SD*
T1 (end Grade 1)	3.58	0.81	3.71	1.11	3.42	1.37
T2 (mid Grade 2)	3.62	0.75	3.59	1.06	3.48	1.28
T3 (end Grade 2)	3.48	0.79	3.50	1.08	3.37	1.24
T4 (mid Grade 3)	3.48	0.76	3.63	1.03	3.43	1.24
T5 (end Grade 3)	3.30	0.79	3.53	1.05	3.27	1.21
T6 (end Grade 4)	3.20	0.71	3.52	1.00	3.26	1.12

First of all, a Wald chi-square test (equivalent to an ANOVA *F* test) indicated that the means of students' mathematics attitude at the six time points differed from each other ($\chi^2(5) = 58.72$, $p < .001$). Pair-wise comparisons of mean attitudes at successive time points revealed significant differences between T2 and T3 ($t = -3.60$, $p < .001$, $d = -0.18$), between

T4 and T5 ($t = -5.00$, $p < .001$, $d = -0.22$), and between T5 and T6 ($t = -2.78$, $p = .005$, $d = -0.14$), each representing a decrease in attitude. The other differences were not significant (T1 vs.T2: $t = 0.95$, $p > .10$, $d = 0.05$; T3 vs.T4: $t = -0.18$, $p > .10$, $d = -0.01$).[3]

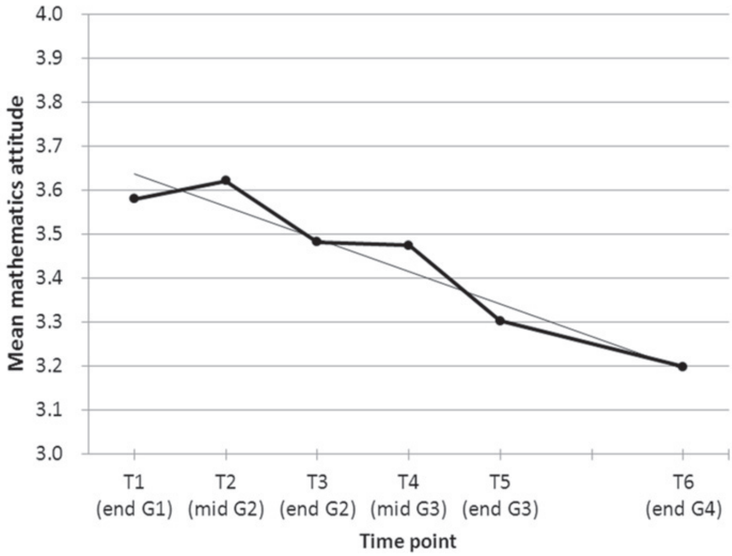


Figure 3. Mean attitude towards mathematics at the six time points, with a linear trend line. G = Grade.

To test whether the developmental pattern reflected an overall decrease in attitude over time, we specified a linear model of the means at the six time points, as is visualized by the thin line in Figure 3. We found a significant negative slope ($B = -0.074$, $SE = 0.006$, $p < .001$), which confirmed our hypothesis that attitude towards mathematics decreases over time. The average decrease was 0.10 standard deviation per half a year. Fit statistics in general showed that the linear model fitted the data reasonably well ($\chi^2(4) = 18.19$, $p = .001$, $RMSEA = .062$, $CFI = .973$, $\chi^2/df = 4.55$), although the χ^2/df ratio was quit low.[4] In addition, a linear regression of the six means resulted in a rather high R^2 value of explained variance of .87. Together, this indicated that the developmental trend for mathematics attitude could well be approximated by a downward linear trend.

[3] The d effect sizes were computed by dividing the difference in means by the average standard deviation over the six time points ($SD = 0.77$).
[4] Fit statistics were computed here by comparison to a null model of no change over time, i.e., with the means of the six time points constrained to be equal (see Widaman & Thompson, 2003).

3.2 Development of mathematics attitude compared to other attitudes

In Figure 4 the developmental pattern of students' mathematics attitude is displayed together with the developmental patterns of students' attitudes towards reading and school (for exact means and standard deviations, see Table 3). First of all, we compared the mean mathematics attitude at the six time points to the means of the other attitudes, using Wald chi-square tests. Averaged over the six time points, we found that mathematics attitude significantly differed from the other attitudes, with students' attitude towards mathematics being lower, on average, than their attitude towards reading ($\chi^2(6) = 60.60$, $p < .001$, average $d = -0.18$), but slightly higher than their attitude towards school ($\chi^2(6) = 31.08$, $p < .001$, average $d = 0.09$).[5] Furthermore, we examined the developmental rate of the attitudes towards reading and school by specifying a linear model of means. As was the case for mathematics attitude, also for the two other attitudes we found a significant negative slope (reading: $B = -0.028$, $SE = 0.011$, $p = .010$; school: $B = -0.039$, $SE = 0.009$, $p < .001$). Comparing the slope of mathematics attitude with the other slopes revealed that the slope for mathematics attitude was significantly steeper than for the other attitudes

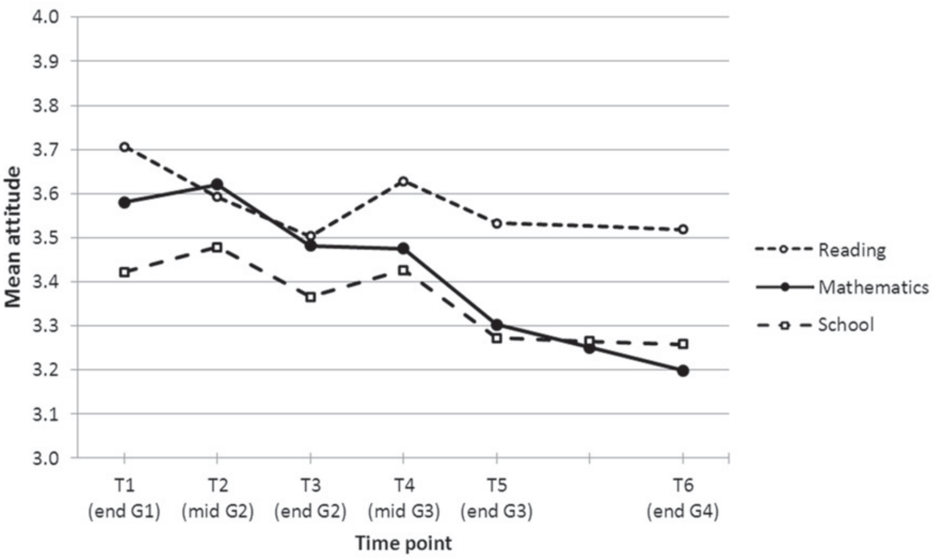


Figure 4. Mean attitude towards mathematics, reading, and school, at the six time points.

[5] Average *d* effect sizes were computed by first, for each time point, dividing the difference between the means of the two attitude scales being compared by the standard deviation of the mathematics attitude scale, and then taking the average over the six time points.

(mathematics vs. reading: $B = -0.046$, $t = -5.17$, $p < .001$; mathematics vs. school: $B = -0.035$, $t = -4.11$, $p < .001$), indicating that mathematics attitude decreased fastest.

3.3 Gender and development of mathematics attitude

Figure 5 displays the mean development of mathematics attitude for boys and girls separately (boys' and girls' means and standard deviations are given in Table 4). From this graph it seems that girls had higher attitudes towards mathematics than boys had, which was confirmed by a Wald chi-square test comparing the mathematics attitude of boys and girls over all time points together ($\chi^2(6) = 13.93$, $p = .030$). In fact, at each individual time point, except T3, girls had a significantly higher mathematics attitude than boys ($p < .05$), as is shown in Table 4. The average *d* effect size over all time points was -0.17.

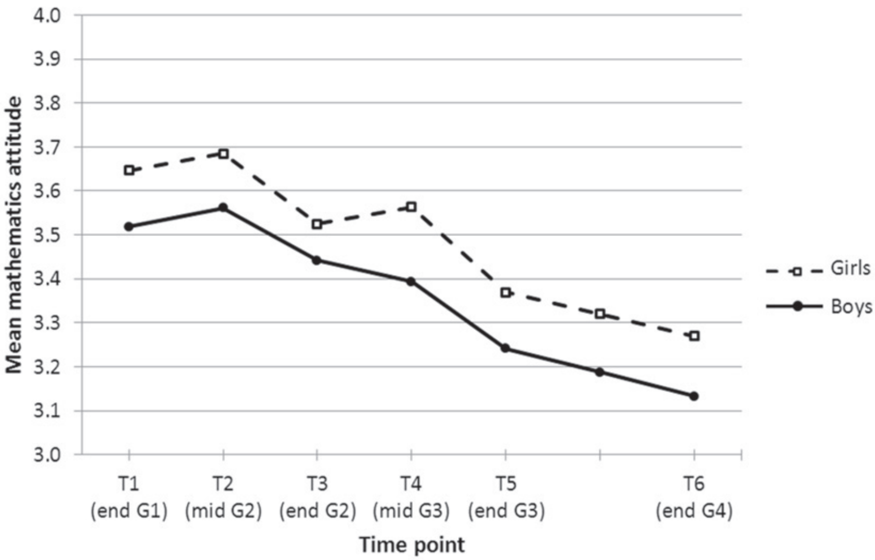


Figure 5. Boys' and girls' mean attitude towards mathematics at the six time points. G = Grade.

Figure 5 suggests that over time the boys' and girls' attitude decreased to the same extent. When we modeled boys' and girls' mean attitudes separately as linear models and compared the slopes, we indeed found no significant difference in slopes (boys: $B = -0.074$, $SE = 0.008$; girls: $B = -0.074$, $SE = 0.008$; $t = -0.00$, $p = .998$). Thus, the difference between boys and girls did not change over time, which was also apparent from the relatively similar effect sizes of gender differences at the different time points, as reported in Table 4.

Table 4
Comparison of boys' and girls' attitude towards mathematics at each time point

	Boys (n = 482)		Girls (n = 450)		Boys vs. Girls[a]	
Time point	*M*	*SD*	*M*	*SD*	*t*	*d*
T1 (end Grade 1)	3.52	0.82	3.65	0.79	-2.30*	-0.16
T2 (mid Grade 2)	3.56	0.78	3.69	0.71	-2.68**	-0.16
T3 (end Grade 2)	3.44	0.81	3.53	0.76	-1.57	-0.11
T4 (mid Grade 3)	3.39	0.81	3.56	0.69	-3.03**	-0.22
T5 (end Grade 3)	3.24	0.82	3.37	0.74	-2.39*	-0.16
T6 (end Grade 4)	3.13	0.71	3.27	0.71	-3.26**	-0.19

[a]A negative value represents a difference in favor of girls.
* $p < .05$. ** $p < .01$. Two-tailed.

To put into perspective our finding that girls liked mathematics more than boys did, we also looked at gender differences for the other attitude scales. Also for reading and school, girls showed a more positive attitude than boys (reading: $\chi^2(6) = 33.20$, $p < .001$, average $d = -0.42$; school: $\chi^2(6) = 70.01$, $p < .001$; average $d = -0.46$). Here, as indicated by the d values, the difference between boys and girls was much larger than for attitude towards mathematics.

3.4 Relation between mathematics attitude and mathematics achievement

Table 5 displays the concurrent correlation between students' attitude towards mathematics and their mathematics achievement at each time point. As expected, correlations were positive and generally statistically significant ($p < .05$), except for the last time point T6. The correlations were, however, quite weak and appeared to decrease over time.

Table 5
Concurrent correlation between mathematics attitude and mathematics achievement at each time point

	Time point					
	T1	T2	T3	T4	T5	T6
Correlation	.158***	.123**	.130***	.074*	.062*	.018

* $p < .05$. ** $p < .01$. *** $p < .001$. One-tailed.

To examine the influence of mathematics attitude on later mathematics achievement, and the influence of mathematics achievement on later mathematics attitude, we specified a cross-lagged path model, as displayed in Figure 6. This model included all paths from attitude and achievement at time point T to attitude and achievement at the next time point $T + 1$, and in this way allowed us to investigate the influence of prior achievement on subsequent attitude, and of prior attitude on subsequent achievement, while controlling for the influence of prior achievement on subsequent achievement and of prior attitude on subsequent attitude. Furthermore, we controlled for the concurrent correlation between attitude and achievement at each time point. To make paths comparable, we did not include time point T6 in this model, as the time distance between T5 and T6 was different from the other time distances.

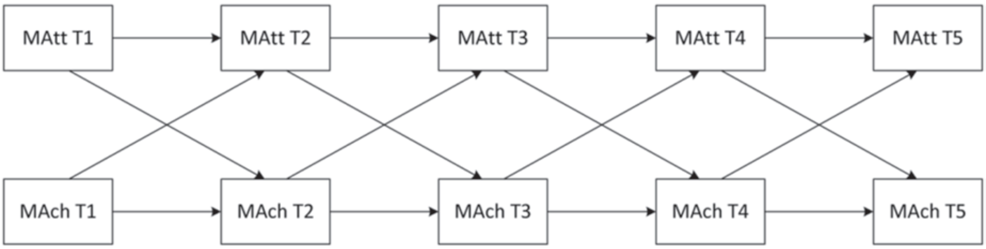


Figure 6. Initial cross-lagged path model of the relations between mathematics attitude (MAtt) and mathematics achievement (MAch).

The model displayed in Figure 6 did not have a good fit, especially when looking at CFI and the χ^2/df ratio ($\chi^2(24) = 123.236$, $p < .001$, $RMSEA = .067$, $CFI = .695$, $\chi^2/df = 5.13$). To improve model fit, we included paths from attitude at time point T to attitude at time point $T + 2$, and from achievement at time point T to achievement at time point $T + 2$. The fit of the thus obtained model was acceptable ($\chi^2(18) = 43.49$, $p < .001$, $RMSEA = .039$, $CFI = .922$, $\chi^2/df = 2.42$). Figure 7 displays the final model, with its standardized path coefficients. Descriptives of all variables in the model and their correlations are provided in Appendix A.

From the path coefficients it appears, firstly, that both mathematics achievement and attitude towards mathematics were rather stable over time. This is evidenced by the large coefficients of the paths from earlier to later measurements of these constructs.

As shown in Figure 7, the cross-lagged paths from mathematics attitude to subsequent mathematics achievement were not significant ($p > .10$), indicating that prior mathematics attitude did not influence later mathematics achievement. Regarding the paths from

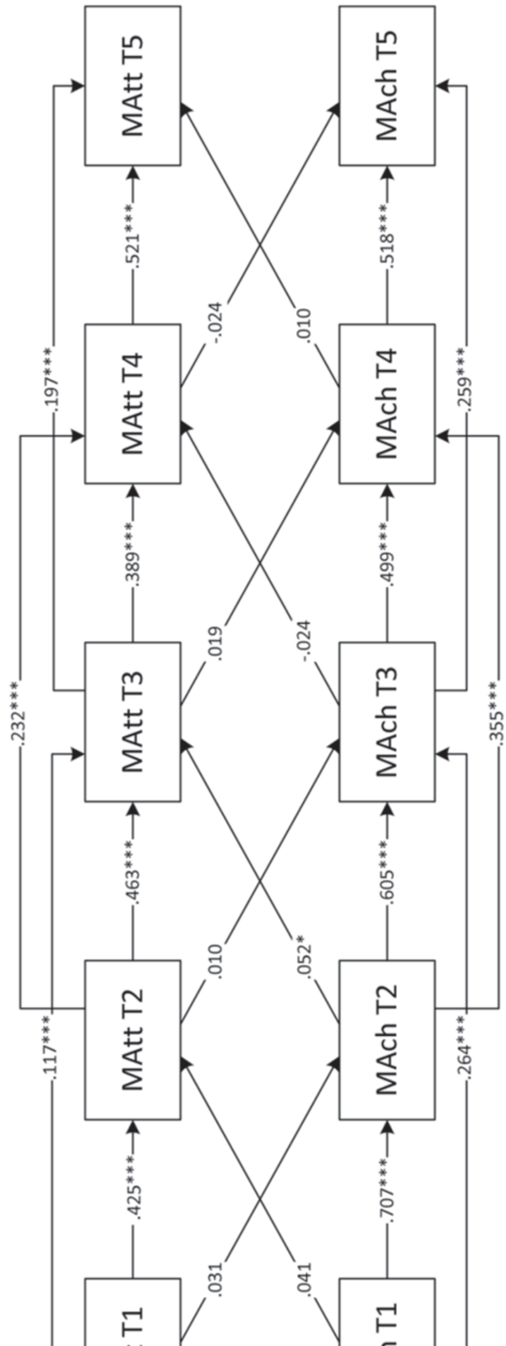


Figure 7. Cross-lagged path model of the relations between mathematics attitude (MAtt) and mathematics achievement (MAch), with standardized path coefficients (β). Concurrent correlations between MAtt and MAch at the same time point were controlled for.
* *p* < .05. *** *p* < .001. One-tailed.

mathematics achievement to subsequent mathematics attitude, we found one significant path: from mathematics achievement at T2 (mid Grade 2) to mathematics attitude at T3 (end Grade 2; $\beta = .052$, $SE = .031$, $p = .046$).

