UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 150

Closing the Gap

How an Adaptive Behavioral Based Program on a Tablet Can Help Low Performing Children Catch Up in Math: a Randomized Placebo Controlled Study

MARTIN HASSLER HALLSTEDT

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

ISSN 1652-9030
ISBN 978-91-513-0200-3
urn:nbn:se:uu:diva-336694

Dissertation presented at Uppsala University to be publicly examined in (13:026) Sydney Alrutz-salen, Blåsen hus, von Kraemers Allé 1A/1E, Uppsala, Friday, 16 February 2018 at 10:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Nancy Jordan (University of Delaware, USA).

Abstract

Hassler Hallstedt, M. 2018. Closing the Gap. *How an Adaptive Behavioral Based Program on a Tablet Can Help Low Performing Children Catch Up in Math: a Randomized Placebo Controlled Study*. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 150. 88 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0200-3.

Early mathematic skills have a substantial impact on later school achievement. Children with poor school achievement are at risk for adverse consequences later in life. Math competencies also have consequences for the economy at large because societies are becoming increasingly dependent on skill sets including mathematics. Proficiency with basic arithmetic, also known as math fact (i.e., 3+8, 12-3), is considered to be a critical early math skill. Intervention research in mathematics have demonstrated that math fact deficits among students with low math performance can be improved with additional targeted, non- technological interventions (i.e., small-group tutoring).

The aim of the present thesis was to investigate the effect, using a randomized placebo controlled design, of addititional adaptive, behavioral based, math training on a tablet on low performing second graders. The first study (study I), investigated if arithmetic skills could be assessed in a reliable and valid way on tablet. The examination showed that arithmetic scales could be transferred from paper-based tests to tablet with comparable psychometric properties, although not for a pictorial scale, and that separate norms are needed for tablet. Study II demonstrated that training on a tablet, for on average 19 hours across 20 weeks, improved basic arithmetic skills after training in the math conditions compared to control/ placebo conditions. The effects were medium sized at post assessment. There was a fadeout of effects at 6 months follow-up, where small effects were shown, and the effects decreased further at 12 months follow-up. Children with lower non-verbal IQ seemed to gain significantly more at follow- ups than children with higher non-verbal IQ. The study found no additional effects of combining working memory training and math training. Study III, using a machine learning analysis, found that children demonstrating a positive response at 6 months follow-up were characterized by having completed 90 % or more of the math program at the default level, in combination with having a fairly favorable socioeconomic background.

In summation, this work demonstrates how an adaptive behavioral based program on a tablet can help low performing children improve critical early math skills.

Keywords: intervention, mathematics, tablet, fluency, short and long-term effects, educational technology

Martin Hassler Hallstedt, Department of Psychology, Box 1225, Uppsala University, SE-75142 Uppsala, Sweden.

ISSN 1652-9030
ISBN 978-91-513-0200-3
urn:nbn:se:uu:diva-336694 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-336694)

To Karin, Astrid, Gösta, Lill-Birgit
och Stor-Birgit

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Hassler Hallstedt, M., & Ghaderi, A. (in press) Tablets Instead of Paper-Based Tests for Young Children? Comparability between Paper and Tablet Versions of the Mathematical Heidelberger Rechen Test 1-4. *Educational Assessment.*

II Hassler Hallstedt, M., Klingberg, K., & Ghaderi, A. (in press) Short and Long-Term Effects of a Mathematics Tablet Intervention for Low Performing Second Graders. *Journal of Educational Psychology*.

III Hassler Hallstedt, M., & Ghaderi, A. Predicting Long-Term Response in a Mathematics Tablet Intervention. *Manuscript submitted for publication.*

Reprints were made with permission from the respective publishers.

Contents

Introduction 11
The impact of math on school achievement and society 11
The impact of early math competencies on later school achievement. 11
The impact of later math competencies on the economy 12
Effects of poor school achievement: adverse consequences for children 13
Risk factors for poor school achievement 14
Critical math skills 17
What is known about essential math skills in compulsory school? 17
Overview of mathematical development 18
Number competence 19
Fluency with whole numbers 19
Fluency with fractions 21
Certain aspects of geometry and measurement 22
Problem solving: applying critical math skills in context 23
Conclusions about critical math skills 23
The contribution of cognitive factors to mathematics achievement 24
Interventions to improve math achievement 25
Effects of math interventions 25
Interventions targeted at students with low math performance 35
Interventions teaching math fact 36
Intervention program: description, development, and theoretical foundations 38
Program description 38
Program development 40
Theoretical foundation 42
Assessment on a tablet 45
Objectives of the thesis 46
Study I - validating tests on a tablet 46
Study II - investigating effects of an intervention 47
Study III - predicting long-term response to intervention 47

Methods 48
Design 48
Study I 48
Study II 48
Study III 48

Participants 48
Study I 48
Study II...49
Study III 50
Instrument 50
Procedure 52
Study I 52
Study II...52
Study III 53
Statistical Analysis 53
Data mining and machine learning 53
Empirical studies 55
Study I - Validating tests..55
Study II - Effects of a math intervention 57
Study III - Predicting positive long-term response 63

Discussion 65
Summary of principle findings 65
Explaining the outcome 65
Interpretation of the magnitude of the intervention effect 65
Why was there a fade-out effect?...66
Why do children with lower non-verbal IQ seem to gain more?.........67
Potential moderating effects 67
Comparing the variable- and person-oriented analyses 67
Absence of effects on problem solving...68
Lack of effects on mental health 69
Limitations...69
Implications 71
Directions for future research 72
Conclusion 73

Acknowledgements 74

References 78

Abbreviations

BRIEF	Behavior Rating Inventory of Executive Function
CAI	Computer Assisted Instruction
CCC	Cover Copy Compare
HRT	Heidelberger Rechen Test 1-4
IRT	Item Response Theory
MA	Mathematics Condition
MA+WM	Mathematic combined with Working Memory Condition
MASC	Multidimensional Anxiety Scale for Children
OECD	Organization of Economic Co-operation and Development
PISA	Programme for International Student Assessment
RTI	Response To Intervention
SDQ	Strength and Difficulties Questionnaire
SES	Socioeconomic Status
SMFQ	Short Mood and Feelings Questionnaire
TIMSS	Trends in International Mathematics and Science Study

Introduction

The impact of math on school achievement and society

The ability to read has long been acknowledged as a skill of outmost importance for the individual and society. Math competencies, on the other hand, have not been viewed as quite as essential to promote success in life and improve communities. However, evidence gathered during the last decade has repeatedly proven the last assumption wrong.

The impact of early math competencies on later school achievement

A meta analysis by Duncan et al. (2007) on six longitudinal data sets across three countries found that math competencies among 5-6 year old are the strongest predictor for general school achievement on standardized tests up to 13-14 years of age. In this study, including well over 40 000 children and controlling for several background factors, math skills were followed by reading and then attention skills in descending order of predictivity. These findings were replicated by Romano, Babchishin, Pagani, and Kohen (2010) who reported that math skills in kindergarten were the most powerful predictor of 3rd-grade school achievement, succeeded by literacy and attention skills. An additional interesting finding in this study was that math performance significantly predicted socio-emotional behavior. Pagani, Fitzpatrick, Archambault, and Janosz (2010) also replicated that math skills were the strongest predictor in a longitudinal study from kindergarten to 2nd-grade, followed by attention skills and receptive language skills. Although not predicting school achievement, a British study by Parsons and Bynner (2006) contributed with findings regarding the relative importance of math skills and literacy. The authors concluded, after examining two cohorts of approximately 17 000 persons followed from birth to 30 years of age, that low math skills (framed as low numeracy) was more strongly associated with negative outcomes on employment, physical and mental health, and legal infringements than low literacy (Parsons & Bynner, 2006).

The impact of later math competencies on the economy

In the book "The Knowledge Capital of Nations" (2015), the economists Hanushek and Woessman summarized their rigorous theoretical and empirical research by claiming that knowledge is the key to a nation's development. They conclude that a country's prosperity is directly related to the skills of its population. Furthermore, the authors claim that the pertinent cognitive skills - called the knowledge capital of a nation - could be adequately assessed by achievement on international math and science tests (e.g., Programme of International Student Assessment, PISA, and Trends in International Mathematics and Science Study, TIMSS). By analyzing historical data sets dating back to the 1960s Hanushek and Woessman demonstrate that fairly modest improvements in these skills can be worth multiples of a nation's current Gross Domestic Product (GDP). Additionally the authors point to examples that educational reform is possible, especially if educational institutions can align incentives with performance, and that countries which successfully reform will systematically depart in economical terms from other countries. More concretely, in a report by the Organization of Economic Co-operation and Development (OECD) the same authors stated, "an improvement of one-half standard deviation in mathematics and science performance at the individual level implies, by historical experience, an increase in annual growth rates of GDP per capita of 0.87%." (OECD, 2010, p. 17). Other long-term projections, this time not directly involving Hanushek and Woessman, indicate that gains for high-income OECD countries could gain 1.5 times their current GDP if every 15-year old acquires at least basic numeracy and literacy skills by 2030. For upper-middle income nations the gains were estimated to be about seven times their GDP (OECD, 2016).

Hanushek and Woessman (2009) have also found that high school students math proficiency drive the economic growth of a country more discernibly than proficiency in other subjects, which again supports the relative importance of math. At the individual level some research, though not conclusive, point to math skills as a better predictor of individual future earnings than other skills learned in high school (Bishop, 1992; Murnane, Willett, & Levy, 1995).

Plausibly, the demand for jobs requiring advanced math competencies is increasing compared to jobs without these demands, indicated by a report stating that growth for math-demanding science and engineering jobs were outpacing overall job growth three-to-one in the US (National Mathematics Advisory Panel, 2008).

Given the substantial influence of math on the prosperity for the society and the individual it is troublesome that almost every fourth student (23 %) in the OECD nations does not acquire the PISA baseline level 2 of mathematics proficiency and are duly defined as low performers. This level is viewed as essential for students to participate effectively and productively in the community (OECD, 2012). The required skills are by no means excessively advanced, for instance being able to interpret an ordinary table of values for different products, or to use a known exchange rate to convert a given amount in one currency to another. As an example in pure economical terms of the costs of poor math skills, the annual cost of low numeracy in the United Kingdom has been estimated to £2.4 billion (Gross, Hudson, & Price, 2009). A simplified summary of the relations presented so far suggest that early math achievement is a predictor of later school achievement, and that later school achievement predict long-term economic growth. Next, the relation between later school achievement and adverse consequences for children, as well as risk and protective factors for later school achievement, will be discussed.

Effects of poor school achievement: adverse consequences for children

While the previous section focused on the effects of math performance on school achievement and economical outcomes, this part presents the adverse consequences for children with poor school achievement (i.e., overall performance, not just achievement in math).

There are many detrimental psychosocial effects for students with poor school achievement. Björkenstam et al. (2011) found that the incidence rate ratio for suicide for students with the lowest grades was 2.7 for women and 4.6 for men compared to those with highest grades. The sample consisted of of a cohort of close to 0.9 million Swedes, followed over a time span of 9-18 years. The study controlled for a number of parental moribidity and socio-demographic variables, and measured final school grades from compulsory school. In a study by Griffin, Botvin, Doyle, Diaz, and Epstein (1999) on 743 seventh-grade students, poor grades were found to predict heavy smoking in twelfth grade. Pitkanen, Kokko, Lyyra, & Pulkkinen (2008) followed 347 persons over a course of 34 years and found that low school success was associated with problems due to drinking in early middle age. Criminality has also been found to be associated with low school performance (Farrington, 2015; Fergusson, Swain-Campbell, & Horwood, 2004). A striking example of the increased risk for criminality was reported by the National Board of Health and Welfare (2010) in Sweden, where students with low or incomplete grades from elementary school were at 8-10 times increased risk of serious criminality in young adulthood compared to those with medium or high grades. This pattern was evident in all socio economic groups. Yet, according to a systematic review the relationship between psychosocial problems and meager academic school performance is reciprocal, not causal, and exhibits a stable pattern from the beginning of school to early adulthood (Gustafsson et al., 2010). This implies that poor psychosocial conditions can lead to poor school outcome, and, conversely, also that low school performance can lead to psychosocial problems.

The same review pointed out that avoiding school failure during the first years of schooling is of particular importance to prevent subsequent negative trajectories.

Risk factors for poor school achievement

Since early math performance is the strongest predictor of later school achievement, it follows that low math performance is a risk factor for poor school achievement (Duncan et al., 2007; Pagani et al., 2010; Romano et al., 2010). This section describes what other factors, especially risk factors, have been found to influence poor school achievement (i.e., math, reading, and science). The concept of risk factor has been used frequently across different scientific disciplines (e.g., medicine, psychology, and education). The definition of risk factor used here is similar to how the term has been framed in other research fields, namely a risk factor is a condition that increases the likelihood of poor school achievement. Although the associated concept of protective factors, defined as an influence that inhibits, reduces, or buffers the probability of a behavior (Brounstein, Zweig, & Gardner, 2001), also applies to the current topic, the main focus here is a brief look at risk factors for poor school achievement.

An interesting finding in a recent PISA-study was that poor achievement at age 15 was not an effect of any single risk factor, but rather a combination and accumulation of different obstacles and disadvantages that have life-long consequences for students (PISA, 2016). While these findings may not be surprising or new to many people, they are important to bear in mind when discussing school policies and educational reform. Furthermore, this data parallels the results found in substance abuse prevention where the accumulation of risk and protective factors, not individual factors, is considered crucial (Brounstein, Zweig, & Gardner 2001).

Looking more specifically on the large amounts of data coming from the PISA studies, gathered and analyzed by the OECD, as well as from a few research groups, an abundance of risk factors for poor achievement at age 15 have been identified. At the individual level the following risk factors for low performance have been identified across the OECD countries: having low socio-economic status (SES), having an immigrant background, speaking a different language than the instructional language at school, living with a single parent, living an a rural area, having a year or less of pre-primary education, having repeated a grade, being on a vocational track, missing classes (truancy), low perseverance/effort in school, and low self-confidence (OECD, 2016; Sirin, 2005). For instance, in the OECD countries students with an immigrant background speaking a different language at home than at school are approximately 2.5 times more likely to be low achievers. Students with low math performance tend to have less self-confidence, perseverance and motivation, and skip school more. The last finding is illustrated by students taking the PISA test where low performers are three times more likely to have skipped school at least once in the two weeks prior to the assessment.

As addressed above, the body of research on risk factors suggests that it is not any single factor that leads to low performance, but instead an interplay and an accumulation of processes and ordeals that impedes learning and accordingly raise the probability of low achievement (e.g., Alexander, Entwisle, & Horsey, 1997; Hao, Hu, & Lo, 2014). Figure 1 shows an example of how the probability of being a low performer increases as a child presents with increasingly more risk factors. The upper line represents socio-economically disadvantaged children, and the lower line represents socio-economically advantaged children. For students not having any of these risk factors, the original risk for being a low achiever is 7 % higher for disadvantaged than advantaged students. On the other hand, among students having all risk factors, the risk of being a low performer is 16 % higher for disadvantaged students. This demonstrates how adding more risk factors leads to differentially accentuated risks for students with low SES compared to student with high SES. Another way to conceive this is that the "penalty" connected to each additional risk factor is somewhat bigger for disadvantaged students than for advantaged students. This pattern is representative for the OECD-countries in general, but other patterns have been identified for certain groups of nations (OECD, 2016).

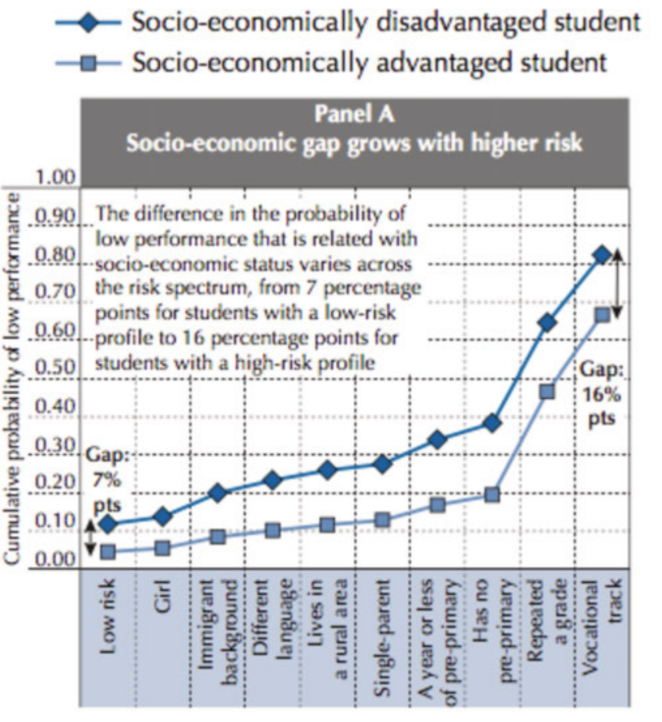


Figure 1. Pattern of Risk Accumulation Across Countries. Cumulative probability of low performance in mathematics. Average of the following 8 countries, which is similar to the OECD average: Croatia, Finland, Iceland, Italy, the Netherlands, the Russian Federation, Spain and Viet Nam. Figure reprinted with permission.

Risk factors are also assembled at the school level, and at an even more aggregate level in the school system. At the school level, a number of risk factors for low achievement have been identified: schools with a high concentration of disadvantaged students, learning environments featuring low expectations for students, unsupportive teachers (low teacher morale), lack of after-school opportunities, uninvolved parents and communities, lack of qualified teachers, and lack of quality educational resources (Gustafson, Nilsen, & Hansen, 2016; OECD, 2016). Being a girl is furthermore a risk factor for low performance in math, while being a boy is a risk factor for low reading achievement. Lastly, at the school systems level a few additional risk factors have been pointed out: lack of school autonomy, early and rigid selection and grouping of students, and unequal allocation across schools. To further illustrate the effects of being socio-economically disadvantaged, individual SES accounts for 14.8 percent of the variation in performance across the OECD. This implies that a less-advantaged student in ninth grade has absorbed almost one year less schooling in mathematics, only due to socioeconomic factors, than a more-advantaged student. Another way to present this difference is that a less-advantaged student has a 39 points lower performance (100 points being one standard deviation) than an advantaged student, only because of socioeconomic factors. At the school level SES this picture is even more pronounced with an average student attending a low SES school scoring 104 points lower compared to an average student going to a socioeconomically advantaged school.