To investigate the average influence, over all time points, of mathematics attitude on subsequent mathematics achievement, and of mathematics achievement on subsequent mathematics attitude, we specified an alternative model constraining the paths from mathematics achievement to subsequent mathematics attitude to be equal, and constraining the paths from mathematics attitude to subsequent mathematics achievement to be equal. For this restricted model, the fit was similar as for the non-constrained model ($\chi^2(24) = 55.25$, $p < .001$, $RMSEA = .037$, $CFI = .904$, $\chi^2/df = 2.30$), indicating that the restricted model was a reasonable alternative. The model revealed a marginally significant average effect of mathematics achievement on subsequent mathematics attitude ($\beta = .019$, $SE = .014$, $p = .090$), indicating that, on average over the period from end Grade 1 to end Grade 3, earlier mathematics achievement tended to predict later mathematics attitude, controlling for earlier mathematics attitude. The *r* effect size, indicated by the β value, was, however, very small. Regarding the influence of attitude towards mathematics on subsequent mathematics achievement we did not find a significant average effect ($\beta = .009$, $SE = .012$, $p > .10$).

3.5 Influence of gameplay behavior on mathematics attitude

To investigate the influence of students' gameplay behavior on students' attitude towards mathematics, we added the gameplay variables Gplay1 to Gplay4 to the model in Figure 7, leading to the model displayed in Figure 8. This model enabled us to examine the influence of gameplay behavior on subsequent mathematics attitude while controlling for influences of earlier attitude on gameplay behavior and for the possible role of mathematics achievement. The paths of interest for our research question are the (thick) paths from Gplay to MAtt.

Because the extent to which the games were played and the way this extent was related to attitude and achievement probably differed across the different gameplay conditions, we specified the model for each of the three gameplay conditions (E1, E2, and E3) separately. Each model was specified two times, once without further restrictions, and once with all paths from Gplay to subsequent mathematics attitude constrained to be equal. The latter model was meant to examine the average influence of gameplay on subsequent mathematics attitude over the four game periods. We also specified two multi-group models, with the three conditions as the groups, in which we constrained the paths from Gplay to MAtt in the three conditions to be equal. With these models we could test per game period, and over all game periods, the average influence of gameplay on mathematics attitude over the three conditions together. All models fitted the data well (all χ^2 values n.s., *RMSEA* range 0–.019, *CFI* range .993–1, χ^2/df range 0.69–1.07). The model results are presented in Table 6. For brevity, this table only displays the coefficients of the paths from Gplay to MAtt. Descriptives of and correlations between all variables in the models are included in Appendix B.

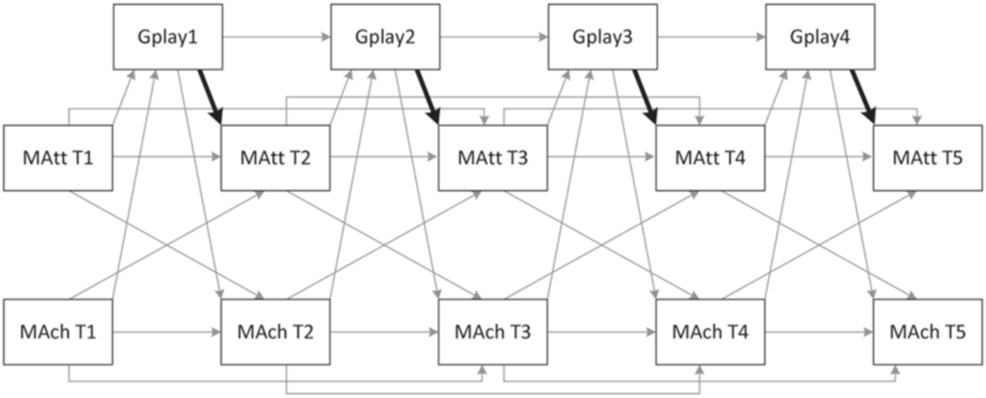


Figure 8. Path model for investigating the influence of gameplay behavior on mathematics attitude.

The model results indicate that, averaged over all three conditions and over all four game periods, gameplay behavior had a significant positive influence on subsequent attitude towards mathematics ($\beta = .044$, $SE = .020$, $p = .014$). The r effect size, indicated by the β value, was, however, very small. When looking at the four game periods separately (averaged over conditions) we see that for each separate game period there was no significant influence of gameplay on subsequent attitude ($p > .05$). However, for Game period 1 the influence was marginally significant ($\beta = .061$, $SE = .038$, $p = .052$). Furthermore, when looking at the three conditions separately (averaged over game periods), the influence of gameplay behavior on mathematics attitude was only significant in the E3 condition, where the games were played at home and afterwards discussed in class ($\beta = .049$, $SE = .021$, $p = .017$). In the E1 condition, where the games were played in class, the influence of gameplay behavior on mathematics attitude over all game periods was marginally significant ($\beta = .039$, $SE = .017$, $p = .066$); for this condition a significant positive influence was found in Game period 1 ($\beta = .100$, $SE = .043$, $p = .018$) and Game period 3 ($\beta = .128$, $SE = .050$, $p = .007$). Finally, in the E2 condition, where the games were played at home without attention in class, there was no significant average influence of gameplay on attitude ($p > .10$). Although the overall influence of gameplay on mathematics attitude seemed to differ between conditions, this difference was not significant (Wald $\chi^2(2) = 0.54$, $p > .10$).

Table 6
Standardized coefficients (β) of the paths from Gplay to subsequent mathematics attitude (MAtt)

Path	Per condition[a] E1 (n = 168)	E2 (n = 253)	E3 (n = 185)	Averaged over conditions[b] (n = 606)
Per game period[c]				
Gplay1 → MAtt T2	.100*	.048	.024	.061[†]
Gplay2 → MAtt T3	-.057	.084	.041	.034
Gplay3 → MAtt T4	.128**	-.047	.032	.022
Gplay4 → MAtt T5	.000	.017	.078[†]	.048
Averaged over game periods[d]				
Gplay → MAtt	.039[†]	.024	.049*	.044*

[a]Separate model for each condition. [b]Multi-group model with Gplay → MAtt paths for the three conditions constrained to be equal. Average β values were obtained by first standardizing each variable by dividing it by its pooled *SD* for the three conditions together. [c]No restrictions on Gplay → MAtt paths. [d]Gplay → MAtt paths of the four game periods constrained to be equal. Average β values were obtained by first standardizing each variable by dividing it by either its condition-specific *SD* or its pooled *SD*, for the per-condition and the averaged-over-condition models, respectively.
[†] $p < .10$. * $p < .05$. ** $p < .01$. One-tailed.

4 Discussion

4.1 Overview and interpretation of our findings

Our results show that Dutch students' attitude towards mathematics in Grade 1 to Grade 4 was moderately positive. This outcome is in agreement with previous research results on primary school students in the Netherlands (e.g., Meelissen et al., 2012) and in other countries (e.g., Dowker et al., 2012; Mullis et al., 2012). Regarding the development of this attitude (Research question 1) we found that, in accordance with our expectation, students' mathematics attitude decreased with age. This result corroborates earlier findings of such a decrease in primary school students (e.g., Krinzinger et al., 2009; Nurmi & Aunola, 2005) and is in line with Fredricks and Eccles' (2002) finding that the decrease in attitude towards mathematics is an ongoing process that already starts early in the school career.

When we compared students' attitude towards mathematics with their attitudes in other domains (Research question 2), we found that attitude towards mathematics was, on average, less positive than attitude towards reading, but slightly more positive than attitude towards school in general. All three attitudes decreased over time, but the decrease in attitude towards mathematics was largest, as was also found by Gottfried et al. (2001).

The comparison of boys and girls (Research question 3) revealed that girls, on average, had a more positive attitude towards mathematics than boys had (average $d = -0.17$). This is a surprising finding, given the fact that in most previous studies gender differences in enjoyment of mathematics were either non-significant (e.g., Meelissen et al., 2012; Viljaranta et al., 2009) or in favor of boys (e.g., Bouffard et al., 2003; Else-Quest et al., 2010; Meelissen & Luyten, 2008; Nurmi & Aunola, 2005). Also for attitude towards reading and towards school in general we found that girls outscored boys, which matches earlier research findings (e.g., Mortimore et al., 1988; West et al., 1997). Apparently, in the early primary school grades in the Netherlands, the more positive attitude of girls towards learning and school is also reflected in their attitude towards mathematics. Another outcome of our study was that boys and girls did not differ in the developmental trend of their mathematics attitude over time; for both genders mathematics attitude declined to the same extent. Thus, the gender difference in favor of girls did not change over time. This result differs from Bouffard et al.'s (2003) finding that girls' attitude towards mathematics decreased from Grade 1 to Grade 3, but boys' attitude did not. It also deviates from Fredricks and Eccles' (2002) finding that attitude toward mathematics decreased more for boys than for girls. Taken together, research findings on gender differences in the development of attitude towards mathematics seem rather inconsistent.

In our examination of the relation between attitude towards mathematics and achievement in mathematics (Research question 4), we, first of all, found small but, most of the times, significant concurrent correlations between mathematics attitude and achievement (rs ranging from .018 to .158). These small relations between attitude and achievement parallel earlier research findings (e.g., Ma & Kishor, 1997). Somewhat remarkably, in our study the relation between attitude and achievement tended to get weaker when children got older. In contrast, other studies showed that this relation slightly increased with age (e.g., Mortimore et al., 1988).

When looking at the cross-lagged relations between mathematics attitude and mathematics achievement, we did not find influences of prior mathematics attitude on subsequent mathematics achievement. We did, however, find a very small, marginally significant average influence ($\beta = .019$) of prior mathematics achievement on subsequent mathematics attitude (this effect was significant for the path from time point T2 to T3). Thus, our study failed to provide evidence for a cyclic relationship in which both achievement and attitude predict each other, as was found by, for example, Aunola et al. (2006) and Viljaranta et al. (2009). Rather, our results are in line with some earlier findings suggesting that mathematics achievement is more predictive of attitude than the other way around (e.g., Gottfried, 1990; Gottfried et al., 2007; Krinzinger et al., 2009).

A final result from our study was that the extent to which students played mathematics computer games (Research question 5) had, on average over the four time points and the three gameplay conditions, a significant positive influence on subsequent attitude towards mathematics (controlling for the possible influence of earlier attitude on extent of gameplay). This influence was very weak though ($\beta = .044$), and was not consistently found for separate game periods or separate gameplay conditions. Nevertheless, the overall result lends some support to the idea that playing mathematics computer games can positively influence students' attitude towards mathematics. This finding adds to the still limited research base on the potential of mathematics computer games in promoting students' mathematics attitude (e.g., Ke, 2008; Ke & Grabowski, 2007).

4.2 Implications of our findings

Our findings indicate that children have a quite positive attitude towards mathematics at the beginning of their school career, but that this attitude gradually decreases over the grades. Such a decrease was also apparent for other school-related attitudes, yet for mathematics the decrease was more pronounced. Although in our study students' attitude towards mathematics did not predict their later achievement in mathematics, having a positive attitude towards mathematics is important for course and career selection (e.g., Nagy et al., 2006; Watt, 2006) and attained educational level (Gottfried et al., 2013). Furthermore, a positive attitude towards mathematics can be considered an important educational outcome in its own right, as it contributes to students' well-being. This all means that the decrease in mathematics attitude is a serious issue. Teachers should be aware of this developmental decline and should think of ways of maintaining students' initial positive attitude towards mathematics. As our results suggest, the use of computer games in mathematics education may help in fostering students' attitude towards mathematics. However, as the influence of playing computer games was found to be very weak, one should, in addition, think of other ways of promoting students' mathematics attitude, such as eliciting changes in practices of teachers (e.g., Aunola et al., 2006; Lerkkanen et al., 2012) or parents (e.g., Gottfried, Fleming, & Gottfried, 1994).

4.3 Limitations and further research

Because our study only involved students in the early primary school grades in the Netherlands, prudence is called for in making generalizations to other grades or other countries. Regarding the relation of gameplay with enjoyment of mathematics, one should bear in mind that our results may depend on the types of games used and the instructional settings in which the games were deployed, which makes generalizations to other games or other instructional settings tentative.

Furthermore, in our study we only focused on students' enjoyment or liking of mathematics. Other affective variables, such as competence beliefs and beliefs about the usefulness of mathematics, are also important factors in mathematics education (e.g., Di Martino & Zan, 2010; Eccles et al., 1983; McLeod, 1992). Further research could extend our findings by looking at the development of early primary school children's beliefs about mathematics, as has been done for example by Fredricks and Eccles (2002) and Jacobs et al. (2002). Especially the relationship between playing mathematics games and children's mathematics beliefs is an interesting direction for further study, as this relationship has not been explicitly studied before.

Another limitation might lie in the fact that we used the same questionnaire at each time point. Although using the same instrument at multiple time points is common practice in longitudinal repeated measurement designs, this might carry the problem that an instrument is more suitable for some age groups than for others. In our case, some of the questionnaire items were illustrated by mathematics problems typical for Grade 1 (see Figure 1) or referred to auxiliary materials most commonly used in the lower grades, such as rods and blocks (Item 8) and the arithmetic rack (Item 10) (see Table 1). Mathematics domains that were introduced in later grades, such as division and decimal numbers, were not included in the questionnaire. This may explain the reported decrease in correlations between students' attitude towards mathematics and their mathematics achievement, as in the later grades the items in mathematics attitude scale were less connected to the contents of the mathematics curriculum. Furthermore, the use of mathematics content from Grade 1 may have influenced our finding of a decrease in mathematics attitude over the grades. Thus, this finding should be taken with some caution. However, for attitude towards reading and towards school, for which the items were not illustrated with pictures characteristic for the lower grades, we also found a decrease. Thus, most likely the found decrease in attitude towards mathematics reflects an actual decline in attitude.

An additional point is that we used only a questionnaire to measure students' attitude towards mathematics. Although questionnaires are most convenient for assessing large numbers of students, other methods such as interviews may provide more in-depth information on students' feelings about mathematics (see, e.g., Leder & Forgasz, 2006; McLeod, 1992).

A final point that should be made is related to our examination of the influence of playing mathematics computer games on students' mathematics attitude. In our study, we investigated the role of gameplay by examining whether students who played more had a higher subsequent mathematics attitude than those who played less (controlling for the possible effect of earlier mathematics attitude on the extent of gameplay). Though this is a valid way of testing the influence of playing mathematics computer games on students' attitude towards mathematics, a more direct way would be to compare students receiving a gameplay intervention with students in a no-game condition. Therefore, to further clarify the possible effect of playing computer games on the enjoyment of mathematics in the early primary school years additional studies are necessary, which include a no-game condition.

In such studies, larger effects may be found than in our study, as the difference between students in their extent of gameplay is, by design, larger.

4.4 Conclusion

We think our study is a valuable contribution to the research field of attitude towards mathematics. To our knowledge, this is the first longitudinal study in the Netherlands on the development of primary school students' enjoyment of mathematics. Our results provide additional evidence of the developmental decrease of students' attitude towards mathematics, and further clarify the relationship between mathematics attitude and mathematics achievement by showing a tendency for a causal path from prior mathematics achievement to later attitude, but not from prior attitude to later achievement. Furthermore, an interesting finding in our study is the girls' advantage in attitude towards mathematics. Finally, our study adds some evidence to the still rather limited research base on the potential of mathematics computer games in fostering students' attitude towards mathematics.

References

Afari, E., Aldridge, J. M., Fraser, B. J., & Khine, M. S. (2013). Students' perceptions of the learning environment and attitudes in game-based mathematics classrooms. *Learning Environments Research, 16,* 131-150.

Ahmed, W., Van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, self-regulated learning, and achievement in mathematics: A growth curve analysis. *Journal of Educational Psychology, 105,* 150-161.

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Aunola, K., Leskinen, E., & Nurmi, J.-E. (2006). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *British Journal of Educational Psychology, 76,* 21-40.

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2014). *Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability.* Manuscript submitted for publication.

Bakker, M., Van den Heuvel-Panhuizen, M., Van Borkulo, S., & Robitzsch, A. (2013). Effecten van online mini-games op multiplicatieve vaardigheden van leerlingen in groep 4 [Effects of online mini-games on second-graders' multiplicative reasoning ability]. *Pedagogische Studiën, 90*(3), 21-36.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective.* Hoboken, NJ: Wiley.

Bouffard, T., Marcoux, M.-F., Vezeau, C., & Bordeleau, L. (2003). Changes in self-perceptions of competence and intrinsic motivation among elementary schoolchildren. *British Journal of Educational Psychology, 73,* 171-186.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Cohen, J. (1988*). Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Di Martino, P., & Zan, R. (2010). 'Me and maths': towards a definition of attitude grounded on students' narratives. *Journal of Mathematics Teacher Education, 13,* 27-48.