Despite the wealth of data just presented and the presence of risk factors at multiple levels, the PISA studies may not have covered a few dimensions where risk factors for low achievement could be expected. In the section on the effects of poor school achievement a reciprocal relation was reported between low performance and psychosocial problems (Gustafson et al., 2010). This finding could eventually suggest that psychosocial problems may be risk factors for low performance as well. It could probably be assumed that since the PISA-studies do not explicitly collect data on the mental health of students, they have not included this dimension in their analyses.

Another dimension that have not been addressed by the PISA-studies is the importance of intelligence as predictor of school achievement. A British prospective longitudinal study of more than 70 000 children, measured on general IQ at age 11, reported that general IQ accounted for 59 % of the variance in mathematics and 48 % of the variance in English on national examinations at age 16 (Deary, Strand, Smith, & Fernandes, 2007). Some additional support was provided by Streze (2007) in a meta-analysis on longitudinal studies, which demonstrated that intelligence was a powerful predictor of grades (considered a proxy for school achievement here) with a correlation of .56.

Lastly, the dimension of heritability should also be considered a risk factor. Having parents with low math performance may be a risk factor for poor school achievement in general, and for math performance in particular since mathematical ability has medium to high levels of heritability. A study by Davis et al. (2008) demonstrated 49 % heritability at age ten and an investigation by Greven, Kovas, Willcutt, Petrill, and Plomin (2014) showed 46 % heritability at age twelve. Kovas et al. (2013) found somewhat higher estimates: 66 % at age six, 73 % at age nine and 56 % at twelve years of age.

After looking at risk factors for school performance, the adverse consequences of poor school achievement, and the importance of math for later school achievement, we will now look into what the critical skills are when acquiring proficiency in mathematics.

Critical math skills

What is known about essential math skills in compulsory school?

The research on mathematical development has been said to pale in comparison with the exhaustive research that has been carried out in the field of reading (Clarke, Nelson, & Shanley, 2016). Fuchs, Fuchs, and Compton (2012) also concluded that less is known about how mathematical abilities evolve than reading, and provided some background for this fact. Measurement studies have identified five basic components in reading: phonological awareness, decoding, fluency, vocabulary, and comprehension (e.g., Mehta, Foorman, Branum-Martin, & Taylor, 2005). Fuchs et al (2012) put forward the need for corresponding measurement studies in the area of mathematics. Additionally, the assumption is that more component skills (e.g., numeration, concepts, math facts, measurement, procedural calculation, and word problems in primary school) exist in mathematics, which is mirrored in the curricula. It is also uncertain if improved performance in one component affects other components (Fuchs et al., 2012). Yet, despite the complexities of the mathematics domain, some more recent research has productively focused on finding important predictors between earlier math skills and later math achievement (e.g., Bailey, Siegler, & Geary, 2014; Siegler et al., 2012), which will be presented in the next section.

Overview of mathematical development

Empirical evidence for mathematical competencies has been found already in infancy, with 9-month-old infants possessing a magnitude-based estimation system (McCrink & Wynn, 2004).

Looking at mathematical development from kindergarten to high school several critical component skills have been suggested, at times leading to intense debates were the relative importance of procedural skills, fluency in basic facts (i.e., addition, subtraction, multiplication, and division), and conceptual knowledge have been disputed (National Mathematics Panel, 2008). However, even if much more research is needed to draw firm conclusions about a comprehensive model of mathematical development, some critical math competencies should be considered as established by research at this point. Early number competence, fluency with whole numbers, fluency with fractions, and particular aspects of measurement and geometry should be viewed as essential mathematical competencies (Bailey et al., 2014; Jordan, Kaplan, Ramineni, & Locuniak, 2009; National Mathematics Panel, 2008; Siegler et al., 2012). Fluency with whole numbers and fractions, and certain aspects of measurement and geometry all serve as distinct foundations for higher order mathematical skills required in algebra which is considered an essential part of high school mathematics. Even if algebra is not the only mathematics component taught at high school, and may not be the only relevant outcome math variable for mathematics at large, it has been viewed to be of profound importance for entry into mathematics-intensive fields (Fuchs et al., 2012; National Mathematics Panel, 2008; Siegler et al., 2012).

Number competence

Early number competence has been studied relatively frequently, at times under the concept of number sense (Jordan, Kaplan, Locuniak, & Ramineni, 2009), and it has been said to describe the ability to understand the meaning of numbers and relationships between numbers (e.g., Gersten, Jordan, & Flojo, 2005; Malofeeva, Day, Saco, Young, & Ciancio, 2004). The origins of number competencies are found already in 6-month-old infants who can differentiate between numerosities of 8 vs. 16 dots using a habituation paradigm (Feigenson, Dehaene, & Spelke, 2004). Meanwhile the development of number competence, as defined in research, typically takes place a few years later before starting kindergarten and lasting approximately up to the first school year. More specifically, number competencies include the capacity to comprehend the value of small quantities instantly, differentiate between numbers and their magnitude (e.g., 5 is closer to 6 than 3), join and separate sets (e.g., 4 and 2 makes 6, and 6 take away 2 is 4), and understand counting principles (e.g., the total number of a set is told by the final number counted) (Jordan et al., 2009). Number competencies also include having a linear representation of small numbers, for instance that a number that comes after is always one more (Siegler & Booth, 2004). The last skill is also known as number line estimation, and performance at age six on extensions of this part of number competence (i.e., students estimate where numbers go on a number line from 0-100) has shown to predict math performance five years later (Geary, 2011). A study by Geary, Hoard, Nugent, and Bailey (2013) also demonstrated that a slightly different version of number competence (termed number system knowledge) at age six predicted functional numeracy at age 13. Ngyuen et al. (2016) provide additional support for the claim that number competencies (more specifically counting and cardinality) in preschool is the strongest predictor of math achievement in fifth grade in a sample of mainly low-income and minority children.

Fluency with whole numbers

Fluency with whole numbers could be considered the next critical math skill after number competencies, although the terms overlap to some degree since fluency with whole numbers include extensions of number competence (National Mathematics Advisory Panel, 2008). Fluency with whole numbers consists of several component skills. Important components are an understanding of the basic operations of addition, subtraction, multiplication, and division, and an ability to automatically recall such math facts (e.g., 7+8=15, or 12/3=4), as well as to compose (e.g., 9+3 could be composed of first 9+1=10, and then 10+2=12) and decompose whole numbers (e.g., 8 could be decomposed to 4+4). Other components included in fluency with whole numbers are an understanding of place value, the use of commutative, associative, and distributive properties, computational dexterity, and application of these skills to problem solving. Fluency with whole numbers also includes a capacity to estimate the results of computations. It has been recommended that students should be fluent with whole numbers using addition and subtraction at the end of third grade, and with multiplication and division at the end of fifth grade (National Mathematics Advisory Panel, 2008).

An important component skill of fluency with whole numbers: Math facts

The component skills described above could each be further elaborated, but the focus here will be on basic operations with addition and subtraction since this component skill is the primary target of the intervention in this thesis. This component skill will subsequently be referred to as math facts, which covers addition and subtraction facts up to 20 in most referenced work hereafter.

Proficiency with math facts is considered an important early skill in mathematical development (e.g., Fuchs et al, 2012; Geary, Hoard, Nugent, & Bailey, 2012; Gersten & Chard, 1999). A study by Bailey, Siegler, and Geary (2014) showed that whole number arithmetic (i.e., math facts) in first grade, together with whole number magnitude knowledge (i.e., number line estimation), predicted fraction knowledge in 7-8 grade controlling for several cognitive abilities, parental education and income, gender and race. Furthermore, Geary (2011) demonstrated that whole number addition at the beginning of first grade predicts mathematics learning through the termination of fifth grade. Competence with math facts is also an indicator of risk for long-term learning disabilities (Geary, Hoard, Nugent, & Bailey, 2012; Goldman et al., l988; Jordan et al., 2003), and viewed as a significant path to more advanced mathematics (e.g., Fuchs, Fuchs, Compton, et al, 2006). The importance of math facts has been explained by the bottle neck hypothesis, which states that students with math facts deficits are slowed down when solving more complex procedural calculation since their attention is consumed by the inherent math fact problems (Fleishner et al, 1982; Fuchs et al., 2012; Fuchs et al., 2009; Geary et al., 1987; Goldman et al., 1988; Jordan et al, 2003; Vasilyeva et al., 2015). Indeed, a study by Fuchs et al. (2013) concluded that improvement in math fact, due to their intervention, had a transfer effect on procedural calculation.

What aspects of math fact skills are most important?

High performing Asian countries (e.g. China, Korea and Japan) focus classroom instruction on intense computational practice in early grades, supporting a quick and accurate retrieval of math facts (Leung, 2001). Investigations of cross-national differences on single-digit addition between Asian and American students at the end of first grade also stress the importance of automatic retrieval of math facts, which was the prevalent strategy among

Asian students, instead of counting which was the dominant method among Americans (Geary, Bow-Thomas, Liu, & Siegler, 1996; Vasilyeva et al., 2015). More specifically, the key predictors of accuracy on single-digit problems were the frequency of correct retrievals and mean retrieval time (Geary et al., 1996). A similar result on single-digit addition was demonstrated in a study on students with a higher level of math performance (Geary et al., 1992). A study comparing students from America and Taiwan reported that the frequency of automatic retrieval and decomposition (*e.g.* sequencing 4 + 5 into 4 + 4 = 8 and then 8 + 1 = 9) on single-digit addition fully mediated cross-national differences on double-digit addition, emphasizing the importance of math facts and decomposition for more complex arithmetic performance (Vasilyeva et al., 2015). Another finding in this study was that successful accomplishment of a decomposition strategy on double-digit tasks seemed to require a high degree of fluency with math facts. Conclusively, it could be stated that research so far has found that fast and correct retrieval of math facts is the most important feature of the math fact component skill. Plausibly, the ability to decompose number is a significant subsequently developed skill within the math fact component.

Fluency with fractions

Being fluent with fractions includes being capable to place positive and negative fractions on a number line, and estimate magnitudes of fractions. It also entails an ability to compare fractions, decimals and percent, and comprehend how finite decimal numbers can be fractions. Fluency with fractions additionally means being proficient in carrying out operations where sums, differences, products, and quotients are handled as fractions. Students should also be able to describe probability, proportionality, and rates when used in a various context. When accurately taught, fractions acquaint the student with how to generalize and use symbolic notation, both being inherent in algebra (National Mathematics Advisory Panel, 2008). Siegler et al. (2012) aimed at finding early predictors of high school mathematics achievement in a study involving more than 4000 students in two longitudinal samples from the United States and the United Kingdom. Student knowledge of fractions and of division at age 10-12 uniquely predicted performance in algebra and overall mathematics achievement 5-6 years later in high school. The study controlled for other mathematical skills, working memory, IQ, and family education and income. Recommendations are that students should be fluent with initial fractions skills at the end of fourth grade (i.e., represent and identify decimals and fractions, and compare them on a number line), and with the completive fractions skills by the end of seventh grade (i.e., operations with positive and negative fractions and integers, and to solve problems including ratio, rate and percent and eventually proportionality) (National Mathematics Advisory Board, 2008).

Why is fluency with fractions so important?

The integrated theory of numerical development provides an interesting explanation of the significance of fraction knowledge (Siegler, Thompson, & Schneider, 2011). The theory points out that a crucial part of numerical development is to learn that many characteristics of whole numbers are not universally true for numbers. For example, whole numbers are countable and have a restricted number of units within any given interval, they have unique successors, and they never decrease with addition and multiplication. For the majority of children, fractions give the first chance to learn that many properties of whole numbers are not true of all numbers. For instance, in fractions students learn that the product in multiplication can be smaller than the multiplicands, for instance 0.8 x 0.8 = 0.64, or that although 8 is greater than 4, 1/8 is smaller than 1/4. Other theories on mathematical development tend to focus on whole numbers and view fractions as of secondary importance, and not give any hints that fluency with fractions is a unique predictor of later mathematics achievement. Rather, these theories normally just describe how learning of fractions is disturbed by experience with whole number (e.g., Gelman & Williams, 1998; Wynn, 1995). Siegler et al. (2012) give an illustrative example of how fraction skills might be essential for later mathematics, and algebra in particular, and the necessity for students to be able to properly estimate answers to simple algebraic equations:

> For example, students who do not understand fractions will not know that in the equation $1/3X = 2/3Y$, X must be twice as large as Y, or that for the equation $3/4X = 6$, the value of X must be somewhat, but not greatly, larger than 6. Students who do not understand fraction magnitudes also would not be able to reject flawed equations by reasoning that the answers they yield are impossible. (Siegler et al., 2012, pp. 692)

Support for this analysis has been provided in studies demonstrating that correct estimation of fraction magnitudes is closely connected to the use of fraction arithmetic procedures (Hecht & Vagi, 2010; Siegler et al., 2012).

Certain aspects of geometry and measurement

According to the National Mathematics Advisory Board (2008) knowledge of similar triangles are of direct importance when learning algebra, since the logic of similar triangles directly corresponds the slope of a straight line used in linear functions. Additional geometric and measurement abilities supposed to contribute to algebra knowledge are: determine area, volume, perimeter, and area using formulas; interpret characteristics of two- and three-dimensional shapes; and find unrevealed lengths, areas, and angles. These skills should incrementally be in place through the end of fifth to the end of seventh grade.

Problem solving: applying critical math skills in context

It could be argued that problem solving should be viewed as a separate critical math skill as well. However, it is not obvious that such a claim is supported by the literature since longitudinal studies on the predictive ability of problem solving skills on later math achievement are hard to find. Perhaps this could be due to how problem solving is conceptually defined. It makes intuitively sense that the importance of math skills, on the individual and societal level, is largely due to effects of utilizing math skills when solving problems. Plausibly, mathematical problem solving abilities may be the consequence of (1) gradually learning the critical math skills just presented, and (2) continually applying these skills to math problems in various contexts. However, when students apply critical math skills in different problem solving situations they do not solely rely on those mathematical skills, but also on language skills as shown by Fuchs et al. (2006). A Finish study demonstrated that reading comprehension in fourth grade predicts mathematical word problem solving ability (i.e., word problems where math facts are presented in context) in seventh grade controlling for text reading fluency and basic calculation ability (Björn, Aunola, & Nurmi, 2016). A correlational study by Decker and Roberts (2015) found that verbal comprehension contributes to explaining variance in word problem solving controlling for calculation skills.

Given that mathematical problem solving depends on both critical math skills and verbal abilities, the argument here is that mathematical problem solving rather is part of an important outcome in mathematics than one particular critical skill that is to be developed within mathematics. Thus, problem solving should be incorporated throughout teaching of critical math skills and seen as a capability that is supportive of learning those skills, as has been suggested by the National Mathematics Advisory Board (2008).

Conclusions about critical math skills

Number competence, fluency with whole numbers, fluency with fractions, and particular aspects of measurement and geometry are arguably the most critical math skills to be developed in compulsory school. Yet, this conclusion should be interpreted with some caution since the field of mathematics development is relatively immature, for instance when compared to the extensive research on foundational skills in reading. In particular, the claim that certain aspects of geometry and measurement is necessary for long-term math achievement needs to be further studied. One reason for the relatively large lack of empirically derived knowledge about mathematical development could be due to difficulties in finding a manifest and broadly accepted outcome criterion for mathematical achievement (e.g., should primarily algebra be in focus or not).

The contribution of cognitive factors to mathematics achievement

The influence of cognitive functions on performance in mathematics has been relatively widely studied with new studies being reported continuously. Given the focus of this thesis, which will be presented in the next section, the center of attention here is on how cognitive capacities, in particular working memory, affect math facts (an essential part of the critical skill fluency with whole numbers), and mathematical word problem solving (e.g., math facts presented in a text that provides a context for the problem at hand).

Working memory influences math achievement

Among cognitive factors relating to mathematics achievement, working memory has been particularly well investigated. There is evidence that working memory capacity is strongly correlated with general mathematics achievement (Geary, 2011) and the ability of working memory to predict math performance has been investigated by several authors. Two longitudinal studies by Li and Geary (2013; 2017) demonstrated how visuospatial memory, considered a part of working memory, uniquely predicts the ease of learning some types of mathematics. The first study (2013) compared how children's visuospatial memory, controlling other variables such as intelligence, predicted achievement gains in mathematics from first to fifth grade and concluded that gains in visuospatial memory stood out as the most important factor. In the follow-up study, Li and Geary (2017) measured the same students on mathematics from sixth to ninth grade, and could finally infer that improvement in visuospatial working memory across elementary school has a unique influence on the ease of learning certain mathematical skills and that this effect increases successively across grades. The types of mathematics measured covered were number system knowledge and math facts in the earlier study on first-to-fifth graders (Li & Geary, 2013), and numerical operations, rational number problems and simple algebra and geometry in the later study on sixth-to-ninth graders (Li & Geary, 2017). Additional support that visuospatial memory predicts later mathematics achievement has been provided by Decker and Roberts (2015), and in another longitudinal study by Geary et al. (2009) that followed children from kindergarten to third grade. Since visuospatial memory is part of working memory it is not surprising that a less differentiated investigation on working memory showed similar influence of working memory on math performance (Kroesbergen & van Dijk, 2015).

Focusing more specifically on how working memory predicts performance on math facts and word problem solving, Viterbori, Usai, Traverso, and De Franchis (2015) demonstrated that working memory ability at age 5 predicts performance on these math outcomes in third grade. Further evidence for the impact of working memory on problem solving ability was demonstrated in studies by Swanson and Beebe-Frankenberger (2004) and Zheng, Wanson, and Marcoulides (2011). Furthermore, Swanson (2006) found that the executive component of working memory predicted outcome in math facts, computation, and problem solving. Lastly, Vasilyeva, Laski, and Shen (2015) pointed out that working memory, besides contributing to computational fluency, also is important when using decomposition strategies, which has been suggested to be a vital part of fluency with whole numbers.