Dotterer, A. M., McHale, S. M., & Crouter, A. C. (2009). The development and correlates of academic interests from childhood through adolescence. *Journal of Educational Psychology, 101,* 509-519.

Dowker, A., Bennett, K., & Smith, L. (2012). Attitudes to mathematics in primary school children, *Child Development Research.* doi:10.1155/2012/124939

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11,* 135-172.

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75-146). San Francisco: W. H. Freeman.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103-127.

Enders, C. K. (2010). *Applied missing data analysis.* New York: Guilford Press.

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education, 7,* 324-326.

Fisher, P. H., Dobbs-Oates, J., Doctoroff, G. L., & Arnold, D. H. (2012). Early math interest and the development of math skills. *Journal of Educational Psychology, 104,* 673-681.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38,* 519-533.

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence, 20,* 507-537.

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology, 82,* 525-538.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (1994). Role of parental motivational practices in children's academic intrinsic motivation and achievement. *Journal of Educational Psychology, 86,* 104-113.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93,* 3-13.

Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. H. (2013). Longitudinal pathways from math intrinsic motivation and achievement to math course accomplishments and educational attainment. *Journal of Research on Educational Effectiveness, 6,* 68-92.

Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H., & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. International *Journal of Behavioral Development, 31,* 317-327.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549-576.

Grootenboer, P., & Hemmings, B. (2007). Mathematics performance and the role played by affective and background factors. *Mathematics Education Research Journal, 19*(3), 3-20.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences, 20,* 169-206.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41,* 111-127.

Hornstra, L., Van der Veen, I., Peetsma, T., & Volman, M. (2013). Developments in motivation and achievement during primary school: A longitudinal study on group-specific differences. *Learning and Individual Differences, 23,* 195-204.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1-55.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14,* 299-324.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73,* 509-527.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenen-wiskunde voor groep 3 tot en met 8* [Scientific justification of the LOVS mathematics tests for Grade 1 to 6]. Arnhem, Netherlands: Cito.

Jonker, V., Wijers, M., & Van Galen, F. (2009, October). The motivational power of mini-games for the learning of mathematics. In M. Pivec (Ed.), *Proceedings of the 3rd European conference on games based learning* (pp. 203–210). Reading, UK: Academic Publishing.

Ke, F. (2008). Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. *Educational Technology Research and Development, 56,* 539-556.

Ke, F., & Grabowski, B. (2007). Gameplaying for maths learning: cooperative or not? *British Journal of Educational Technology, 38,* 249-259.

Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment, 27,* 206-225.

Leder, G. C., & Forgasz, H. J. (2006). Affect and mathematics education: PME perspectives. In A. Gutiérrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 403-427). Rotterdam, the Netherlands: Sense.

Lerkkanen, M.-K., Kiuru, N., Pakarinen, E., Viljaranta, J., Poikkeus, A.-M., Rasku-Puttonen, H., ... & Nurmi, J.-E. (2012). The role of teaching practices in the development of children's interest in reading and mathematics in kindergarten. *Contemporary Educational Psychology, 37,* 266-279.

Ma, X., & Cartwright, F. (2003). A longitudinal analysis of gender differences in affective outcomes in mathematics during middle and high school. *School Effectiveness and School Improvement, 14,* 413-439.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education, 28,* 26-47.

Ma, X., & Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education, 110,* 256-280.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4,* 333–369.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: Vol. 3 Conative and affective process analyses* (pp. 223-253). Hillsdale, NJ: Lawrence Erlbaum.

Marsh H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11,* 320-341.

Mata, M. L., Monteiro, V., & Peixoto, F. (2012). Attitudes towards mathematics: Effects of individual, motivational, and social support factors. *Child Development Research.* doi:10.1155/2012/876028

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). New York: Macmillan.

McLeod, D. B. (1994). Research on affect and mathematics learning in the JRME: 1970 to the present. *Journal for Research in Mathematics Education, 25,* 637-647.

Meelissen, M., & Luyten, H. (2008). The Dutch gender gap in mathematics: Small for achievement, substantial for beliefs and attitudes. *Studies in Educational Evaluation, 34,* 82-93.

Meelissen, M. R. M., Netten, A., Drent, M., Punter, R.A., Droop, M., & Verhoeven, L. (2012). PIRLS- en TIMSS-2011. *Trends in leerprestaties in lezen, rekenen en natuuronderwijs* [PIRLS and TIMSS 2011. Trends in learning outcomes in reading, mathematics, and science]. Nijmegen/Enschede, the Netherlands: Radboud Universiteit/Universiteit Twente.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters.* Berkeley, CA: University of California Press.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user’s guide* (6th ed.). Los Angeles: Author.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12,* 323-345.

Neale, D. C. (1969). The role of attitudes in learning mathematics. *The Arithmetic Teacher, 16,* 631-640.

Nurmi, J.-E., & Aunola, K. (2005). Task-motivation during the first school years: A person-oriented approach to longitudinal data. *Learning and Instruction, 15,* 103-122.

Pilli, O., & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in North Cyprus. *Computers & Education, 62,* 62-71.

Prensky, M. (2001). Digital Game-Based Learning. New York: McGraw-Hill.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54-67.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74. Retrieved from http://www.dgps.de/fachgruppen/methoden/mpr-online

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles, 50,* 241-252.

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1-67. Retrieved from http://www.jstatsoft.org

Van Eck, R. (2006). The effect of contextual pedagogical advisement and competition on middle-school students' attitude toward mathematics using a computer-based simulation game. *Journal of Computers in Mathematics and Science Teaching, 25,* 165-195.

Viljaranta, J., Lerkkanen, M. K., Poikkeus, A. M., Aunola, K., & Nurmi, J. E. (2009). Cross-lagged relations between task motivation and performance in arithmetic and literacy in kindergarten. *Learning and Instruction, 19,* 335-344.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34,* 229-243.

Watt, H. M. G. (2006). The role of motivation in gendered educational and occupational trajectories related to maths. *Educational Research and Evaluation, 12,* 305-322.

West, A., Hailes, J., & Sammons, P. (1997). Children's attitudes to the national curriculum at Key Stage 1. *British Educational Research Journal, 23,* 597-613.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8,* 16-37.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105,* 249-265.

Wu, W., West, S. G., & Taylor, A. B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods, 14,* 183-201.

Appendix A

Descriptives of and correlations between mathematics attitude scores (MAtt) and mathematics achievement scores (MAch) (N = 932)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. MAtt T1	3.58	0.81	–								
2. MAtt T2	3.62	0.75	**.43**	–							
3. MAtt T3	3.48	0.79	**.33**	**.52**	–						
4. MAtt T4	3.48	0.76	**.34**	**.43**	**.51**	–					
5. MAtt T5	3.30	0.79	**.32**	**.42**	**.46**	**.62**	–				
6. MAch T1	43.96	15.33	**.16**	**.11**	**.12**	.04	**.07**	–			
7. MAch T2	53.89	15.10	**.14**	**.12**	**.13**	**.07**	**.08**	**.71**	–		
8. MAch T3	64.43	14.98	**.15**	**.11**	**.13**	.05	**.07**	**.70**	**.79**	–	
9. MAch T4	73.65	15.23	**.12**	**.11**	**.13**	**.07**	**.08**	**.69**	**.75**	**.78**	–
10. MAch T5	80.68	15.52	**.09**	.07	**.08**	.03	.06	**.57**	**.64**	**.66**	**.72**

Note. Bolded correlations are significant at the $\alpha = .05$ level (two-tailed).

Appendix B

Table B1

Descriptives of and correlations between mathematics attitude scores (MAtt), mathematics achievement scores (MAch), and the gameplay variables (Gplay), for gameplay condition E1 (n = 168)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MAtt T1	3.66	0.76	–												
2. MAtt T2	3.66	0.77	**.44**	–											
3. MAtt T3	3.62	0.73	**.41**	**.59**	–										
4. MAtt T4	3.48	0.79	**.38**	**.52**	**.59**	–									
5. MAtt T5	3.31	0.76	**.35**	**.37**	**.48**	**.59**	–								
6. MAch T1	41.49	16.81	**.15**	-.01	.08	.10	.10	–							
7. MAch T2	52.72	15.22	**.13**	-.02	.09	.08	.07	**.64**	–						
8. MAch T3	65.10	15.43	**.15**	.03	.07	.03	.08	**.63**	**.78**	–					
9. MAch T4	72.52	14.74	.09	-.04	.02	.02	.07	**.59**	**.75**	**.79**	–				
10. MAch T5	80.19	15.64	**.14**	-.06	-.01	.01	.08	**.50**	**.67**	**.71**	**.73**	–			
11. Gplay1	-0.06	0.46	.08	**.12**	.06	.07	.03	**.23**	**.21**	**.32**	**.13**	**.13**	–		
12. Gplay2	0.04	0.39	**.14**	**.13**	.05	.04	.04	**.14**	**.17**	**.25**	**.09**	.08	**.91**	–	
13. Gplay3	0.13	0.17	.03	.03	.04	.14	**.07**	**.31**	**.29**	**.34**	**.30**	**.26**	**.55**	**.52**	–
14. Gplay4	0.01	0.39	-.06	-.16	-.06	.10	.05	**.20**	**.24**	**.15**	**.17**	**.26**	.18	.09	**.42**

Note. Bolded correlations are significant at the α = .05 level (two-tailed).

Table B2

Descriptives of and correlations between mathematics attitude scores (MAtt), mathematics achievement scores (MAch), and the gameplay variables (Gplay), for gameplay condition E2 (n = 253)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MAtt T1	3.46	0.85	–												
2. MAtt T2	3.58	0.73	**.38**	–											
3. MAtt T3	3.43	0.81	**.26**	**.50**	–										
4. MAtt T4	3.42	0.77	**.28**	**.43**	**.48**	–									
5. MAtt T5	3.27	0.81	**.28**	**.45**	**.42**	**.58**	–								
6. MAch T1	42.49	15.09	**.17**	**.15**	**.15**	.05	**.16**	–							
7. MAch T2	53.20	15.94	.11	**.10**	**.17**	.10	**.13**	**.76**	–						
8. MAch T3	62.61	14.47	.11	**.09**	**.17**	.10	**.16**	**.75**	**.81**	–					
9. MAch T4	72.13	16.26	.12	**.11**	**.16**	**.12**	**.12**	**.71**	**.79**	**.81**	–				
10. MAch T5	79.43	15.44	.04	.03	**.13**	.07	.08	**.63**	**.67**	**.74**	**.78**	–			
11. Gplay1	-0.80	0.56	**.15**	**.11**	.12	-.00	.10	**.18**	**.16**	**.17**	**.15**	.12	–		
12. Gplay2	-0.80	0.41	**.13**	**.23**	**.21**	.06	.14	**.14**	.13	**.17**	**.17**	**.17**	**.60**	–	
13. Gplay3	-0.85	0.31	.09	**.18**	.07	.02	.07	.07	.03	.03	.01	.03	**.27**	**.43**	–
14. Gplay4	-0.71	0.24	-.02	.07	.08	.04	.05	-.05	**-.12**	-.08	-.09	-.09	**.16**	**.24**	**.22**

Note. Bolded correlations are significant at the α = .05 level (two-tailed).

Table B3

Descriptives of and correlations between mathematics attitude scores (MAtt), mathematics achievement scores (MAch), and the gameplay variables (Gplay), for gameplay condition E3 (n = 185)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MAtt T1	3.64	0.83	–												
2. MAtt T2	3.62	0.80	**.47**	–											
3. MAtt T3	3.46	0.78	**.38**	**.50**	–										
4. MAtt T4	3.48	0.78	**.38**	**.40**	**.41**	–									
5. MAtt T5	3.36	0.81	**.31**	**.38**	**.42**	**.71**	–								
6. MAch T1	43.07	14.38	**.20**	**.17**	**.14**	.00	.06	–							
7. MAch T2	53.90	15.05	**.19**	**.20**	**.16**	.02	.04	**.78**	–						
8. MAch T3	63.35	14.58	**.15**	.15	**.18**	-.02	.00	**.69**	**.78**	–					
9. MAch T4	72.54	13.69	.12	**.17**	.11	-.03	-.01	**.71**	**.76**	**.80**	–				
10. MAch T5	81.34	16.26	**.18**	.16	.03	-.03	.02	**.55**	**.64**	**.62**	**.66**	–			
11. Gplay1	-0.64	0.52	.07	.07	**.12**	.01	-.02	**.25**	**.31**	**.22**	**.30**	.13	–		
12. Gplay2	-0.64	0.44	.06	**.16**	**.12**	.01	.00	.17	**.21**	.15	**.20**	.08	**.60**	–	
13. Gplay3	-0.63	0.41	.03	**.16**	.05	.08	.00	**.17**	**.17**	.13	**.24**	.13	**.36**	**.46**	–
14. Gplay4	-0.64	0.31	.02	.10	.07	.01	.09	.00	.06	-.05	.08	.05	**.27**	**.37**	**.33**

Note. Bolded correlations are significant at the $\alpha = .05$ level (two-tailed).

Chapter 7

Summary and conclusion

Summary and conclusion

In spite of the many promises of computer games for education, the empirical research base on the effectiveness of educational computer games is still quite sparse. This is the case for educational computer games in general, as well as for the subject of mathematics specifically. Many earlier studies investigating effects of games were small-scale, did not include a control group, or did not use random assignment to conditions. Recent review articles call for large-scale, longitudinal experiments carried out in the school practice. This recommendation is in line with the recent policy of evidence-based education: educational innovations should be based on rigorous empirical research of what works in educational practice.

To obtain such empirical evidence for the effectiveness of computer games in the mathematics domain of multiplicative reasoning, we carried out a longitudinal research project, called the BRXXX-project. We examined the effects of online mini-games, which were focused on gaining multiplicative fact knowledge and operation skills (through practicing), as well as on gaining insight in multiplicative number relations (through exploration and experimentation in the games). The mini-games we used were mostly adapted versions of multiplicative mini-games selected from the Dutch mathematics games website Rekenweb (www.rekenweb.nl, English version: www.thinklets.nl). In the research project, in which over 1000 primary school students participated, students were followed from the end of Grade 1 (Dutch *groep 3*) to the end of Grade 4 (Dutch *groep 6*). We employed a cluster randomized controlled trial, including three experimental conditions (playing the games at school, playing the games at home, and playing the games at home with afterwards debriefing at school) and a control condition.

Apart from the effectiveness of the mini-games in enhancing students' multiplicative reasoning ability, in our research project we also examined students' starting knowledge of multiplicative reasoning, just before they start receiving formal instruction on this domain. Another focus in the project was on the development of students' attitude towards mathematics, and the influence of playing mathematics games on this attitude. The different studies carried out in the course of our longitudinal research project were reported in the previous five chapters.

In the following, I summarize the research findings of the studies reported in this thesis. After that, I provide some practical implications for mathematics education, and I suggest some directions for further research. I end with the main conclusion of our research project.

1 Summary of research findings

1.1 Students' starting knowledge of multiplicative reasoning

In Chapter 2 we described our study on the starting point of students' formal multiplicative reasoning, that is, the students' knowledge base available just before formal instruction on multiplicative reasoning commences. This knowledge, which we referred to as "pre-instructional knowledge", is important to lay bare, as it can be used by teachers to build on when formal instruction starts. Our analysis including 1176 students from 53 Grade 1 classes revealed that, at the end of Grade 1, children already have a considerable amount of knowledge on multiplicative reasoning. Specifically, we found that children, on average, correctly solved 58% of the multiplicative problems presented to them. As we noted, this pre-instructional knowledge can be seen as a form of informal knowledge, that is, knowledge gained through everyday experiences rather than through formal instruction. Our study showed that children can even display this knowledge when assessed in a relatively formal test setting (without the help of an interviewer or teacher and without the use of physical materials). Many students could even solve some bare number problems, presented in the form of a doubling problem or with the × symbol replaced by the word *times*.

When we examined children's performance on different types of problems, we found that problems which include a picture with countable objects were easiest to solve. In addition, the semantic structure of multiplicative problems influenced their difficulty level, with equal groups problems (e.g., 3 boxes with 4 cookies each) being easiest. No clear difference in difficulty was found between multiplication and division problems, which is in line with earlier research findings indicating that, before formal instruction, children intuitively link these two operations and use the same strategies for both.

Finally, we looked at the influence of student and class characteristics. We found no difference in knowledge of multiplicative reasoning between boys and girls, but students with higher-educated parents displayed more multiplicative knowledge than did students with lower-educated parents. Furthermore, students' performance in multiplicative reasoning was found to be related to the mathematics textbook used in class.

1.2 Effects of mini-games on students' multiplicative reasoning ability

The process of learning from games can be thought of as similar to the process through which the abovementioned pre-instructional, or informal, knowledge was acquired, as students can learn from their experiences in playing games. The specific mini-games used in our study encouraged the use of informal, context-related procedures to solve multiplicative problems. In this way, the learning from mini-games in the beginning of formal instruction of multiplicative reasoning may nicely connect to children's informal knowledge and to the way children are used to learn before formal instruction in this domain starts.

Chapters 3, 4, and 5 focused on the effectiveness of multiplicative mini-games in enhancing students' multiplicative reasoning ability. For examining this effectiveness, we used a cluster randomized controlled trial, in which the participating schools were randomly distributed over four conditions:

E1 playing multiplicative mini-games at school, integrated in a lesson

E2 playing multiplicative mini-games at home, without attention at school

E3 playing multiplicative mini-games at home, with debriefing at school

C playing at school mini-games on other mathematics domains (control group)

In all conditions, the teachers were requested to keep the total in-class lesson time spent on each mathematics domain the same as would have been the case if the school had not been participating in the study. In this way, we could compare a multiplicative reasoning curriculum including mini-games (in the E conditions) with the regular multiplicative reasoning curriculum without these games (in C). The pseudo-intervention in the control group was included to prevent the effect of the mini-games interventions to be obscured by a possible positive effect of the mere participation in a research project (Hawthorne effect). In each condition, there were four 10-week game periods, two in Grade 2 and two in Grade 3. Students' development of multiplicative reasoning ability was measured using multiplicative ability tests administered at the end of Grade 1, Grade 2, and Grade 3. To measure the effects of the games as accurately as possible, in each of the three studies on the effects of the mini-games interventions we included only those schools/classes in which more than half of the intervention was carried out.

In Chapter 3 we reported a preliminary analysis on the effects of the games in the first year of our research project, that is, in Grade 2. We examined students' improvement on a general multiplicative ability test, which assessed a combination of multiplicative operation skills (procedural knowledge) and insight in multiplicative concepts and number relations (conceptual knowledge). Included in the analysis were 1005 students from 46 schools. When all three experimental conditions together were compared to the control group, we did not find a significant effect of the mini-games on students' learning gains. When testing each experimental condition separately, we found a marginally significant effect for the E3 condition, where the games were played at home and debriefed at school ($d = 0.23$). For playing at school (E1) and playing at home without attention at school (E2) no effect was found.

In Chapter 4 we reported on the effectiveness of the mini-games interventions in both Grade 2 and Grade 3. Here, we examined the effects of the interventions on three different aspects of multiplicative reasoning ability: multiplicative number fact *knowledge* (declarative knowledge), multiplicative operation *skills* (procedural knowledge), and *insight* in multiplicative concepts and number relations (conceptual knowledge). Included were 719 students from 35 schools (the smaller sample as compared to the analysis reported in Chapter 3 was caused by some schools dropping out of the research project and by some teachers having performed less than half of the intervention in Grade 3). In line with our findings in Chapter 3, we found that the mini-games were most effective when they were played at home and afterwards debriefed at school. When deployed in this way, the games positively affected students' multiplicative skills as well as their insight, as compared to the control group (*d*s ranging from 0.22 to 0.29). Also when the games were played at school, integrated in a lesson, they were found to be effective, but only for enhancing insight, and only in Grade 2 ($d = 0.35$). The games were not effective when they were played at home without attention at school.

The advantage of playing at home with debriefing at school (E3) can be explained by this intervention having the combined benefit of playing at home (extended learning time, more learner control) and playing at school (debriefing sessions). The extended time spent on multiplicative reasoning, and the larger amount of freedom students have in exploring in the games when they play at home, may only be effective when the experiences in the game are reflected upon in debriefing sessions at school. Through such debriefing sessions the student can generalize what they have learned in the games, such that it can also be applied outside the games. Another possible role of the in-class debriefing sessions in the E3 intervention is encouraging students to play the games at home (which was indeed more often done in the E3 condition than in the E2 condition).

In Chapter 4 we also examined the role of students' gameplay behavior (time and effort spent on the games), and their gender and prior mathematics ability, in the effectiveness of the games. We found that students' gameplay behavior was often related to learning outcomes, which further indicates the potential of the games in learning multiplicative reasoning. Regarding gender, we found that in Grade 2 the games were more effective for boys than for girls, whereas this gender difference disappeared, and occasionally reversed, in Grade 3. Finally, for the E2 intervention, in which the games were played at home without attention at school, an influence of prior mathematics ability was found: it turned out that students with above-average mathematics ability did profit from this intervention. Apparently, these students did not need debriefing sessions to learn from the games.

In Chapter 5, we examined the effects of the mini-games in special primary education. Here, we studied the effectiveness of a one-year intervention, in the special education equivalent of Grade 2. We started with the same four conditions as we did with the regular education schools, but it turned out that in the home-playing conditions (E2 and E3) the special education teachers did not manage to carry out the intervention as intended. Possibly, such a more independent learning method lies too far from the normal educational practice in special education: teachers may not be used to let students work at home independently and, thus, are less inclined to do so. In our special education study, then, we only investigated the effectiveness of playing the mini-games at school, integrated in a lesson (E1). The study included 81 students from 5 schools for special primary education. We found that the mini-games intervention was effective in enhancing students' multiplication number fact knowledge (declarative knowledge) as compared to the control group ($d = 0.39$). On a combined test of skills (procedural knowledge) and insight (conceptual knowledge), we did not find a difference between the experimental and the control group. For skills and insight, then, the inclusion of mini-games in the multiplicative reasoning curriculum did not have an added value as compared to the regular curriculum without these games, but can still be considered equally effective as the regular approach.

1.3 Students' attitude towards mathematics and the influence of playing mini-games

Chapter 6 addressed students' attitude towards mathematics, which we conceptualized as students' liking, or enjoyment, of the subject of mathematics. In this analysis, 932 students from 45 regular primary schools were included. Students' attitude towards mathematics, and their general mathematics achievement, were measured each half year from end Grade 1 to end Grade 3, and once again at the end of Grade 4.

First of all, we found that students in Grade 1 generally have a moderately positive attitude towards mathematics. However, this attitude was found to gradually decrease in the later grades. This decreasing pattern is in line with what is reported in many earlier studies. We also found a decline for attitude towards reading and towards school in general, but attitude towards mathematics had the largest decrease.

Secondly, we investigated the relation of mathematics attitude with gender. In contrast to many earlier research findings, in our study girls turned out to have a more positive attitude towards mathematics than boys had. The decrease of mathematics attitude over time was the same for boys and girls, which means that the difference between the genders did not change over time.

A third focus was on the relationship between mathematics attitude and mathematics achievement. As expected, we found significant correlations between mathematics attitude and achievement at almost each time point. Furthermore, we found that, averaged over all time points, mathematics achievement was a marginally significant predictor of later mathematics attitude, whereas mathematics attitude did not predict later mathematics achievement.

Finally, we examined the influence of playing the mathematics mini-games on students' mathematics attitude, by investigating gameplay behavior (time and effort spent on the games) as a predictor of later mathematics attitude. Averaged over the four game periods and the three experimental conditions in our study, we found a significant influence of gameplay behavior, be it very small. This finding indicates that, apart from having a positive learning effect, playing mathematics mini-games also has potential in fostering students' attitude towards mathematics.

2 Practical implications for mathematics education

From our research findings we can draw some recommendations for the practice of mathematics education in primary school, especially in the domain of multiplicative reasoning.

Build on available knowledge. When multiplication and division are formally introduced, students already have available a considerable amount of knowledge of multiplicative reasoning. Teachers should be aware of this knowledge, such that they can connect the formal multiplication and division to children's informal procedures and understanding. To be able to optimally build the teaching of formal multiplicative reasoning on the knowledge students bring with them, teachers should assess students' multiplicative knowledge before the start of formal instruction in this domain. Our research suggests that it is possible to assess this knowledge through the use of a computer-based test.

Use mini-games for enhancing multiplicative ability. Our results show that mini-games can effectively be employed in mathematics education to enhance students' multiplicative reasoning ability, both in regular and in special primary education. In regular education, the games can best be played at home and afterwards debriefed at school. Yet, also when the games are played at school, integrated in a lesson, they can foster students' multiplicative insight. Students with high prior mathematics ability can also profit from playing the mini-games at home without attention at school. In special primary education, playing multiplicative mini-games at school can promote students' multiplicative number fact knowledge. An intervention in which the games are played at home appeared less feasible in special primary schools (at least for the schools in our study).

Try to maintain students' initial positive attitude towards mathematics. As we found in our study, in the beginning of primary school, students have a positive attitude towards mathematics. This positive disposition should be nourished, such that it does not get lost. Having a positive attitude towards mathematics is important, because attitude towards mathematics is related to mathematics achievement. Moreover, as we discussed in Chapter 6, students' mathematics attitude can influence students' later mathematics course selection and their educational attainment, and it is important in its own right, as it contributes to students' well-being. Unfortunately, consistent with earlier research, we found that students' attitude towards mathematics decreased over the years. This is a serious issue. Teachers should strive to diminish this decline and attempt to maintain students' initial positive attitude towards mathematics. Our results suggested that playing mathematics games can play a role here.

3 Suggestions for further research

During the BRXXX project several new research questions emerged. A first question concerns which exact aspects of the mini-games interventions contributed to their effectiveness. We found that, in regular primary education, the mini-games were most effective in the E3 condition, in which the games were played at home and debriefed at school. However, the large-scale nature of the project did not make it possible to investigate in detail what characteristics of the debriefing sessions and the games, and the interactions between both, played a crucial role in bringing about the found effects. For example, an intriguing question that could not be answered in the current study was on the precise function of the debriefing sessions in the E3 intervention, which may be a combination of stimulating reflection and encouraging students to play the games at home. Moreover, one may wonder how such debriefing sessions can be effective even when not all students played the games, as was the case in our study. Further research should clarify this.

A second direction for further study relates to the somewhat disappointing results for special education students, where the games only had added value for promoting number fact knowledge, but not for fostering multiplicative skills or insight. Possibly for special education students more cues in the games are needed to come to grasp the concepts and relations embedded in the games. Future research should also investigate the specific requirements for classroom discussions in special education, such that the learning opportunities in the games can optimally be realized.

Finally, given the abovementioned importance of students' attitude towards mathematics, it is worthwhile to seek for interventions to slow down the decrease in this attitude over the years. Our findings indicate that playing mathematics computer games may be helpful, but studies comparing game groups with non-game groups are needed to further examine this possibility.

4 Conclusion

The BRXXX research project was set up to experimentally investigate the effectiveness of deploying multiplicative mini-games in primary school mathematics education. From our research findings, we conclude that multiplicative mini-games can effectively be employed both in regular and in special primary education. In regular primary education, the games were found to be most effective when they were played at home and afterwards debriefed at school. When employed in this way, mini-games can contribute to the regular multiplicative reasoning curriculum in promoting students' multiplicative operation skills as well as their insight in multiplicative concepts and number relations. Also when the games were played at school, integrated in a lesson, they were found to have added value as compared to the regular multiplicative reasoning without these games, but only for enhancing insight. In special primary education, we found that playing multiplicative mini-games at school can support students' knowledge of multiplicative number facts.

In the course of our research project, our experience was that it is quite difficult to carry out a large-scale longitudinal intervention study in the school practice. It was not easy to find teachers willing to participate in the study, and to motivate teachers in subsequent grades to continue their participation. Yet, I think that our research was well worth the effort. We collected valuable evidence of the effectiveness of mini-games in learning multiplicative reasoning. In this way, we contributed to the still small knowledge base on the effectiveness of computer games in mathematics education. Moreover, as the interventions were carried out by the regular class teachers, our research findings are directly applicable to the practice of primary education.

Though our research specifically focused on the effectiveness of mini-games in the domain of multiplicative reasoning, I believe that also some more general implications of our findings can be drawn, which may hold for employing computer games in other mathematics domains or in other school subjects in primary education. The most prominent one is that teacher guidance, for example in the form of debriefing sessions, is important in learning from games. Such guidance may be especially helpful for educational games based on experiential learning, where students can learn new concepts and relations through exploring and experimenting in the games. For students with high prior knowledge, teacher guidance seems less necessary. Another implication from our research is that games can be employed to extend the learning time beyond the time that is available during school hours Yet, from our findings, I conclude that such out-of-school learning activities need to be related to in-school activities (e.g., debriefing sessions) to be effective, at least for students with lower prior knowledge.

Appendix:

The BRXXX mini-games

Appendix: The BRXXX mini-games

The games used in the BRXXX project (in the experimental conditions) were mostly adapted versions of multiplicative mini-games selected from the Dutch mathematics games website Rekenweb (www.rekenweb.nl, English version: www.thinklets.nl). Table 1 shows which games were used in each of the four game periods in our intervention. In the following, we provide a screenshot of each game, together with a description of the game and its learning objectives. It should be noted that these short descriptions are just meant to give the reader an idea of each of the games. The descriptions and instructions the teachers in our research project received were more elaborate than what is provided here.

As a service to the schools in the Netherlands that like to use the games, the complete program of BRXXX mini-games for multiplication and division, with the accompanying instruction videos and teacher manuals, is available at the BRXXX website: www.fisme.science.uu.nl/briks

Table 1
Games per game period

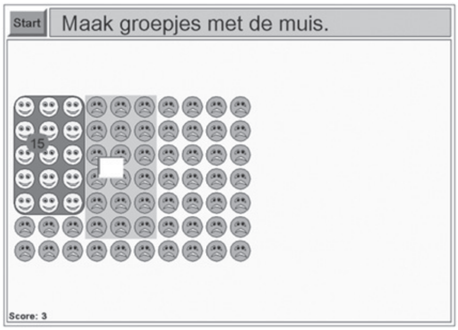
Game period 1 (fall Grade 2)	Game period 2 (spring Grade 2)
1-1 Catching	2-1 Choosing money 1
1-2 Making groups 1	2-2 Making groups 2
1-3 Stamps	2-3 Frog
1-4 Easy problem	2-4 Quick problems 2
1-5 Clothesline	2-5 Falling problems 1
1-6 Quick problems 1	2-6 Wall of numbers 1
1-7 Which of three? 1	2-7 Number factory
1-8 Three in a row	2-8 Four in a row
Game period 3 (fall Grade 3)	**Game period 4 (spring Grade 3)**
3-1 Which of three? 2	4-1 Four in a row
3-2 Falling problems 2	4-2 Choosing money 2
3-3 Art floor	4-3 Wall of numbers 2
3-4 Magic book	4-4 Number factory 2
3-5 Money problems	4-5 Frog
3-6 Fair sharing	4-6 Pay the exact amount 2
3-7 Pay the exact amount 1	4-7 Magic book
3-8 Enlargement	4-8 Falling problems 3

1-1 Catching

Game description
The student has to determine the number of ladybirds on the screen. However, as the ladybirds crawl around, they cannot easily be counted. Ladybirds can be caught in jars to make groups of ladybirds that can later be counted in groups.

Learning objectives
- Realizing that it is easier to count a number of objects when you have groups of equal size
- Developing strategies for finding the total number of objects when having equal size groups (e.g., counting in steps)
- Realizing that some group sizes are easier than others (e.g., counting in steps of 5 is easier than counting in steps of 6)

1-2 Making groups 1



Game description
The student has to make a rectangular arrays of faces, and has to determine the number of faces in each array. The entire field has to be completed.

Learning objectives
- Developing strategies for determining the total amount of objects in an array (e.g., counting in steps, doubling, multiplication)
- Realizing that for some arrays it is easier to determine the total amount than for others (e.g., rows of 5 is easier than rows of 6)
- Realizing that differently shaped rectangular arrays can have the same total amount (e.g., $3 \times 4 = 4 \times 3 = 2 \times 6$)
- Realizing that known amounts of earlier-found arrays can help to find the amount in a new array (e.g. by adding one row or by doubling)

1-3 Stamps

Game description
An envelope with a number of equal value stamps is shown. The student has to determine the total value of the stamps. The accompanying multiplication problem (with *times* instead of the × symbol) is shown. The student can request a related number fact for help (“Hulpsom”), which is then shown on the envelope in another color. The student can choose to work with stamps of value 2, 5, and 10.

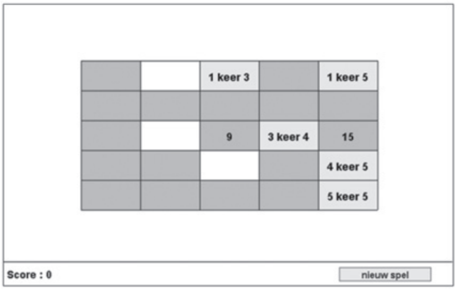
Learning objectives

- Realizing that a number of equal groups can be represented by a bare multiplication problem
- Practicing multiplication problems
- Realizing that related multiplication facts can be used to calculate the answer to new multiplication problems (doubling, using a neighbour problem)
- Realizing that multiplication problems can be calculated in parts (e.g. first adding together the upper 5 stamps and then adding the lower 3 stamps)

1-4 Easy problem

Game description

The student has to solve bare number multiplication problems (with *times* instead of the × symbol), which are structured in an array. In this way, the student can easily use answers to neighbour problems in calculating new multiplications.