Other cognitive factors influencing math achievement

Although not to the same extent, studies on other cognitive functions have also documented influence on math achievement. Fuchs et al. (2006) demonstrated relationships between inattentive behavior and math facts, algorithmic computation, and problem solving skills. Phonological ability has also been considered as significant, since phonological processing has been shown to be important for word problem solving (Swanson & Beebe-Frankenberger, 2004), and phonological decoding was found to be an independent predictor of math facts (Fuchs et al., 2006). The latter study also reported that processing speed was a unique predictor of math facts and algorithmic computation, and Decker and Roberts (2015) found that processing speed was a significant predictor of word problem solving. Finally, fluid intelligence or reasoning has demonstrated a relationship with word problem solving (Swanson & Bebee-Frankenberger, 2004), and to predict subsequent word problem solving capacity (Decker & Roberts, 2015).

Interventions to improve math achievement

Effects of math interventions

Because the focus of this thesis primarily to teach math facts to students with low math performance, using technology that will be further described later, this section will attempt to describe (1) what constitutes a high quality study in the field of educational interventions, (1) the effects of math interventions and especially educational technology interventions, (2) interventions targeted at students with low math performance, and (3) interventions teaching math facts.

To put the intervention evaluated in this thesis in context, it is necessary to know what counts as reliable evidence according to methodological research in the field of educational interventions.

A comprehensive methodological article by Cheung and Slavin (2015) examined 645 studies from 12 recent reviews of evaluations of mathematics, reading, and science programs, including technological and non-technological applications. The authors reported approximately twice as large effect sizes for small-scale trials (i.e., < 250 students), published documents and experimenter-made measures, than for large-scale studies, unpublished articles and independent measures. Independent measures typically included standardized tests, but also experimenter-made measures if they were comprehensive and not only measured objectives inherent to the objectives of the intervention (but unlikely to be stressed in the control group). Findings also suggested that quasi-experiments had significantly larger effect sizes than randomized experiments.

In this methodological study, Cheung and Slavin (2015) suggested three possible explanations for their results. First, small-scale studies are often more closely controlled than large-scale studies, and therefore more to likely to preserve high implementation fidelity, or provide extra help that could hardly be replicated in a large trial. Consequently, this type of strengthened control in small-scale studies is more likely to produces positive outcomes. Another potential explanation is that researcher-developed tests are more frequently utilized in small studies. These measures may have a narrower scope and are possibly more sensitive to treatments than standardized test, which are often used in large-scale trials. This explanation is supported by a meta-analysis by Li and Ma (2011) that found larger effects among studies using non-standardized tests compared to investigations utilizing standardized tests. The last explanation is that the limited statistical power in small studies leads to over-reporting of high effect sizes since large effects are required to get significant results. It is conceivable that small trials without significant effects are discarded as "pilots" and put in the drawer by researchers, or are rejected by journal editors (Cheung & Slavin, 2015).

A closer examination of educational technology interventions for enhancing math achievement revealed similar methodological effects (Cheung & Slavin, 2013). As mentioned in the previous section, randomized trials have produced significantly smaller effect sizes than quasi-experimental trials within the larger educational field. This finding should be kept in mind when interpreting the results in the meta-analysis by Cheung and Slavin (2013) on educational-technological (ed-tech) math interventions, where only 35 % (or 26 out of 74) of the studies were randomized. The average effect size for quasi-experimental trials was more than twice that of randomized trials in this meta-analysis. The effects of small studies (< 250 students) were about twice that of large studies, which repeated the previously reported findings on technological and non-technological interventions in math, science and reading. Because Cheung and Slavin (2013) only included published studies, but not studies with experimenter-made measures, these two features could not be investigated in the meta-analysis on ed-tech interventions for improving mathematical skills. Furthermore, interventions using short durations may produce biased results since brief durations tend to produce larger effects than long-duration investigations. Consequently, Cheung and Slavin (2013) only included studies with duration of 12 weeks or more in their meta-analysis. The authors concluded that there is an "urgent need for more practical randomized studies in the area of educational technology applications for mathematics" (Cheung & Slavin, 2013, p. 101). A similar point was made by Dowker (2017) in an overview of interventions for primary school children with difficulties in mathematics, where she noted the discrepancy between the number of computer games developed for intervention purpose compared to the relatively few systematic evaluations of these programs.

More recently published educational technology studies with similar scope as the current study have improved somewhat on these issues, for example larger samples (Burns, Matthew, Rebecca, & Megan, 2012; Carr, 2012; Pitchford, 2015; Riconscente, 2013). Nevertheless, the same studies are obscured by the following limitations: short duration with only one week (Carr, 2012), eight (Pitchford, 2015) or nine weeks of training (Riconscente, 2013), quasi-experimental design (Burns et al., 2012; Carr, 2012), experimenter-made measures (Pitchford, 2015) or lack of psychometric information (Carr, 2012), lack of control of extraneous variables with significantly fewer students practicing simultaneously in the intervention group than in the comparison groups (Pitchford, 2015), or utterly few classes included (Riconscente, 2013), and scarcity of participants characteristics which limits generalization (Burns et al., 2012).

Many studies have investigated the effects of different types of interventions targeted at mathematical skills. There are plenty of options on how to categorize the various interventions (e.g., Clarke, Nelson, & Shanley, 2016; Dowker, 2017), where format of delivery arguably would be an adequate way since the ease of implementation and scaling of an intervention largely depends on how the intervention is delivered. The argument here is that a research-based intervention needs to be able to scale relatively well in order to motivate the large resources spent in this type of research (i.e., randomized controlled studies). Categorizing interventions according to delivery format yields three distinct types of interventions: whole-class instruction, individual or small group instruction with a designated tutor, and different versions of computer assisted training. The last category could also be labeled educational technology interventions, and will subsequently be referred to as ed-tech interventions.

Whole-class instruction

In general, teaching mathematics to whole-classes has not been regarded as part of intervention research, but rather as part of curriculum development and relatively few studies have investigated whole-class instruction using a rigorous design. A study by Jitendera et al (2015) utilized whole-class instruction and it could serve as a good example of a well designed study within this context, because it had a large sample, randomization at the class level, used multilevel analysis to account for nesting effects and effective handling of intention to treat, high fidelity of implementation, and standardized measures. The purpose of the study was to promote proportional problem-solving among seventh-graders including 1999 students, and 82 teachers/classrooms that were cluster randomized. The teachers assigned to the intervention condition participated in a two-day professional development course on how to teach schema-based instruction to a whole-class. Students in the schema-based instruction group scored significantly better than the control group at post assessment and at retention tests after nine weeks, with a small-medium sized effect (d=0.46). This shows how a whole-class approach can be used to teach specific skills, and that the requirements are very high for this type of research: recruitment of nearly two thousand students across 50 school districts, implementation of the program in 40 classrooms, and assessments at four time points is an impressive work. Yet, to randomize correctly, using classes as unit for randomization, such large numbers of students are required to achieve the power necessary to detect relevant levels of significance.

Since many countries have become increasingly concerned with improving mathematical performance, there is an increasing engagement in examining the potential of whole-class programs in reducing the shares of children with mathematical difficulties (Dowker, 2017). As reported in Dowker (2017), the British program Mathematics Mastery is an example of this. The program is partly inspired by Singapore mathematics education and focus on teaching a few topics in depth compared to regular curricula. An evaluation (not peer-reviewed) by the charity Education Endowment Fund, including more than 4000 students in 82 schools, reported that students in grade 1 in the schools randomized to Mathematics Mastery made 2 months gain in mathematics knowledge compared to control schools. Meanwhile, these gains were not significant and the effect size reported was d=0.10.

Individual or small group instruction

Several effective intervention programs using individual or small group instruction have been identified within the responsiveness-to-intervention (RTI) framework which has been a significant educational reform in the United States in the 21th century (Johnson & Street, 2012). The first level of RTI is general education (i.e., primary prevention), where students are regularly screened to identify children with risk of poor learning outcomes. At-risk learners are then provided more intensive instruction, usually involving at least one round of small-group tutoring (i.e., second level of prevention), and eventually individualized instruction if needed (i.e., third level of prevention). The small-group interventions are often led by a designated tutor following a manual, conducted in first to third grade, and take place about three times per week during 15-20 weeks (Fuchs, Fuchs, and Compton, 2012).

One illustrative example of a RTI study with small group instruction was conducted by Fuchs, Compton et al. (2005) where 537 first graders, whereof 127 were at-risk for mathematics difficulty, were randomized to tutoring or control conditions. Tutoring took place three times weekly, each session was 40 minutes, and the intervention lasted 16 weeks. Tutoring was led by a student with a university degree following a script after completing a 1-day training session. The tutoring program was based on the concrete-representational-abstract model (Mercer, Jordan & Miller, 1996), which departs from concrete objects to facilitate conceptual learning. The topics covered were introduced sequentially: writing and identifying numbers to 99; identifying less, more, and equal with objects; sequencing numbers; using the symbols =, <, and >; skip counting by 2s, 5s, and 10s; understanding place value; identifying operations; place value 0-50; writing number sentences; place value 0-99; addition facts with sums to 18; subtraction facts with minuends to 18; review of addition and subtraction facts; place value; 2-digit addition; 2-digit subtraction; and missing addends. Results demonstrated that tutoring improved computation and concepts/applications, but no fact fluency.