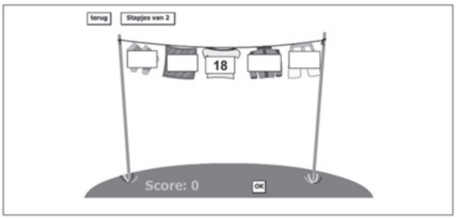
Learning objectives

- Practicing multiplication problems
- Realizing that neighbour problems can be used to calculate new multiplications
- Exploring the pattern of numbers that appears in the game (multiples of numbers)

1-5 Clothesline

Game description

The student has to fill in numbers on all cloths on the clothsline, by counting in steps of 2, 5, or 10.

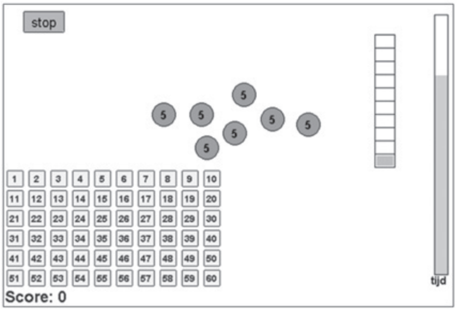


Learning objectives

- Practicing counting in steps
- Realizing the regularity in numbers in the multiplication table of 5 (all numbers end with 5 or 0) and the multiplication table of 10 (all numbers end with 0)

1-6 Quick problems 1

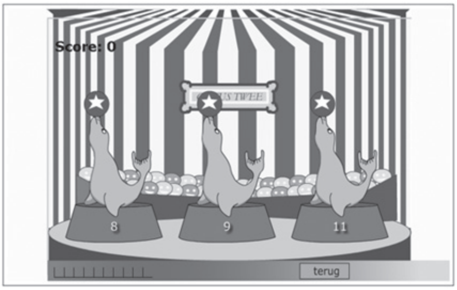
Game description
A number of equal value coins is shown. The student has to determine the total value as quickly as possible. This has to be done ten times, within total time limit of 75 seconds. Coins of 2, 5, and 10 are included.



Learning objectives
- Practicing the quick calculation of multiplication problems presented as a number of equal value coins
- Realizing that strategies can help to quickly calculate multiplication problems (e.g., counting in steps, doubling)

1-7 Which of three? 1

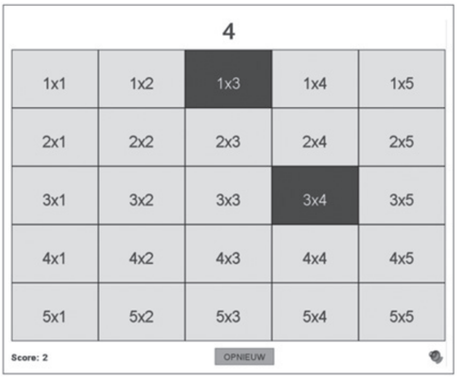
Game description
From three numbers, the student has to select the number that is in a given multiplication table. This game involves the multiplication tables of 2 and 5.



Learning objectives
- Learning to recognize the numbers from the multiplication tables of 2 and 5
- Realizing that numbers in the multiplication table of 2 are all even numbers
- Realizing that numbers in the multiplication table of 5 all end with 5 or 0

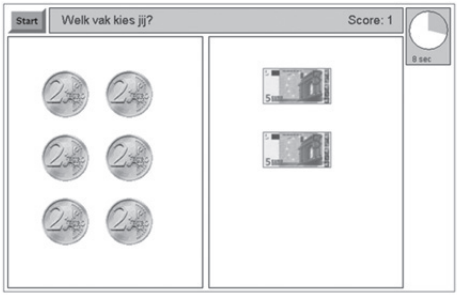
1-8 Three in a row

Game description
A target number is given, and the students has to select a multiplication problem that has this number as the outcome. Successively selected multiplication problems should form a row of three as quickly as possible.



Learning objectives
- Practicing multiplication facts
- Realizing that different multiplication problems can have the same outcome (e.g., $12 = 3 \times 4 = 4 \times 3 = 2 \times 6$)
- Exploring the pattern of multiplication problems in the game (including the multiplication tables up to 5×5)

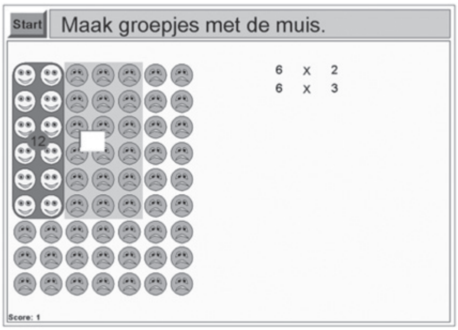
2-1 Choosing money 1



Game description
Two amounts of money are shown, and the student has to quickly choose the amount with the highest value. Each amount is presented as a structured set of multiple coins or banknotes of only one or two types, such that the student is encouraged to use multiplicative relations to solve the problems.

Learning objectives
- Practicing the quick calculation of multiplication problems
- Learning to use structured representations to quickly calculate multiplication problems (e.g., calculating 6 × 2 as 3 × 4)
- Learning to use related number facts to quickly make estimations (e.g., 5 × 2 = 10, so 6 × 2 must be more than 10)

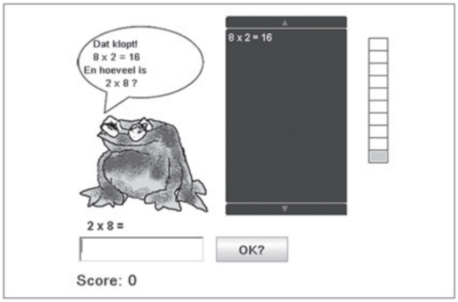
2-2 Making groups 2



Game description
The student has to make rectangular arrays of faces and has to determine the number of faces in each array. The accompanying multiplication problem appears on the right. After the entire field has been completed, the student has to match each multiplication problem with the corresponding rectangular array.

Learning objectives
- Developing strategies for determining the total amount of objects in a rectangular array (e.g., doubling, using a known multiplication fact)
- Laying connections between rectangular arrays and corresponding multiplication problems
- Practicing multiplication problems
- Realizing that differently shaped rectangular arrays can have the same total amount (e.g., 3 × 4 = 4 × 3 = 2 × 6)
- Realizing that known amounts of earlier-found arrays can help to find the amount in a new array (e.g., by adding one row or by doubling)

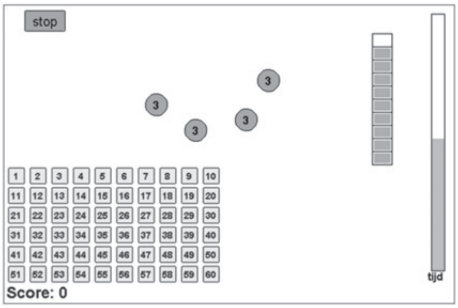
2-3 Frog



Game description
The student is asked to enter a known multiplication problem and its outcome. Then, the frog asks for the outcome of a related multiplication problem, including problems with reversed order of numbers (commutative property), one more/less (distributive property), doubles, halves, and tenfolds (associative property).

Learning objectives
- Practicing recall and calculation of multiplication problems
- Developing insight in how relations between multiplication problems can help in solving new problems, and which relations are useful in which circumstances

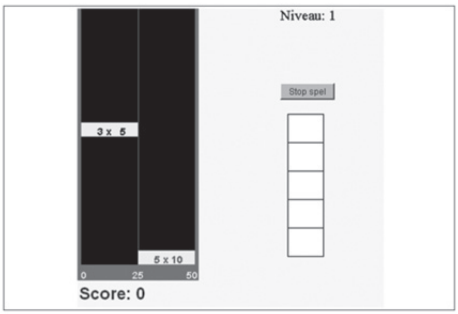
2-4 Quick problems 2



Game description
A number of equal value coins is shown. The student has to determine the total value as quickly as possible. This has to be done ten times, within total time limit of 75 seconds. Coins of 2, 3, 4, 5, and 10 are included.

Learning objectives
- Practicing the quick calculation of multiplication problems presented as a number of equal value coins
- Realizing that strategies can help to quickly calculate multiplication problems (e.g., counting in steps, doubling)

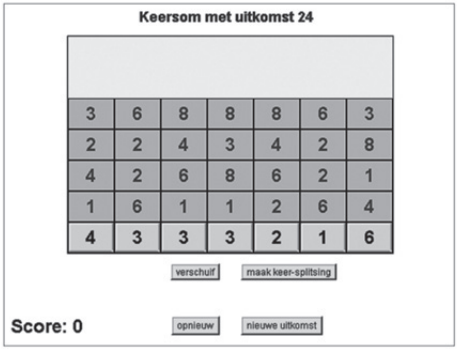
2-5 Falling problems 1



Game description
Multiplication problems are falling down, and the student has to decide whether the outcome is below or above 25 before the problem hits the ground. The falling speed increases during the game.

Learning objectives
- Practicing the quick calculation or estimation of the outcome of multiplication problems
- Realizing that multiplication problems with higher numbers have higher outcomes
- Realizing that relations between multiplication problems can be helpful in quickly calculating or estimating a multiplication outcome (e.g., 5 × 5 = 25, so 3 × 5 should be less than 25)

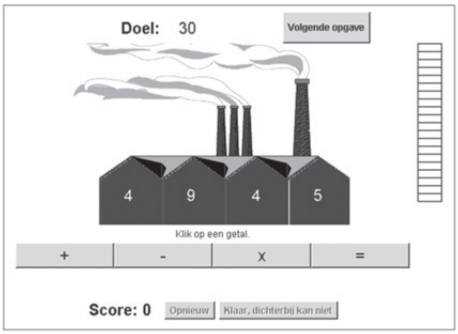
2-6 Wall of numbers 1



Game description
The student has to select two or more numbered bricks in the bottom row of a wall, that together multiply to a given target number (e.g., 24). If correct, the selected bricks disappear and the other bricks fall down. The goal is to demolish the entire wall. The game includes the target numbers 12, 16, 18, 20, 24, and 36.

Learning objectives
- Practicing multiplication problems
- Getting acquainted with multiplication problems with more than two terms
- Learning which different multiplication problems have the same outcome
- Realizing that the order in which multiplication problems are calculated does not matter for the outcome obtained (commutative property, associative property)
- Realizing that when a number is multiplied by 1, it remains the same

2-7 Number factory



Game description
A target number is given. The student has to combine the numbers in the factory, using addition, subtraction, and multiplication, to come as close as possible to the target number. Not all numbers in the factory have to be used. The numbers presented have been chosen in such a way that at least one multiplication is needed to come close to the target number.

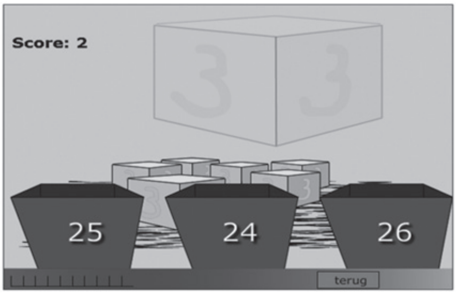
Learning objectives
- Practicing multiplication problems (and addition and subtraction problems)
- Developing insight in using numbers and operations to create particular numbers

2-8 Four in a row

Game description
A target number is given, and the student has to select a multiplication problem that has this number as the outcome. Successively selected multiplication problems should form a row of four as quickly as possible.

Learning objectives
- Practicing multiplication facts
- Developing insight in multiplication problems that have the same outcome (e.g., $4 \times 5 = 5 \times 4 = 10 \times 2$)
- Recognizing the pattern of multiplication problems in the game (including all multiplication tables)

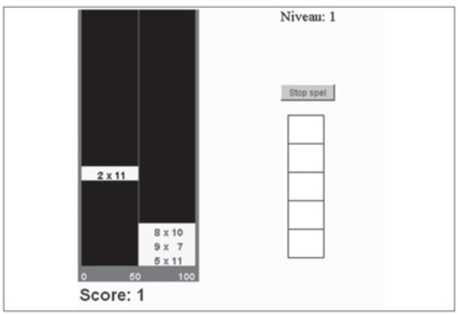
3-1 Which of three? 2



Game description
From three numbers, the student has to select the number that is in a given multiplication table. This game involves the multiplication tables of 2 to 9.

Learning objectives
- Learning to recognize the numbers from the multiplication tables of 2 to 9
- Practicing multiplication tables
- Realizing the regularity in certain multiplication tables (e.g., tables of even numbers only contain even numbers, tables of uneven numbers contain both even and uneven numbers, numbers in the multiplication table of 5 always end with 5 or 0)

3-2 Falling problems 2

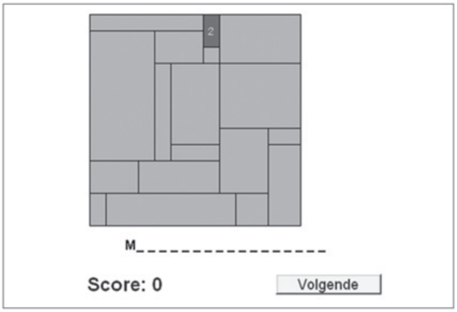


Game description
Multiplication problems are falling down, and the student has to decide whether the outcome is below or above 50 before the problem hits the ground. The falling speed increases during the game.

Learning objectives
- Practicing the quickly calculation or estimation of the outcome of multiplication problems
- Realizing that multiplication problems with higher numbers have higher outcomes
- Practicing the use of strategies in quick calculation or estimation (e.g., one more/less, doubling)

3-3 Art floor

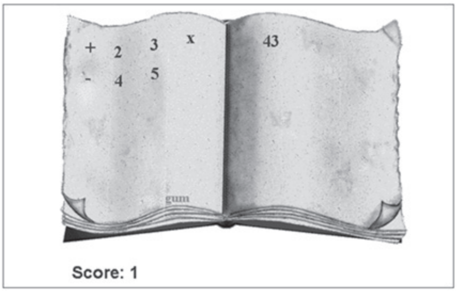
Game description
A floor with differently shaped rectangular tiles is shown. The student has to determine the area of each tile, based on an area that is known. The student first can find a tile that has a similar shape as the known tile, then a tile that is the double of it and so on. When the areas of all tiles have been determined, a word appears.



Learning objectives
- Recognizing how rectangular shapes are composed
- Developing different strategies for finding an area (e.g., using multiples of a known area)
- Practicing the application of multiplication problems

3-4 Magic book

Game description
The student has to combine the four given numbers, using addition, subtraction, and multiplication, to get exactly to the target number. When the target number is obtained, a hidden picture is shown. For each target number, at least one multiplication is needed.



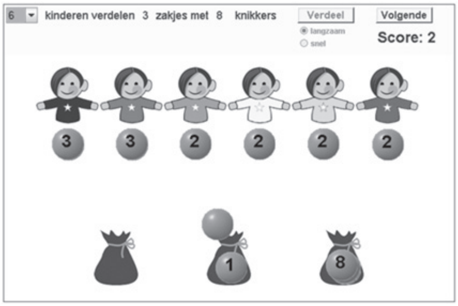
Learning objectives
- Practicing multiplication problems
- Developing insight in using numbers and operations to create particular numbers

3-5 Money problems

Game description
The student has to solve multiplication problems with money amounts beyond €10. If an incorrect answer is given, the student can request a structured representation of the problem with banknotes and coins, which stimulates the use of strategies.

Learning objectives
- Practicing the calculation of multiplication problems above the tables of 1 to 10
- Realizing that multidigit multiplication problems can be calculated in steps: e.g., 3×12 can be solved by first calculating 3×10 and then adding 3×2 (distributive property)

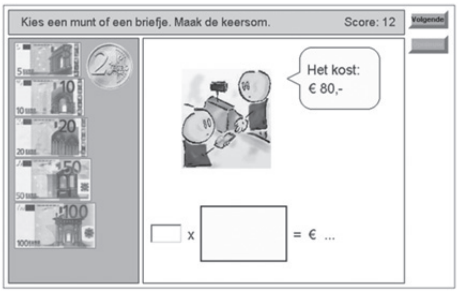
3-6 Fair sharing



Game description
The student has to select, from a drop-down list, a number of children among whom a given number of bags with each the same number of marbles can be evenly divided. After the student has selected a number of children, an animation is shown in which the marbles are divided one by one over the children. This animation shows whether the number of children was correctly choosen.

Learning objectives
- Practicing division
- Realizing that an amount can have multiple divisors
- Developing insight in relations between multiplication problems (e.g., $3 \times 8 = 6 \times 4$)
- Developing insight in divisibility and factors of numbers

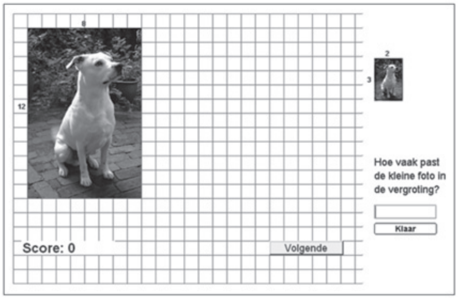
3-7 Pay the exact amount 1



Game description
The student has to pay a certain amount of money. Only one type of coin or banknote can be used. The student has to select a coin or banknote and indicate how many of these are needed. If there are other possible solutions, the “Opnieuw” (*again*) button is highlighted and the student can give another solution to the same problem, using another coin or banknote.