Another example of a RTI study using small group instruction was provided by Fuchs, Powell et al. (2009) where 133 third graders with mathematics difficulties were randomly assigned to one of three conditions: control (no tutoring), tutoring on automatic retrieval of math fact (i.e., a scripted program called Math Flash), or tutoring on word problems focusing on foundational skills of math fact, procedural calculations, and algebra (i.e., a scripted program called Pirate Math). In the randomization, students were stratified across two sites and mathematics difficulty status (mathematics difficulty only vs. mathematics and reading difficulty). As in the study by Fuchs et al. (2005) tutoring occured three times per week during 16 weeks, although the duration of a session was slightly shorter being 20-30 minutes. The tutoring group was led by a research assistant who completed a checklist to identify the percentage of essential objectives covered in each session. All sessions were audiotaped and a sample was analysed showing that the mean percentage of objectives covered varied between 98.1-99.5 % across sites.

The condition training automatic retrieval of math fact (i.e., Math Flash) practiced 200 math facts with addends and subtrahends from 0 to 9. For the major part of the session the tutor, following a protocol (not reading verbatim), thaught strategies (e.g., counting up) and provided explanations while students were practicing. A minor part of the session, up to 7.5 minutes, was devoted to computerized training (thereby the name Math Flash because numbers flashed on the screen), where students were briefly shown a math fact on the screen and then required to type the complete math fact. If the answer was correct the student heared applause and earned a point, and if the answer was incorrect the student had another chance to answer correctly.

The condition practing word problems (i.e., Pirate Math) consisted of four units where the first unit focused on training math fact (not using computer); review double-digit addition and subtraction; solve for X in any position in easy algebraic equations; and teach students to check their word problem work. The other three units consisted of training word problems of gradually ascending complexity, while integrating skills from the first unit. Each unit introduced a category of word problems, where the types taught were: Total (combining two or more amounts), Change (an initial amount that decreases or increases), and Difference (comparison of two amounts). Previously thaught types were systematically revisited in the last units.

Results showed that Math Flash improved fluency with math fact and transfer to procedural computation, but no transfer to algebra or word problems. Pirate Math enhanced word problem, as well as and fluency with math fact, procedural computation, and algebra. No differential effects of tutoring was found for students' with differing mathematics difficulty status (Fuchs et al., 2008).

Fuchs, Fuchs, and Compton 2012 summarized the findings from the described studies (Fuchs, Compton et al., 2005; Fuchs, Powell et al., 2009) and two other RTI-studies (Fuchs, Fuchs et al., 2008; Fuch, Geary et al., 2013) by concluding that these interventions make meaningful differences in decreasing the achievement gap, or at least stopping the growth of the achievement gap, between at-risk students and normally developing children. Effect sizes reported in these studies were medium sized (d=0.50-0.79).

Programs using individualized instruction have also been found to be effective in remediating students with low performance in mathematics. The Mathematics Recovery program, designed in Australia, is an example of a program where teachers provide individualized intensive instruction to low performing 6-7 year olds during 30 minutes per day over the course of 12-14 weeks (Wright, Martland, & Stafford, 2006). The effects of Mathematics Recovery have been documented in a study by Smith, Cobb, Farran, Cordray, & Munter (2013) with very small to small effect sizes (d=0.15-0.30) compared to controls at post assessment. Number Counts is another program targeted at 6-7 year olds with low math attainment. Children participating in the program receive on average 40 hours of individualized instruction by a teacher during one semester. According to a report (not peer-reviewed) by Torgerson et al. (2011) where 409 children were randomized to either Number Counts or a wait list, children participating in Number Counts improved significantly with an effect size of d=0.33.

It should be kept in mind that intensive individualized interventions, such as Mathematics Recovery and Number Counts, are estimated to be needed by only 5 % of mathematics learners (Dowker, 2017). Since at least 20 % of adult people have persistent numeracy problems, with serious economical, social, and practical consequences, there is clearly a need for less intensive interventions as well, such as the small-groups interventions described above.

To sum up, well designed studies on whole-class instruction are rare but some studies have been conducted showing significant, but very small to small effects. Instruction in small-group has been more thoroughly studied and typically medium sized effects have been found. Individualized intensive instruction has been less well investigated, but findings indicate that individualized training can improve students learning significantly.

Interventions using educational technology (ed-tech interventions)

Interventions using technology can of course also be delivered in the previously described formats (i.e., whole-class, small-group or individualized instruction), but the point here is the difference in how instructions are given. In ed-tech interventions most or even all instructions and feedback are provided by a computer instead of a teacher or tutor as described earlier.

Looking at the bigger picture of effects of ed-tech interventions in mathematics, Cheung and Slavin (2013) concluded that educational technology programs for enhancing mathematics achievement are only making a modest difference compared to traditional teaching methods without educational technology. They authors reported an estimate of overall effect size of d=0.15, including various types of ed-tech math interventions. To put this in context, What Works Clearinghouse considers an effect size of d=0.20 to be meaningful when comparing programs with equal dosage (What Works Clearinghouse (2014). However, since the meta-analysis by Cheung and Slavin (2013) calculated one average effect size for all outcome measures in each study we do not know the effects on the best outcome measure in each study. This implies that despite the very small effects reported in this meta-analysis there could be effects on specific, and important outcomes in ed-tech math interventions that are not put forward. Still, the meta-analysis by Cheung and Slavin (2013) is the best, and most recent, available source on such effects and will therefore be used when looking into a more detailed analysis on different types of ed-tech math interventions.

What types of ed-tech interventions exist within mathematics education? In their meta-analysis on ed-tech interventions in mathematics, Cheung and Slavin (2013) derived three categories from the literature: (1) supplemental computer assisted instruction (supplemental CAI programs), (2) computer-managed learning systems, and (3) comprehensive models. Examples of supplemental CAI programs were Jostens, PLATO, Larson Pre-Algebra and SRA Drill and Practice, where the program supplements traditional classrom instruction with additional instruction at students' assessed levels. Computer-managed learning systems included only Accelerated Math, which continuously measures students' math progress and performance, and assigns appropriate math material accordingly. Comprehensive models were programs such as Cognitive Tutor and I Can Learn that use computer-assisted instruction in conjunction with non-computer activities (for instance cooperative learning and teacher-led instruction).

What are the effects of these different types of ed-tech interventions? Supplemental CAI was found to be the most effective type among these three approaches, with an effect size of d=0.18 compared to d=0.08 for the computer-management program and d=0.07 for comprehensive programs (Cheung & Slavin, 2013). Kulik, Kulik, and Bangert-Browns (1985) and Niemiec, Samson, Weinstein, and Walberg (1987) found similarly small effects for computer-management programs in mathematics. However, it is interesting to note the superiority of supplemental CAI programs to comprehensive programs since a meta-analysis on the effects of ed-tech programs on reading performance by Cheung and Slavin (2011) showed that programs integrating computer and non-computer instruction produced larger effects (d=0.28) compared to supplemental programs (d=0.11). This suggests that a more integrated approach is more beneficial in reading than in mathematics (Cheung and Slavin, 2013), where the evidence points in favor of supplemental CAI programs.

A study by Hotard and Cortez (1983) provided an early and representative example of a study on a supplemental CAI program, in terms of effect size and not being peer reviewed (Cheung & Slavin, 2013). A total of 190 students were ordered in matched pairs and randomized to training CAI 10 minutes daily (content not described by authors), or receiving regular teaching using the same total dose of mathematics teaching. The sample were low performing students in third to sixth grade, recruited from two schools in lower socioeconomic areas, who had successfully received compensatory teaching, but were still behind in mathematics. The duration of training was 6 months, and effects sizes were d=0.19.

What are the characteristics of effective ed-tech interventions? Ed-tech math interventions are more effective for elementary students than secondary students (Niemiec et al., 1987; Slavin & Lake, 2008; Slavin & Smith, 2009). Cheung and Slavin (2013) found a similar result when comparing effect sizes, although the difference was not significant. Already in 1985, Kulik et al. concluded that high school students have less need for computer drills and tutorials with a very structured and adaptive format. Another characteristic of successful ed-tech programs are that program practiced more than 30 minutes a week have bigger effects than those practiced less than 30 minutes a week (Cheung & Slavin, 2013).

Even if it has been assumed that the effects of educational technology would increase in more recent studies along with the advancement of technology, meta-analyses has found no such evidence (Cheung & Slavin, 2013; Christmann & Badgett, 2003; Liao, 1998). Despite these shortcomings, Cheung and Slavin (2013) stated that educational technology will continue to play an increasingly important role in the near future and that the question no longer is whether teachers should use technology or not, but how to best implement programs in the classroom. This conclusion may be surprising given the small or minimal effect sizes presented above, but several arguments have been made. Common arguments for the use of mobile technology in teaching are reduced costs, increased adaptivity that facilitates learning, and scalability (Ozdamli, 2012). Furthermore, the National Council of Teachers of Mathematics (2011), a significant educational institution within mathematics, consider technology to be essential for teaching and learning mathetmatics, because it affects mathematics teaching and improves students' learning. One could also argue that technology may make it easier to identify struggling students in need of help, and that technology could facilitate home work. These assumptions have many proponents, and there is general trend of schools being encouraged to implement digital methods in their daily practice. This is manifested in Sweden by a recent policy document from the Swedish government stating that schools have a obligation to promote digital competence among students (Government Offices of Sweden, 2017). However, at the moment the past evidence does not support these assumptions. The future will tell if the promise of technology improving schools will be fulfilled.