Learning objectives
- Practicing multiplication and division (above the tables of 1 to 10)
- Realizing that an amount can have multiple divisors
- Developing insight in relations between multiplication problems (e.g., $8 \times 10 = 4 \times 20 = 16 \times 5$)
- Developing insight in divisibility and factors of numbers

3-8 Enlargement



Game description
The student has to determine how many times the small photograph fits into the enlarged photograph. The small photo can be moved over the enlargement to get an idea of their relative sizes. If initially an incorrect answer is given, the students get the possibility to lay multiple copies of the small photo onto the enlargement.

Learning objectives
- Practicing the application of multiplication problems in the context of arrays (e.g., four rows with each four small photos)
- First exploration of calculation area using length times width (e.g., $3 \times 2 = 6$ and $12 \times 8 = 96$; $96 \div 6 = 16$; so 16 small photos in enlarged photo)
- Experiencing the change of area in the context of enlargements

4-1 Four on a row
This game is the same as game 2-8.

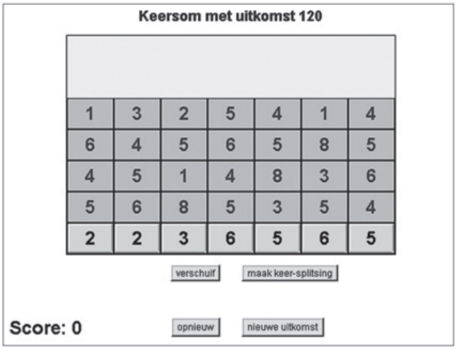
4-2 Choosing money 2

Game description
Two amounts of money are shown, and the student has to quickly choose the amount with the highest value. Each amount is presented as a structured set of multiple coins or banknotes of only one or two types, such that the student is encouraged to use multiplicative relations to solve the problems. In *Choosing money 2*, more difficult sets of money are included than in *Choosing money 1* (game 2-1).

Learning objectives
- Practicing the quick calculation of multiplication problems
- Learning to use structured representations to quickly calculate multiplication problems (e.g., $3 \times 15 = 3 \times 10 + 3 \times 5$)
- Learning to use relations between number facts to quickly make estimations (e.g., $2 \times 10 = 20$ and $10 + (3 \times 5)$ is more than 20, so together this is more than 2×20)

4-3 Wall of numbers 2

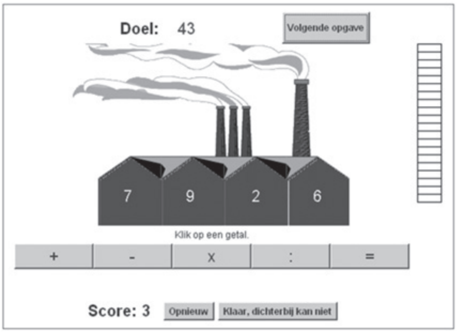
Game description
The student has to select two or more numbered bricks in the bottom row of a wall that together multiply to a given target number (e.g., 120). If correct, the selected bricks disappear and the other bricks fall down. The goal is to demolish the entire wall. The game includes the target numbers 24, 36, 48, 54, 60, 72, and 120.



Learning objectives
- Practicing multiplication problems
- Practicing multiplication problems with more than two terms
- Learning which different multiplication problems have the same outcome
- Realizing that the order in which multiplication problems are calculated does not matter for the outcome obtained (commutative property, associative property)
- Realizing that when a number is multiplied by 1, it remains the same

4-4 Number factory 2

Game description
A target number is given. The student has to combine the numbers in the factory, using addition, subtraction, multiplication, and division, to come as close as possible to the target number. Each time at least one multiplication is needed to come close.



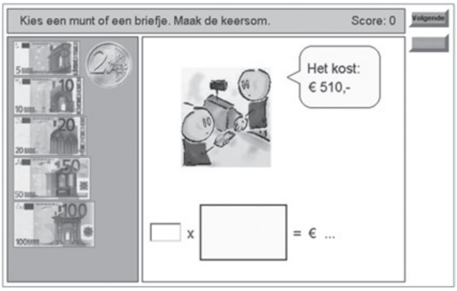
Learning objectives
- Practicing multiplication and division problems (and addition and subtraction problems)
- Developing insight in using numbers and operations to create particular numbers
- Realizing that dividing "is not always possible" (i.e., sometimes the division has a remainder)

4-5 Frog
This game is the same as game 2-3.

4-6 Pay the exact amount 2

Game description

The student has to pay a certain amount of money. Only one type of coin or banknote can be used. The student has to select a coin or banknote and indicate how many of these are needed. If there are other possible solutions, the "Opnieuw" (*again*) button is highlighted and the student can give another solution to the same problem, using another coin or banknote. The amounts to be paid are higher and/or more difficult than in *Pay the exact amount 1* (game 3-7).



Learning objectives

- Practicing multiplication and division above the tables of 1 to 10
- Realizing that an amount can have multiple divisors
- Developing insight in relations between multiplication problems above the tables of 1 to 10 (e.g., $51 \times 10 = 102 \times 5$)
- Developing insight in divisibility and factors of numbers

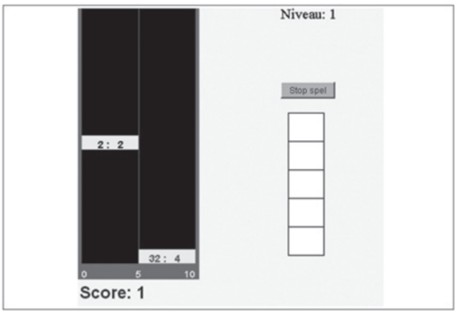
4-7 Magic book

This game is the same as game 3-4.

4-8 Falling problems 3

Game description

Division problems are falling down, and the student has to decide whether the outcome is below or above 5 before the problem hits the ground. The falling speed increases during the game.



Learning objectives

- Practicing the quickly calculation or estimation of the outcome of division problems
- Realizing that division problems with a larger divisor have smaller outcomes, and that division problems with a smaller divisor have larger outcomes
- Realizing that strategies can be used to quickly calculate or estimate division outcomes (e.g., thinking of a multiplication fact, using a neighbour division problem to get close to the outcome)

Samenvatting

Dankwoord

Curriculum vitae

Overview of BRXXX publications

Overview of BRXXX presentations

FIsme Scientific Library

ICO Dissertation Series

Samenvatting

Vermenigvuldigen en delen, ook wel multiplicatieve vaardigheden genoemd, zijn een belangrijk onderdeel van het reken-wiskundecurriculum in het basisonderwijs. Bij het leren van vermenigvuldigen en delen is het van belang om de tafelfeiten te automatiseren, en om vaardigheden te ontwikkelen in het berekenen van vermenigvuldig- en deelopgaven. Naast deze tafelkennis (declaratieve kennis) en rekenvaardigheden (procedurele kennis), moeten leerlingen ook begrip, of inzicht, ontwikkelen in de achterliggende concepten en de getalrelaties bij vermenigvuldigen en delen (conceptuele kennis). Een mogelijke manier om zowel de tafelkennis en rekenvaardigheden als het inzicht van leerlingen te verbeteren is door middel van educatieve computerspelletjes.

Sinds de opkomst van de computerspelletjes is vaak gewezen op het mogelijke nut ervan in het onderwijs, omdat computerspelletjes vaak erg motiverend zijn voor leerlingen. Door deze motiverende werking kan het gebruik van educatieve computerspelletjes ervoor zorgen dat leerlingen meer tijd en aandacht aan het leren besteden, wat kan leiden tot hogere leeruitkomsten. Veel kinderen zijn ook bereid om educatieve spelletjes thuis te spelen, in hun vrije tijd. Dit wijst op de mogelijkheid van het uitbreiden van de leertijd door het aanbieden van educatieve computerspelletjes om thuis te spelen. Buiten de motivationele kenmerken hebben computerspelletjes het voordeel dat ze directe feedback kunnen geven. Als spelletjes worden gebruikt voor het oefenen van tafelfeiten en rekenvaardigheden is deze feedback nuttig omdat leerlingen meteen kunnen zien of hun antwoord goed is. Verder kan directe feedback, in combinatie met de relatief veilige omgeving die een computer biedt, leerlingen aanmoedigen om te exploreren en experimenteren in een reken-computerspel. Door middel van dit zogenoemde ervaringsleren kunnen leerlingen nieuwe rekenwiskundige concepten en strategieën ontdekken, waarmee het inzicht wordt verhoogd. Naast de mogelijkheden van computerspelletjes voor het leren, kan het spelen van computerspelletjes mogelijk ook bijdragen aan een positieve attitude ten opzichte van schoolvakken, in dit geval het vak rekenen-wiskunde.

Ondanks de beloften van computerspelletjes voor het onderwijs is er nog weinig empirisch bewijs voor de effectiviteit van educatieve computerspelletjes. Dit geldt voor educatieve computerspelletjes in het algemeen, en ook specifiek voor computerspelletjes voor rekenen-wiskunde. Eerdere studies naar de effecten van computerspelletjes waren vaak kleinschalig, hadden geen controlegroep, of maakten geen gebruik van random toewijzing aan condities. Auteurs van recente reviewartikelen geven aan dat grootschalige longitudinale studies in de schoolpraktijk nodig zijn. Deze aanbeveling is in lijn met de aandacht die er momenteel is voor *evidence-based education*: educatieve innovaties moeten gebaseerd zijn op grondig empirisch onderzoek naar wat werkt in het onderwijs.

Om dit empirisch bewijs te verkrijgen voor de effectiviteit van computerspelletjes in het reken-wiskundedomein van vermenigvuldigen en delen, is een longitudinaal onderzoeksproject opgezet: het BRXXX-project. Dit project is gerealiseerd binnen het programma OnderwijsBewijs van het Ministerie van Onderwijs, Cultuur en Wetenschap. In het BRXXX-project, waaraan meer dan 1000 basisschoolleerlingen hebben deelgenomen, werden leerlingen gevolgd van eind groep 3 tot eind groep 6. Er werd gewerkt met een speciaal voor dit onderzoek ontwikkeld programma van mini-games (korte, gerichte spelletjes), waarbij gebruik is gemaakt van spelletjes van het Rekenweb (www.rekenweb.nl). De mini-games waren zowel gericht op het oefenen van tafelkennis en vaardigheden in het oplossen van vermenigvuldig- en deelopgaven, als op het verkrijgen van inzicht in multiplicatieve getalrelaties (via ervaringsleren). We onderzochten de effectiviteit van drie verschillende manieren van het inzetten van de mini-games: op school spelen geïntegreerd in een les, thuis spelen, en thuis spelen met nabespreking op school.

Naast de effectiviteit van de computerspelletjes op de multiplicatieve vaardigheden van de leerlingen hebben we in het onderzoeksproject ook gekeken naar wat leerlingen al weten van vermenigvuldigen en delen net voordat ze hier formeel les in krijgen. Een andere aandachtsgebied in het project was de ontwikkeling van de attitude van leerlingen ten opzichte van het vak rekenen-wiskunde, en de relatie met het spelen van reken-computerspelletjes. De verschillende studies die zijn uitgevoerd binnen het BRXXX-project zijn beschreven in hoofdstuk 2 tot 6 van dit proefschrift.

In *hoofdstuk 2* zochten we een antwoord op de vraag: wat weten leerlingen aan het eind van groep 3 al op het gebied van vermenigvuldigen en delen, vlak voor ze formele instructie krijgen in dit domein? Om deze ‘informele kennis’ bloot te leggen, hebben we gekeken naar de prestaties van de leerlingen op de eerste multiplicatieve toets in het onderzoeksproject. In totaal werden 1176 leerlingen van 53 groepen 3 meegenomen in de analyse. De resultaten lieten zien dat de leerlingen al behoorlijk wat multiplicatieve opgaven konden oplossen. Gemiddeld werd 58% van de opgaven goed gemaakt. Zelfs van kale keer-opgaven (kale sommen met het × symbool vervangen door het woord ‘keer’) werd gemiddeld meer dan de helft goed beantwoord. Er is dus al aardig wat informele multiplicatieve kennis aanwezig bij leerlingen in groep 3. Bovendien konden de leerlingen in ons onderzoek deze kennis in een relatief formele toetssituatie laten zien (een online toets zonder gebruik van hulpmaterialen). Leerkrachten kunnen op deze reeds bestaande multiplicatieve kennis voortbouwen bij het aanleren van vermenigvuldigen en delen.

Toen we keken naar de prestaties van de leerlingen op verschillende typen multiplicatieve opgaven, vonden we dat opgaven gemakkelijker op te lossen waren wanneer ze een plaatje bevatten met mogelijkheden om te tellen, en wanneer het om een multiplicatieve situatie met gelijke groepen ging (bijv. 3 dozen van 4 koeken). We vonden geen duidelijk verschil in moeilijkheid tussen vermenigvuldig- en deelopgaven. Dit komt overeen met eerder onderzoek waaruit bleek dat kinderen, voordat ze formele instructie hebben gehad in multiplicatieve vaardigheden, een intuïtieve verbinding leggen tussen vermenigvuldigen en delen en voor beide dezelfde strategieën gebruiken.

We vonden geen verschil in informele multiplicatieve kennis tussen jongens en meisjes. Wel bleek dat leerlingen met hoger opgeleide ouders meer kennis van vermenigvuldigen en delen hadden dan leerlingen met lager opgeleide ouders. Verder waren de leerlingprestaties gerelateerd aan de rekenmethode die in de klas gebruikt werd. Dit kan mogelijk verklaard worden door verschillen tussen de rekenmethodes in de hoeveelheid en het soort informele, voorbereidende multiplicatieve activiteiten die in groep 3 worden aangeboden.

In *hoofdstuk 3, 4, en 5* richtten we ons op de effectiviteit van mini-games bij het verbeteren van de multiplicatieve vaardigheden. Om deze effectiviteit te onderzoeken werden de deelnemende scholen random verdeeld over vier condities:

- *E1* het op school spelen van multiplicatieve mini-games, geïntegreerd in een les
- *E2* het thuis spelen van multiplicatieve mini-games, zonder aandacht op school
- *E3* het thuis spelen van multiplicatieve mini-games, met een nabespreking op school
- *C* controlegroep: het op school spelen van mini-games over andere reken-wiskundeonderwerpen (ruimtelijke orientatie, optellen, aftrekken)

In alle condities werd de leerkrachten gevraagd om de totale lestijd besteed aan elk rekenonderwerp hetzelfde te houden als wanneer ze niet mee zouden doen aan het onderzoek. Op deze manier konden we het reguliere lesprogramma voor multiplicatieve vaardigheden (in de controlegroep) vergelijken met een lesprogramma waar het spelen van mini-games deel van uitmaakte (in de experimentele condities). De pseudo-interventie in de controlegroep was bedoeld om te controleren voor het mogelijke positieve effect dat het deelnemen aan een onderzoeksproject op zichzelf al kan hebben (Hawthorne effect).

In elke conditie waren er vier spelletjesperiodes van tien weken; twee periodes in groep 4 en twee periodes in groep 5. De ontwikkeling van de leerlingen op het gebied van vermenigvuldigen en delen werd gemeten met online multiplicatieve toetsen aan het eind van groep 3, groep 4, en groep 5. Om de effecten van de mini-games zo nauwkeurig mogelijk te kunnen meten, hebben we in elk van de drie studies naar de effecten van de spelletjes alleen die scholen/klassen meegenomen waar tenminste de helft van de spelletjes was behandeld.

Hoofdstuk 3 handelt over de effecten van de spelletjes in groep 4. We onderzochten hier de leerwinst van de leerlingen op een multiplicatieve toets, gericht op vaardigheden in het berekenen van vermenigvuldig- en deelopgaven (procedurele kennis), en inzicht in multiplicatieve getalrelaties (conceptuele kennis). In de analyse werden 1005 leerlingen van 46 scholen meegenomen. Regressie-analyses lieten zien dat, gemiddeld over de drie experimentele groepen samen, de mini-games interventie geen effect had op de leerwinst van de leerlingen. Toen we voor elk experimentele conditie apart het effect bekeken, vonden we een marginaal significant effect van de E3 interventie (p = .07, d = 0.23); in de E1 en E2 conditie vonden we geen effect. De resultaten lijken er dus op te wijzen dat het thuis spelen met een nabespreking op school de meeste potentie heeft.

In hoofdstuk 4 onderzochten we de effecten van de mini-games interventies in zowel groep 4 als groep 5. Hier keken we naar effecten op de drie verschillende aspecten van multiplicatieve vaardigheden: tafelkennis, vaardigheden in het berekenen van vermenigvuldig- en deelopgaven, en inzicht in multiplicatieve getalrelaties. In de analyse werden 719 leerlingen van 35 scholen meegenomen (het kleinere aantal leerlingen/scholen in vergelijking met de studie in hoofdstuk 3 werd veroorzaakt door de uitval van scholen en het niet voldoende uitvoeren van de groep 5 interventie door sommige leerkrachten). Met pad-analyses onderzochten we het effect van de interventies in groep 4 en groep 5. In lijn met onze bevinding in hoofdstuk 3 vonden we dat de mini-games het meest effectief waren wanneer ze thuis werden gespeeld en op school werden nabesproken (E3). Wanneer de spelletjes op deze manier werden ingezet, hadden ze, in vergelijking met de controlegroep, een positief effect op zowel de rekenvaardigheden in het vermenigvuldigen en delen, als het inzicht in multiplicatieve getalrelaties (d effectgroottes van 0.22 tot 0.29). Ook wanneer de spelletjes op school werden gespeeld, geïntegreerd in een les (E1), waren ze effectief, maar alleen voor het bevorderen van inzicht, en alleen in groep 4 (d = 0.35). Voor de E2 interventie vonden we geen effect.

Het feit dat het thuis spelen van de spelletjes met een nabespreking op school het meest effect had, kan verklaard worden doordat deze manier van het inzetten van de spelletjes het gecombineerde voordeel heeft van thuis spelen (extra leertijd, meer ‘learner control’) en op school spelen (nabespreking). De extra tijd besteed aan vermenigvuldigen en delen, en de grotere mate van vrijheid die leerlingen hebben in het exploreren in de spelletjes wanneer ze ze thuis spelen, zijn mogelijk alleen effectief wanneer op de ervaringen uit de spelletjes wordt gereflecteerd in nabesprekingen op school. Door zulke nabesprekingen kunnen leerlingen generaliseren wat ze in de spelletjes hebben geleerd, zodat het geleerde ook buiten de spelletjes kan worden toegepast. Een andere mogelijke rol van de nabesprekingen in de E3 interventie is het aanmoedigen van leerlingen om de spelletjes thuis te spelen (dit werd inderdaad meer gedaan in de E3 conditie dan in de E2 conditie).

In hoofdstuk 4 hebben we ook gekeken naar de rol van het speelgedrag van leerlingen in de spelletjes, oftewel, de tijd en aandacht die ze aan de spelletjes besteed hebben. Deze gegevens werden bijgehouden via inlogaccounts voor elke leerling. Het bleek dat de mate waarin de leerlingen met de spelletjes hebben gespeeld soms, maar soms ook niet, gerelateerd was aan de leeruitkomsten van de leerlingen. Verder hebben we gekeken naar de rol van geslacht en rekenniveau bij de effectiviteit van de spelletjes. De resultaten lieten zien dat in groep 4 de spelletjes effectiever waren voor jongens dan voor meisjes. In groep 5 verdween dit verschil. Verder vonden we voor de E2 interventie (thuis spelen zonder aandacht op school) een invloed van het rekenniveau. Al had deze interventie geen significant effect voor alle leerlingen samen, voor leerlingen met een bovengemiddeld rekenniveau was deze interventie wel effectief. Blijkbaar hadden deze leerlingen geen nabesprekingen nodig om van de spelletjes te leren.

In hoofdstuk 5 bekeken we de effecten van de mini-games in het speciaal basisonderwijs (SBO). Hier hebben we het effect van een één-jarige interventie onderzocht, in het SBO-equivalent van groep 4. We zijn gestart met dezelfde vier condities als in het reguliere basisonderwijs, maar het bleek dat in de thuisspeel-condities (E2 en E3) de SBO-leerkrachten de interventie voor minder dan de helft hebben uitgevoerd (mogelijk zijn leerkrachten in het SBO niet gewend om leerlingen zelfstandig thuis te laten werken). Daarom konden we in onze SBO studie alleen kijken naar het effect van het op school spelen van de mini-games, geïntegreerd in een les (E1). De studie bevatte 81 leerlingen van 5 scholen. We vonden dat de mini-games interventie effectief was in het verbeteren van de tafelkennis van de leerlingen (declaratieve kennis), in vergelijking met de controlegroep ($d = 0.39$). Op een toets van vaardigheden in het berekenen van vermenigvuldig- en deelopgaven en inzicht in multiplicatieve getalrelaties (procedurele en conceptuele kennis) vonden we geen verschil met de controlegroep. Voor rekenvaardigheden en inzicht kunnen we dus zeggen dat het inzetten van de mini-games als onderdeel van het lesprogramma voor vermenigvuldigen en delen weliswaar geen toegevoegde waarde had ten opzichte van het normale lesprogramma in het SBO, maar nog steeds gezien kan worden als een 'veilige' lesmethode (leeruitkomsten zijn hetzelfde als met het normale programma).

Hoofdstuk 6 beschrijft onze laatste studie, over de ontwikkeling van de attitude ten opzichte van het vak rekenen-wiskunde. We hebben de reken-wiskundeattitude hier geconceptualiseerd als: hoe leuk vinden leerlingen het vak rekenen-wiskunde? Door middel van een vragenlijst werd deze attitude zes keer gemeten: elk halfjaar van eind groep 3 tot eind groep 5, en nog een keer aan het eind van groep 6.

De resultaten lieten zien dat leerlingen in groep 3 redelijk positief zijn over rekenen-wiskunde. Deze attitude bleek echter af te nemen in de latere schooljaren. Dit afnemende patroon is in overeenstemming met wat in eerdere studies werd gevonden. We vonden ook een afname van de attitude ten opzichte van lezen en ten opzichte van school, maar de afname van reken-wiskundeattitude was het sterkst. In tegenstelling tot bevindingen van eerdere studies bleek in onze studie dat meisjes een positievere reken-wiskundeattitude hadden dan jongens. De afname van de attitude over de jaren was hetzelfde voor jongens als voor meisjes.

Verder onderzochten we de relatie tussen reken-wiskundeattitude en reken-wiskundeprestaties. We keken hiervoor naar de scores van de leerlingen op de reken-wiskundetoetsen van het Cito leerlingvolgsysteem. Zoals verwacht vonden we significante correlaties tussen de reken-wiskundeattitude en de reken-wiskundeprestaties op hetzelfde meetmoment. Verder bleek dat, gemiddeld over alle meetmomenten, reken-wiskundeprestatie een marginaal significante voorspeller was van reken-wiskundeattitude op een later meetmoment, terwijl attitude geen voorspeller was van latere prestaties.

Tot slot hebben we gekeken naar de invloed van het speelgedrag in de spelletjes (tijd en aandacht besteed aan de spelletjes) op de latere reken-wiskundeattitude van de leerlingen. Gemiddeld over de vier spelletjesperiodes en de drie experimentele condities vonden we een significante, maar erg kleine, invloed van speelgedrag. Deze bevinding suggereert dat het spelen van de mini-games, naast het gevonden leereffect, ook kan bijdragen aan het bevorderen van de reken-wiskundeattitude van leerlingen. In vervolgonderzoek zou deze mogelijkheid verder moeten worden onderzocht.

Dankwoord

De weg naar dit proefschrift was er een met hobbels, kronkels en doodlopende zijweggetjes. Het eind zou ik niet hebben bereikt zonder de hulp, steun, en het advies van vele mensen.

Allereerst bedank ik Marja van den Heuvel-Panhuizen, mijn promotor. Onze vele discussies waren niet altijd makkelijk, maar ze hebben zeker bijgedragen aan de kwaliteit van het proefschrift. Marja, dankjewel voor je deskundigheid op het gebied van rekenwiskundedidactiek, en voor al je hulp en grote inspanningen, niet alleen op het instituut maar ook 's avonds en in het weekend via Skype.

Een tweede belangrijk persoon bij de totstandkoming van dit proefschrift is Sylvia van Borkulo, postdoc in het BRXXX-project en lange tijd mijn kamergenoot op het instituut. Sylvia, dank je voor de fijne samenwerking, en ook voor je geduld en de fijne gesprekken als het eens wat minder ging.

Then there is Alexander Robitzsch, our statistics guy from Salzburg. Alexander, thank you for your expertise, your critical view and your humor. I hope we will have a nice dinner at Mr. Jack's once again.

Hanneke Loomans was onze leerkracht-onderzoeker. Dankjewel Hanneke, voor je nieuwsgierigheid, je frisse blik, en de gezellige thee-momenten op woensdagmiddag. Ebitha Antuma en Peter Kooy, jullie hebben helaas niet zo lang aan het project kunnen meewerken, maar jullie enthousiasme was er niet minder om, dank jullie wel. Mirte van Galen en Joyce Hagens, ook jullie bedankt voor je bijdrage aan dit project. En niet te vergeten Peter Boon en Wim van Velthoven, dank voor jullie hulp bij het beschikbaar maken van de spelletjes en de toetsen in de DWO.

Heel belangrijk in het BRXXX-project waren natuurlijk de deelnemende scholen, leerkrachten, en leerlingen. Ik dank de meer dan 100 leerkrachten en meer dan 1000 leerlingen van harte voor hun medewerking en enthousiasme. Zonder jullie bijdrage was dit onderzoek niet mogelijk geweest.

Naast de mensen die direct meewerkten aan het onderzoek, wil ik ook andere collega's bedanken. Marjolijn, bij jou kon ik altijd even langs met een vraag of voor een praatje. Ada en Miranda, jullie waren samen met Sylvia mijn wandelmaatjes in de lunchpauzes. Heel fijn dat jullie mij af en toe even wegtrokken van mijn computer! En Nathalie, dankjewel voor je hulp bij het drukklaar maken van het proefschrift. Ook gaat mijn dank uit naar de ontwikkelaars van de spelletjes van Rekenweb die (in aangepaste versie) in het BRXXX-project zijn gebruikt. Bedankt Vincent, Frans en anderen. Verder bedank ik alle andere FI-collega's die interesse toonden in het BRXXX-onderzoek en mij af en toe een steuntje in de rug gaven.

Ook in mijn privéleven zijn veel mensen belangrijk voor me geweest tijdens de lange weg naar dit proefschrift. Jan Henk, mijn beste vriendje, dankjewel. Voor alles kon ik bij jou terecht, van leuke uitjes, lange gesprekken, en vakanties, tot weekendenlang bij jou aan mijn proefschrift werken terwijl jij voor me kookte en teksten nalas. Het kon niet anders dan dat jij mijn paranimf zou zijn. Leticia, na het afronden van het boek kon ik je eindelijk bezoeken in Berlijn, voor een revival van onze Utrechtse eet-en-klets afspraken. Wat leuk dat ook jij mijn paranimf wil zijn! Mijn lieve zusjes Jikke en Saskia, jullie zijn de beste! Cathelijne, dank je voor alle fijne avonden en muzikale afleiding. Rob, Alexander, Danni, Jacqueline, Susanne en alle anderen, dank voor jullie steun en de vele gezellige momenten. Tot slot, mijn ouders Niek en Mariëtte: jullie zijn er altijd, dank jullie wel.

Curriculum vitae

Marjoke Bakker was born on January 27, 1982, in Deventer (the Netherlands). After completing here secondary schooling in 2000 at the Lorentz Casimir Lyceum in Eindhoven, she started her university studies in Computer Science at the Eindhoven University of Technology. She obtained a bachelor's degree, with honours, in 2004. Then, after a year of studying Linguistics at the Radboud University Nijmegen, she decided to pursue a master's degree in Psychology, at the University of Twente. She graduated with honours in 2008. In her master's program, Marjoke specialized in the psychology of knowledge and learning. Her master thesis focused on the effectiveness of a game-like learning material in enhancing primary school students' spatial ability. After a short employment as a junior researcher at the University of Twente, Marjoke started her PhD research at the Freudenthal Institute in September 2009. The PhD research was carried out under the supervision of Prof. dr. Marja van den Heuvel-Panhuizen, and was performed within the BRXXX research project, funded by the OnderwijsBewijs program of the Dutch Ministry of Education. In the course of the BRXXX project, Marjoke presented the research findings at several national and international conferences, and published in national and international journals. Marjoke was accepted as a PhD member of the ICO research school.

Overview of BRXXX publications
(up to February 2014)

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2014). First-graders' knowledge of multiplicative reasoning before formal instruction in this domain. *Contemporary Educational Psychology, 39,* 59-73. doi: 10.1016/j.cedpsych.2013.11.001

Bakker, M., Van den Heuvel-Panhuizen, M., Van Borkulo, S., & Robitzsch, A. (2013). Effecten van online mini-games op multiplicatieve vaardigheden van leerlingen in groep 4 [Effects of online mini-games on second-graders' multiplicative reasoning ability]. *Pedagogische Studiën, 90*(3), 21-36.

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2013). What children know about multiplicative reasoning before being taught. In Lindmeier, A. M. & Heinze, A. (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education, Vol. 2* (pp. 49-56). Kiel, Germany: PME.

Bakker, M., Van den Heuvel-Panhuizen, M., & Loomans, H. (2013). Voortbouwen op informele kennis: Een goed rekenbegin [Building on informal knowledge: A good math start]. *Jeugd in School en Wereld (JSW), 98*(4), 6-9.

Loomans, H., & Van den Heuvel-Panhuizen, M. (2013). Voorbereiding tafels: Leren met een computerspel [Preparation multiplication tables: Learning with a computer game]. *Jeugd in School en Wereld (JSW), 98*(3), 40-43.

Bakker, M., Van den Heuvel-Panhuizen, M., Van Borkulo, S., & Robitzsch, A. (2012). Effects of mini-games for enhancing multiplicative abilities: A first exploration. In S. De Wannemacker, S. Vandercruysse & G. Clarebout (Vol. Eds.), *Communications in computer and information science: Vol. 280. Serious games: The challenge* (pp. 53-57). Berlin: Springer. doi: 10.1007/978-3-642-33814-4_7

Van Borkulo, S., Van den Heuvel-Panhuizen, M., Bakker, M., & Loomans, H. (2012). One mini-game is not like the other: Different opportunities to learn multiplication tables. In S. De Wannemacker, S. Vandercruysse & G. Clarebout (Vol. Eds.), *Communications in computer and information science: Vol. 280. Serious games: The challenge* (pp. 61-64). Berlin: Springer. doi: 10.1007/978-3-642-33814-4_9

Other articles have been submitted or are in preparation.

The complete BRXXX program with mini-games for multiplication and division, with accompanying instruction videos and teacher manuals, is available at the BRXXX website: www.fisme.science.uu.nl/briks

Overview of BRXXX presentations
(up to February 2014)

Bakker, M., & Van den Heuvel-Panhuizen, M. (2014, January). *Effecten van Rekenweb-spelletjes bij het leren vermenigvuldigen en delen* [Effects of Rekenweb games in the learning of multiplication and division]. Paper presentation at the 32th Panama Conference, Noordwijkerhout, the Netherlands.

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2013, November). *The effectiveness of mathematics computer games in enhancing primary school students' multiplicative reasoning ability*. Paper presented at ICO National Fall School 2013, Maastricht, the Netherlands. (nominated best paper award)

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2013, July). *What children know about multiplicative reasoning before being taught*. Paper presentation at the 37th Conference of the International Group for the Psychology of Mathematics Education, Kiel, Germany.

Bakker, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2013, March). *Knowledge of multiplicative relations before being taught*. Paper presentation at the Fifth Expert Meeting on Mathematical Thinking and Learning, Walferdange, Luxembourg.

Bakker, M., Van den Heuvel-Panhuizen, M., Robitzsch, A., & Van Borkulo, S. (2012, June). *Effecten van RekenWeb games op multiplicatieve vaardigheden in groep 4* [Effects of RekenWeb games on multiplicative abilities in Grade 2]. Paper presented at Onderwijs Research Dagen, Wageningen, Netherlands.

Van Borkulo, S., & Loomans, H. (2012, March). *Het ene spelletje is het andere niet. Leerzame kenmerken van rekencomputerspelletjes* [One game is not like the other: Learning opportunities of mathematics computer games]. Workshop at Nationale Rekendag, Zeist, the Netherlands.

Van Borkulo, S., Van den Heuvel-Panhuizen, M., Bakker, M., & Loomans, H. (2012, January). *Het ene spelletje is het andere niet. Leerzame kenmerken van rekencomputerspelletjes* [One game is not like the other: Learning opportunities of mathematics computer games]. Paper presentation at the 30th Panama Conference, Noordwijkerhout, the Netherlands.

Bakker, M., Van den Heuvel-Panhuizen, M., Van Borkulo, S., & Robitzsch, A. (2011, October). *Effects of mini-games for enhancing multiplicative abilities: A first exploration.* Poster presented at Serious Games: The Challenge, Gent, Belgium.

Van Borkulo, S., Van den Heuvel-Panhuizen, M., & Bakker, M. (2011, October). *One mini-game is not like the other: Different opportunities to learn multiplication tables.* Poster presented at Serious Games: The Challenge, Gent, Belgium.

Bakker, M., Van den Heuvel-Panhuizen, M., & Van Borkulo, S. (2011, June). *Multiplicatieve vaardigheden in groep 3: Resultaten van een online toets* [Multiplicative reasoning ability in Grade 1: Results of an online test]. Paper presented at Onderwijs Research Dagen, Maastricht, the Netherlands.