Because the intervention in this thesis uses a tablet to teach mathematics, it is of particular interest to look more specifically at investigations on the effects of this relatively new technology. In a systematic review by Haßler, Major and Hennessy (2016), including 23 studies across various subjects for students between 5 and 20 years old, the authors concluded that the scarcity of rigorous studies made it difficult to draw firm conclusions about the effects of using tablets. Among the few intervention studies in mathematics using tablets, a study by Schachter and Jo (2017) is probably the most well designed. The study tested a comprehensive math program called Math Shelf with 4 to 5 year old preschoolers and found significant improvements after 22 weeks of training with large effects (d=0.96), and even larger effects for low performing children were found in a moderation analysis. Some limitations of the study should be noted: although classes were randomized the analysis was conducted on an individual level, data on fidelity of implementation was not provided, and the psychometric quality of the instruments could be questioned.

Among the few intervention studies in mathematics using tablets, a study by Schachter and Jo (2017) is probably the most well designed, because it had a large sample (N=433), a randomized design, and a long duration of training (22 weeks). However, some limitations of the study should be noted: although classes were randomized the analysis was conducted on an individual level, data on fidelity of implementation was not provided, and the psychometric quality of the instruments could be questioned.

More concretely, the study tested a comprehensive math program called Math Shelf with 4 to 5 year old preschoolers. Initially, each child took a placement test to set the child on an appropriate mathematics learning trajectory in the linear progression of 180 unique activities in the program. The program thaught sorting and matching, one-to-one counting, matching different quantity representations, numeral identification from 1 to 3, applying the cardinal principle, subitizing, sequencing numberals and collections, matching numerals to quantities from 1 to 6, and then 1 to 9. Significant improvements with large effects (d=0.96) were found, and even larger effects for low performing children were found in a moderation analysis.

Taken together, the above presented studies show that ed-tech interventions only are demonstrating very small overall effects. Supplemental computer assisted instruction, where the program supplements traditional classroom instruction with further instruction at the child's assessed levels, is the most effective type of ed-tech intervention. Ed-tech math programs are more effective when targeted at elementary school students, and when students practice more than 30 minutes a week. Despite the advancement of technology no increases in effects have been found in more recent studies on ed-tech interventions in mathematics. Still, experts in the field expect educational technology to play an increasingly important role in the classrooms (Cheung & Slavin, 2013). Lastly, although the effects of using tablets are unknown studies are beginning to accumulate (Haßler & Hennessy, 2016). Let us now turn to the important aspect of long-term effects of math interventions.

Lack of research on long-term effects

Even though math interventions, presumably, are created to make a lasting difference, surprisingly few studies have actually investigated the long-term outcomes of math programs, especially when counting studies using a high quality design. Furthermore, no ed-tech studies measuring long-term outcomes were found when searching the literature. Consequently all subsequent studies are on non-technological interventions. Accordingly, Dowker (2017) noted that very long outcomes are not evaluated in most intervention studies in mathematics.

The little evidence that exist point to a fadeout of effects after training has ended, with control students catching up to students in the treatment group. Clements, Sarama, Wolfe, and Spitler (2013) found that initial effects after training faded out as children were followed across three years (the first year being the intervention phase), from pre-Kindergarten to first-grade. Furthermore, Bailey et al (2016) conducted another analysis on this sample and found that 72 % of the fadeout effects were attributable to preexisting differences between students in the intervention and control groups with the same performance level after training. An intervention study by Smith et al. (2013) reported similar fadeout effects when following students from the beginning of first grade, when the intervention phase took place, to the end of second grade. As noted earlier, these two studies were both non-technological.

Interestingly, to the best of our knowledge there are no studies on older children that measure effects beyond the end of the intervention using high methodological standards (e.g., independent measures, large samples, randomized design). If long-term effects are investigated, follow-ups typically take place within 4-9 weeks for these older students (e.g., Codding, Chan-Iannetta, Palmer, & Lukito, 2009; Jitendra et al., 2015).

The same evidence of fadeout effects has been found when looking at the broader spectrum of interventions aimed at increasing cognitive skills, socio-emotional skills, or behaviors among child and adolescents (Bailey, Duncan, Odgers, & Yu, 2017; U.S. Department of Health and Human Services, 2010), Although one overview demonstrated that effects may persist at close to full strength for 1-2 years beyond the termination of the programs, a fadeout effect was demonstrated after that (Leak et al., 2010).

In essence, most math intervention studies do not investigate long-term effects, and the few studies that do give evidence of a fadeout effect after training has ended when followed up to second grade. No studies on long-term effects of educational technology programs have been documented so far. After this exposé of different types of interventions and their effects, the focus will now turn to the effects of interventions for students with low math performance, since these students are in focus of the current investigation.

Interventions targeted at students with low math performance

Some interventions aimed at helping at-risk children have already been described under the heading Individual or small group instruction. The purpose of this section is to present additional meta-analytic findings on what intervention elements are effective for students with low math performance (whereof at-risk children is a part).

Chodura, Kuhn, and Holling (2015) conducted a meta-analysis on interventions for children with mathematical difficulties that included children diagnosed as dyscalculic or learning-disabled, as well as at-risk dyscalculic children (defined as having a percentile rank below the 26th percentile). The 35 studies enclosed children 6-12 years old, and the interventions could be technology based or non-technological. The average effect size reported was 0.85 using Hedges g, which could be interpreted as Cohen's *d* but also more conservatively according to some researchers, plausibly indicating a large effect. Yet, the findings in this meta-analysis should be viewed with caution because 76 % of the variance could be attributed to differences between studies, and children with differing levels of mathematical difficulties were treated as one population.

Bearing this important limitation in mind, it might still be interesting to consider that effective interventions for children with mathematical difficulties were characterized by direct instruction (compared to self instruction and strategy instruction), a focus on basic arithmetic competencies, being computer based, and utilizing a single-subject setting (Chodura, Kuhn, & Holling, 2015). Direct instruction, as interpreted in the meta-analysis just referenced, is used in the same sense as explicit instruction, which is a method where the teacher provides clear models for solving a class of problems utilizing a set of examples and non-examples on how to solve this type of problems. Newly acquired strategies and skills are extensively practiced, and students are provided with opportunities to think aloud in order to talk through the decision they make. Students receive extensive feedback throughout the practice. Learning disabled students, students with mathematical difficulties, and students who perform in the lowest third of a typical class, have consistently demonstrated positive outcomes on achievement with computation and word problems (National Mathematics Panel, 2008).

Furthermore, evidence suggests that age at intervention has little effect on level of improvement for children with mathematical difficulties, at least within the primary school age range (Chodura, Kuhn, & Holling, 2015; Dowker, 2017). This proposes that there is no critical period for introducing interventions, in line with evidence that age of starting school has little influence on the final outcome (Mullis, Martin, Foy, & Arora, 2012).

The above description suggests that math interventions are effective for students with low performance, although the magnitude of the effects may be somewhat uncertain. The most effective interventions are characterized by a focus on basic math arithmetic competencies, direct or explicit instruction, computer based instruction, and teaching in a single-subject setting. The investigations in this thesis utilize most of these characteristics to teach primarily math facts, and secondarily problem solving and prerequisite skills to math facts. The subsequent text will thus provide a more comprehensive description of interventions targeted at math facts.

Interventions teaching math fact

Several studies have demonstrated the importance of timed practice when acquiring math facts using a non-technological approach. Fuchs et al. (2013) isolated the effects of speeded practice and found that at-risk students receiving timed practice made ground in catching up to not-at-risk classmates, while students receiving untimed training did not. The authors concluded that timed practice plays a substantial role in learning math facts.

Additional evidence has been accumulated by behavior analytical studies, often using detailed analysis on smaller samples, on a technique called Explicit Timing where students are provided with a set of problems, usually in paper form. A tutor or a teacher gives the student a time interval to finish the given tasks and then times the student who tries to complete as many problems as possible (Codding et al., 2007; Rhymer, Skinner, Henington, D´Reaux, & Sims, 1998; Rhymer, Henington, Skinner, & Looby, 1999; Rhymer & Morgan, 2005, van Houten & Thompson, 1976). Following these findings Gersten et al (2009) recommended that 10 minutes in each session should be devoted to building fluent retrieval of basic arithmetic facts (including addition, subtraction, multiplication, and division) at all grade levels. A more comprehensive, and highly structured, behavioral based intervention that also utilizes timed practice is the Morningside Math Facts: Multiplication and Division curriculum (Johnson, 2008). An investigation by McTiernan, Holloway, Healy, and Hogan (2016), including 28 students, demonstrated that 9-11 year old students receiving the Morningside curriculum increased on fluency outcomes compared to non-fluency based teaching. The students also showed improvement on measures of endurance and stability, as well as on a subtest of the Wechsler Individual Achievement test of mathematical ability.

Effective learning of math facts has also included goal setting (Gross et al., 2014), reward (Freeland & Noell, 1999), and immediate feedback (Duhon, House, Hastings, Poncy, & Solomon, 2015).