Bakker, M. (2011, April). *Spelletjes voor het ontwikkelen van multiplicatieve vaardigheden* [Games for developing multiplicative reasoning ability]. Presentation at VOR-ICT Promovendimiddag, Utrecht, the Netherlands.

Bakker, M., Van den Heuvel-Panhuizen, M., & Van Borkulo, S. (2011, January). *Multiplicatieve vaardigheden in groep 3* [Multiplicative reasoning ability in Grade 1]. Paper presentation at the 29th Panama Conference, Noordwijkerhout, the Netherlands.

Bakker, M., Van den Heuvel-Panhuizen, M., & Van Borkulo, S. (2010, November). *First-graders' performance on multiplicative problems as measured by an online test: A first exploration.* Paper presented at ICO Toogdag 2010, Amsterdam.

More presentations on the BRXXX project are planned.

FIsme Scientific Library

(formerly published as CD-β Scientific Library)

82. Hăng, N. V. T. (2014). *Design of a social constructivism-based curriculum for primary science education in Confucian heritage culture.*
81. Sun, L. (2014). *From rhetoric to practice: enhancing environmental literacy of pupils in China.*
80. Mazereeuw, M. (2013). *The functionality of biological knowledge in the workplace. Integrating school and workplace learning about reproduction.*
79. Dierdorp, A. (2013). *Learning correlation and regression within authentic contexts.*
78. Dolfing, R. (2013). *Teachers' Professional Development in Context-based Chemistry Education. Strategies to Support Teachers in Developing Domain-specific Expertise.*
77. Mil, M. H. W. van (2013). *Learning and teaching the molecular basis of life.*
76. Antwi, V. (2013). *Interactive teaching of mechanics in a Ghanaian university context.*
75. Smit, J. (2013). *Scaffolding language in multilingual mathematics classrooms.*
74. Stolk, M. J. (2013). *Empowering chemistry teachers for context-based education. Towards a framework for design and evaluation of a teacher professional development programme in curriculum innovations.*
73. Agung, S. (2013). *Facilitating professional development of* Madrasah *chemistry teachers. Analysis of its establishment in the decentralized educational system of Indonesia.*
72. Wierdsma, M. (2012). *Recontextualising cellular respiration.*
71. Peltenburg, M. (2012). *Mathematical potential of special education students.*
70. Moolenbroek, A. van (2012). *Be aware of behaviour. Learning and teaching behavioural biology in secondary education.*
69. Prins, G. T., Vos, M. A. J. & Pilot, A. (2011). *Leerlingpercepties van onderzoek & ontwerpen in het technasium.*
68. Bokhove, Chr. (2011). *Use of ICT for acquiring, practicing and assessing algebraic expertise.*
67. Boerwinkel, D. J. & Waarlo, A. J. (2011). *Genomics education for decision-making. Proceedings of the second invitational workshop on genomics education, 2-3 December 2010.*
66. Kolovou, A. (2011). *Mathematical problem solving in primary school.*
65. Meijer, M. R. (2011). *Macro-meso-micro thinking with structure-property relations for chemistry. An explorative design-based study.*
64. Kortland, J. & Klaassen, C. J. W. M. (2010). *Designing theory-based teaching-learning sequences for science. Proceedings of the symposium in honour of Piet Lijnse at the time of his retirement as professor of Physics Didactics at Utrecht University.*
63. Prins, G. T. (2010).*Teaching and learning of modelling in chemistry education. Authentic practices as contexts for learning.*
62. Boerwinkel, D. J. & Waarlo, A. J. (2010). *Rethinking science curricula in the genomics era. Proceedings of an invitational workshop.*
61. Ormel, B. J. B. (2010). *Het natuurwetenschappelijk modelleren van dynamische systemen. Naar een didactiek voor het voortgezet onderwijs.*

60. Hammann, M., Waarlo, A. J., & Boersma, K. Th. (Eds.) (2010). *The nature of research in biological education: Old and new perspectives on theoretical and methodological issues – A selection of papers presented at the VIIth Conference of European Researchers in Didactics of Biology*.
59. Van Nes, F. (2009). *Young children's spatial structuring ability and emerging number sense*.
58. Engelbarts, M. (2009*). Op weg naar een didactiek voor natuurkunde-experimenten op afstand. Ontwerp en evaluatie van een via internet uitvoerbaar experiment voor leerlingen uit het voortgezet onderwijs*.
57. Buijs, K. (2008). *Leren vermenigvuldigen met meercijferige getallen*.
56. Westra, R. H. V. (2008). *Learning and teaching ecosystem behaviour in secondary education: Systems thinking and modelling in authentic practices*.
55. Hovinga, D. (2007). *Ont-dekken en toe-dekken: Leren over de veelvormige relatie van mensen met natuur in NME-leertrajecten duurzame ontwikkeling*.
54. Westra, A. S. (2006). *A new approach to teaching and learning mechanics*.
53. Van Berkel, B. (2005). *The structure of school chemistry: A quest for conditions for escape*.
52. Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education: The case of water quality*.
51. Doorman, L. M. (2005). *Modelling motion: from trace graphs to instantaneous change*.
50. Bakker, A. (2004). *Design research in statistics education: on symbolizing and computer tools*.
49. Verhoeff, R. P. (2003). *Towards systems thinking in cell biology education*.
48. Drijvers, P. (2003). *Learning algebra in a computer algebra environment. Design research on the understanding of the concept of parameter*.
47. Van den Boer, C. (2003). *Een zoektocht naar verklaringen voor achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs*.
46. Boerwinkel, D. J. (2003). *Het vormfunctieperspectief als leerdoel van natuuronderwijs. Leren kijken door de ontwerpersbril.*
45. Keijzer, R. (2003). *Teaching formal mathematics in primary education. Fraction learning as mathematising process*.
44. Smits, Th. J. M. (2003). *Werken aan kwaliteitsverbetering van leerlingonderzoek: Een studie naar de ontwikkeling en het resultaat van een scholing voor docenten*.
43. Knippels, M. C. P. J. (2002). *Coping with the abstract and complex nature of genetics in biology education – The yo-yo learning and teaching strategy*.
42. Dressler, M. (2002). *Education in Israel on collaborative management of shared water resources.*
41. Van Amerom, B. A. (2002). *Reinvention of early algebra: Developmental research on the transition from arithmetic to algebra*.
40. Van Groenestijn, M. (2002). *A gateway to numeracy. A study of numeracy in adult basic education.*
39. Menne, J. J. M. (2001). *Met sprongen vooruit: een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – een onderwijsexperiment*.
38. De Jong, O., Savelsbergh, E. R. & Alblas, A. (2001). *Teaching for scientific literacy: context, competency, and curriculum.*

37. Kortland, J. (2001). *A problem-posing approach to teaching decision making about the waste issue.*
36. Lijmbach, S., Broens, M., & Hovinga, D. (2000). *Duurzaamheid als leergebied; conceptuele analyse en educatieve uitwerking*.
35. Margadant-van Arcken, M. & Van den Berg, C. (2000). *Natuur in pluralistisch perspectief – Theoretisch kader en voorbeeldlesmateriaal voor het omgaan met een veelheid aan natuurbeelden*.
34. Janssen, F. J. J. M. (1999). *Ontwerpend leren in het biologieonderwijs. Uitgewerkt en beproefd voor immunologie in het voortgezet onderwijs*.
33. De Moor, E. W. A. (1999). *Van vormleer naar realistische meetkunde – Een historisch-didactisch onderzoek van het meetkundeonderwijs aan kinderen van vier tot veertien jaar in Nederland gedurende de negentiende en twintigste eeuw*.
32. Van den Heuvel-Panhuizen, M. & Vermeer, H. J. (1999). *Verschillen tussen meisjes en jongens bij het vak rekenen-wiskunde op de basisschool – Eindrapport MOOJ-onderzoek.*
31. Beeftink, C. (2000). *Met het oog op integratie – Een studie over integratie van leerstof uit de natuurwetenschappelijke vakken in de tweede fase van het voortgezet onderwijs*.
30. Vollebregt, M. J. (1998). *A problem posing approach to teaching an initial particle model*.
29. Klein, A. S. (1998). *Flexibilization of mental arithmeticsstrategies on a different knowledge base – The empty number line in a realistic versus gradual program design*.
28. Genseberger, R. (1997). *Interessegeoriënteerd natuur- en scheikundeonderwijs – Een studie naar onderwijsontwikkeling op de Open Schoolgemeenschap Bijlmer*.
27. Kaper, W. H. (1997). *Thermodynamica leren onderwijzen*.
26. Gravemeijer, K. (1997). *The role of context and models in the development of mathematical strategies and procedures*.
25. Acampo, J. J. C. (1997). *Teaching electrochemical cells – A study on teachers' conceptions and teaching problems in secondary education*.
24. Reygel, P. C. F. (1997). *Het thema 'reproductie' in het schoolvak biologie*.
23. Roebertsen, H. (1996). *Integratie en toepassing van biologische kennis – Ontwikkeling en onderzoek van een curriculum rond het thema 'Lichaamsprocessen en Vergift'*.
22. Lijnse, P. L. & Wubbels, T. (1996). *Over natuurkundedidactiek, curriculumontwikkeling en lerarenopleiding*.
21. Buddingh', J. (1997). *Regulatie en homeostase als onderwijsthema: een biologie-didactisch onderzoek*.
20. Van Hoeve-Brouwer G. M. (1996). *Teaching structures in chemistry – An educational structure for chemical bonding*.
19. Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*.
18. Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity*.
17. De Jong, O., Van Roon, P. H. & De Vos, W. (1995). *Perspectives on research in chemical education.*
16. Van Keulen, H. (1995). *Making sense – Simulation-of-research in organic chemistry education*.
15. Doorman, L. M., Drijvers, P. & Kindt, M. (1994). *De grafische rekenmachine in het wiskundeonderwijs*.

14. Gravemeijer, K. (1994). *Realistic mathematics education*.
13. Lijnse, P. L. (Ed.) (1993). *European research in science education*.
12. Zuidema, J. & Van der Gaag, L. (1993). *De volgende opgave van de computer*.
11. Gravemeijer, K, Van den Heuvel Panhuizen, M., Van Donselaar, G., Ruesink, N., Streefland, L., Vermeulen, W., Te Woerd, E., & Van der Ploeg, D. (1993). *Methoden in het reken-wiskundeonderwijs, een rijke context voor vergelijkend onderzoek*.
10. Van der Valk, A. E. (1992). *Ontwikkeling in Energieonderwijs*.
9. Streefland, L. (Ed.) (1991). *Realistic mathematics education in primary schools*.
8. Van Galen, F., Dolk, M., Feijs, E., & Jonker, V. (1991). *Interactieve video in de nascholing reken-wiskunde.*
7. Elzenga, H. E. (1991). *Kwaliteit van kwantiteit*.
6. Lijnse, P. L., Licht, P., De Vos, W. & Waarlo, A. J. (Eds.) (1990). *Relating macroscopic phenomena to microscopic particles: a central problem in secondary science education*.
5. Van Driel, J. H. (1990). *Betrokken bij evenwicht*.
4. Vogelezang, M. J. (1990). *Een onverdeelbare eenheid*.
3. Wierstra, R. F. A. (1990). *Natuurkunde-onderwijs tussen leefwereld en vakstructuur*.
2. Eijkelhof, H. M. C. (1990). *Radiation and risk in physics education*.
1. Lijnse, P. L. & De Vos, W. (Eds.) (1990). *Didactiek in perspectief*.

ICO Dissertation Series

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, Erasmus University Rotterdam, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Antwerp, University of Ghent, University of Groningen, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University (and formerly Radboud University Nijmegen and Tilburg University).

Recent publications in the ICO Dissertation Series:

281. Cviko, A. (19-12-2013) *Teacher roles and pupil outcomes in technology-rich early literacy learning*. Enschede: University of Twente
280. Kamp, R. J. A. (28-11-2013) *Peer feedback to enhance learning in problem-based tutorial groups*. Maastricht: Maastricht University
279. Lucero, M. L. (21-11-2013) *Considering teacher cognitions in teacher professional development: Studies involving Ecuadorian primary school teachers*. Ghent: Ghent University
278. Dolfing, R. (23-10-2013) *Teachers' professional development in context-based chemistry education. Strategies to support teachers in developing domain-specific expertise*. Utrecht: Utrecht University
277. Popov, V. (8-10-2013) *Scripting intercultural computer-supported collaborative learning in higher education*. Wageningen: Wageningen University
276. Bronkhorst, L. H. (4-10-2013) *Research-based teacher education: Interactions between research and teaching*. Utrecht: Utrecht University
275. Bezdan, E. (04-10-2013) *Graphical overviews in hypertext learning environments: When one size does not fit all*. Heerlen: Open University of the Netherlands
274. Kleijn, R. A. M. de, (27-09-2013) *Master's thesis supervision. Feedback, interpersonal relationships and adaptivity*. Utrecht: Utrecht University
273. Pillen, M. T. (12-09-2013) *Professional identity tensions of beginning teachers*. Eindhoven: Eindhoven University of Technology
272. Meeuwen, L. W. van (06-09-13) *Visual problem solving and self-regulation in training air traffic control*. Heerlen: Open University of the Netherlands
271. Keuvelaar-Van den Bergh, L. (26-06-2013) *Teacher feedback during active learning: The development and evaluation of a professional development programme*. Eindhoven: Eindhoven University of Technology.
270. Hornstra, T. E. (17-06-2013) *Motivational developments in primary school. Group-specific differences in varying learning contexts*. Amsterdam: University of Amsterdam
269. Vandyck, I. J. J. (17-06-2013) *Fostering community development in school-university partnerships*. Amsterdam: VU Universtiy Amsterdam
268. Milliano, I. I. C. M. de (24-05-2013) *Literacy development of low-achieving adolescents. The role of engagement in academic reading and writing*. Amsterdam: University of Amsterdam

267. Taminiau, E. M. C. (24-05-2013) *Advisory models for on-demand learning*. Heerlen: Open University of the Netherlands

266. Azkiyah, S. N. (23-5-2013) *The effects of two interventions – on teaching quality and student outcome*. Groningen: University of Groningen

265. Diggelen, M. R. van (21-05-2013) *Effects of a self-assessment procedure on VET teachers' competencies in coaching students' reflection skills*. Eindhoven: Eindhoven University of Technology

264. Knol, M. H. (19-04-2013) *Improving university lectures with feedback and consultation*. Amsterdam: University of Amsterdam

263. Dekker-Groen, A. (19-04-2013) *Teacher competences for supporting students' reflection. Standards, training, and practice*. Utrecht: Utrecht University

262. Verberg, C. P. M. (18-04-2013) *The characteristics of a negotiated assessment procedure to promote teacher learning*. Leiden: Leiden University

261. Jong, R. J. de (11-04-2013) *Student teachers' practical knowledge, discipline strategies, and the teacher-class relationship*. Leiden: Leiden University

260. Belo, N. A. H. (27-03-2013) *Engaging students in the study of physics*. Leiden: Leiden University

259. Bijker, M. M. (22-03-2013) *Understanding the gap between business school and the workplace: Overconfidence, maximizing benefits, and the curriculum*. Heerlen: Open University of the Netherlands

258. Noroozi, O. (11-01-2013) *Fostering argumentation-based computer-supported collaborative learning in higher education*. Wageningen: Wageningen University

257. Van der Linden, P. W. J. (20-12-2012) *A design-based approach to introducing student teachers in conducting and using research*. Eindhoven: Eindhoven University of Technology

256. Timmermans, A. C. (27-11-2012) *Value added in educational accountability: Possible, fair and useful?* Groningen: University of Groningen

255. De Feijter, J. M. (09-11-2012) *Learning from error to improve patient safety*. Maastricht: Maastricht University

254. Kenbeek, W. K. (31-10-2012) *Back to the drawing board. Creating drawing or text summaries in support of System Dynamics modeling*. Enschede: University of Twente

253. Doppenberg, J. J. (24-10-2012) *Collaborative teacher learning: settings, foci and powerful moments*. Eindhoven: Eindhoven University of Technology

252. Peltenburg, M. C. (24-10-2012) *Mathematical potential of special education students*. Utrecht: Utrecht University

251. Phielix, C. (28-09-2012) *Enhancing collaboration through assessment & reflection*. Utrecht: Utrecht University

250. Van de Pol, J. E. (28-09-2012) *Scaffolding in teacher-student interaction. Exploring, measuring, promoting and evaluating scaffolding*. Amsterdam: University of Amsterdam

249. Vrijnsen-de Corte, M. C. W. *Researching the teacher-researcher. Practice-based research in Dutch professional development schools*. Eindhoven: Eindhoven University of Technology

248. Spanjers, I. A. E. (05-07-2012) *Segmentation of animations: Explaining the effects on the learning process and learning outcomes*. Maastricht: Maastricht University

247. Van Andel, J (22-06-2012) *Demand-driven education. An educational-sociological investigation*. Amsterdam: VU University Amsterdam