Immediate feedback is a corner stone in another behavioral analytic technique called Cover Copy Compare that has been applied to various subjects, including math facts training. In Cover Copy Compare (CCC) students are taught to view a problem and an answer on the left half of a paper, cover up the problem and answer, write the problem and answer on the right half of the paper, and uncover to see if their answer was correct (Joseph et al., 2012). Joseph et al. (2012) conducted a relatively small meta-analytical investigation on CCC, using mostly single-case studies on students with some kind of impairment (e.g., ADHD, or mental retardation). The authors concluded that CCC improved academic performance across subjects, especially when combined with other evidence-based techniques such as goal setting, rewards, and opportunities to respond. Codding, Chan-Iannetta, Palmer, and Lukito (2009) found that a combination of CCC and goal setting, with a specified number of problems correct, led to steeper slopes and better final performance on math facts compared to most other treatment conditions, with retained effects at one month follow-up.

An interesting comparison between Explicit Timing (i.e., including timed practice) and Cover Copy Compare (i.e., including untimed practice) was made by Codding et al. (2007). Ninety-eight second and third graders were randomly assigned to control, ET, or a CCC condition. Both training conditions focused on single digit subtraction using the training methods described above. Training occurred twice weekly across six weeks during recess time, the duration of training was only 5 minutes for each training condition, groups had a mean size of nine students, and the two treatment conditions attended 10 sessions on average. Before each session progress monitoring, measuring digits correct per minute respectively digits incorrect per minute, was used to assess level of fluency.

Results showed that students initial level of fluency influenced intervention effectiveness. For students with relatively low levels of fluency (framed as being in the frustrational range) Cover Copy Compare was most effective, whereas students with relatively high levels of fluency (being in the instructional range) benefited more from Explicit Timing. This points to the importance of untimed practice in initial training on math facts when fluency levels are low, and the value of timed practice once higher fluency have been acquired. However, these results should be interpreted with caution, because the study suffered from several limitation: only students from one elementary school were included, duration of training was short which could increase the risk for spurious results (Cheung & Slavin, 2013), and the attrition was 31 %. Despite these shortcomings, the study by Codding et al. (2007) provides some valuable initial empirical guidance for how to best tailor training with respect to level of fluency.

To sum up, the studies presented above suggest that several non-technological techniques and interventions targeted at students with low math performance have shown effects in improving math facts skills. Critical program components are immediate feedback without timing for students with low levels of math fact fluency, and timed practice for students with higher levels of fluency. A more comprehensive curriculum, including timed practice, possible enhances not only fluency, but also endurance and stability. These findings will be somewhat more elaborated in the next section where the intervention program in the current thesis will be described, along with the theoretical foundations on which the program was built on.

Intervention program: description, development, and theoretical foundations

Program description

The intervention program, built and evaluated in Study II and Study III within this thesis, is called Chasing Planets (Planetjakten in Swedish). The primary goal of the program is to teach addition and subtraction facts (i.e., math facts) up to 12, with a maximum of one double-digit number per math fact. The secondary focus of the program is to teach prerequisite skills to math facts and word problem solving including math facts.

The motivation to create a new math program was that the combination of behavior analysis, Relational Frame Theory (De Houwer, 2013), instructional design (Tiemann & Markle, 1991) and Precision Teaching (i.e., teaching students to a set mastery criterion; Johnson & Street, 2012), had not been tested before in mathematics, especially not in combination with modern technology. However, the reading program HeadSprout had shown that these theories and methods could be combined successfully when delivered on computer (Huffstetter, King, Onwuegbuzie, Schneider, & Powell-Smith, 2010). The many paper-based programs of Direct Instruction had also shown that Precision Teaching and instructional design could be combined fruitfully when teaching mathematics (Gary & Engelmann, 1996). Altogether, these various inputs made an argument that a math program based on these methods and delivery format could prove viable.

Chasing Planets is an application on iPad with a graphical interface made up of 261 planets on a space map. Each planet has unique math exercise which has to be completed in order to unlock the next planet. In the first session students design their own avatar and then start on the first planet. Every planet included at least one of three phases used in all training: the modeling, guide, and fluency phase. These three distinct phases of learning is a well-established sequence that has been utilized in explicit instruction (Archer & Hughes, 2011), and the Morningside Model of Generative Instruction (Johnson & Layng, 1994).

In the model phase, a new skill or concept was introduced using audial instructions along with animations and hand-drawn examples that included frequent student responding. One example of a model phase was when the student was instructed on how to add 1 on the number line. A number line from 0 to 10 would appear, and the student would be instructed by a voice to tap 6 on the number line. After responding correctly (or receiving a prompt if responding incorrectly), 6 was circled. A hand drawn line then started from 6 and ended on 7, while the voice simultaneously said "Now I will jump ooone on the number line. Where did I land? Tap the number!". This sequence would then be repeated with a few more numbers. This type of responding only involves previously acquired skills, which is typical for the model phase.

The guide phase aimed at helping students acquire the just modeled skill, with the ultimate goal of achieving high levels of accuracy, using untimed practice and scaffolding (i.e., giving new instructions tailored to specific errors committed). Using the example with adding 1 on the number line, the student would practice multiple tasks where the student had to add 1 on the number line, with gradually fading levels of scaffolding.

Lastly, the fluency phase involved timed practice and aimed at developing proficiency, or fluency, with the skill at hand by using a set frequency aim and immediate feedback. If the student was to build fluency on the skill adding 1 on the number line, having the student avatar race against a computer avatar would do this, with the goal of completing as many corrects as required to win over the avatar in a specific race. An algorithm that took previous responses into consideration set the frequency aim for each race.

As might have been recognized, the guide phase shares the features of untimed practice and immediate feedback with the Cover Copy Compare technique described above, which was found to be more effective for learners with emerging levels of fluency. Likewise, the fluency phase shares the timed practice characteristic with the Explicit Timing technique, also presented earlier in the text, which was found to be more benevolent for students with relatively high levels of fluency.

Gamified elements were used during the guide and fluency phases. In the fluency phase, for instance, the students´ avatar competed against a computer avatar. When the student avatar won the race, a new race against another opponent was initiated. An algorithm that took the outcome of the preceding race into consideration set the difficulty level of the new opponent. Furthermore, the student avatar received stars for correct answers during training which were later used to purchase edibles to feed a so called Knowledge Monster which surfaced at the end of the daily practice.

The primary goal of the program was to teach math facts. Math facts were trained as number families, for instance the numbers 257 were taught as a family with two addition facts (2+5=7 and 5+2=7) and two subtraction facts (7-2=5 and 7-5=2). The training of number families involved multiple components including the three phases (i.e., modeling, guide and fluency phases) of training. First, a model phase where a new family was introduced, an untimed guide or scaffolding phase with the goal of achieving a high percent corrects, and lastly a timed fluency practice with a set frequency aim.

The secondary goals of the program were teaching of prerequisite skills to math facts and word problem solving including math facts. Some examples of the prerequisite components were: rapid tapping to facilitate swift responding on math facts, matching number to text (e.g., see eight, tap 8), counting amounts from 3 to 20, and discriminating among smaller and bigger numbers ranging from 3 to 15. Word problem solving skills were: labeling problem parts, turning words to number problems, and vocabulary (e.g., at a very early level equating "take away" with "minus" or "-").

Program development

The program was created using a behavior analytic framework originally developed by Tiemann and Markle (1991) and later refined by Twyman, Layng, Stikeleather, and Hobbins (2004) when creating the online programs Headsprout Early Reading and Headsprout Reading Comprehension. The programming process can be summarized as follow (Twyman et al, 2004):

1. Analyze the content to be taught
2. Decide what the critical instructional objectives are
3. Determine the criterion tests (i.e., set explicit measurable goals for each objective)
4. Define the required behaviors to enter the program
5. Build the instructional sequence
6. Test the instructional sequence and continually adjust the instructional sequence (5) unit it meets the objectives (2)
7. Build additional components that maintain the progress made

Although our initial intention was to follow this programming process closely, lack of resources and time constraints made compromises necessary. The initial content analysis (1) was that math facts is a critical skill in mathematical development, and that problem solving using math facts is an important application of this skill. When deciding what the critical instructional objectives were (2), we recognized that math facts skills could be divided into several composite skills (e.g., discriminating between smaller and larger numbers) that could be further decomposed to component skills were the instructional objectives became more precise (e.g., student sees "15" and "12", then hears "Which one is bigger?" and taps "15"). When determining the criterion test (3), we specified what fluency rates the student had to master for each instructional objective. The required behaviors to enter the program (4) were to be able to count from 1 to 20 and a few other skills, although in retrospect these requirements were probably far too liberal.

Designing the instructional sequence (5) and continually testing and revising it (6) was a major challenge in the whole process of building Chasing Planets. Several iterations were made with students to try out various instructional sequences (e.g., how to model a number family effectively, or how to present multiple math facts simultaneously without distracting side effects). Even more subtle features, such as deciding the optimal size of number elements at the bottom of the screen that students used for giving a response, required attention as well. The building of components that maintained the progress (7) was not touched upon at all, due to limited resources. In summary, the programming process outlined by Twyman et al (2004) served as useful guidance when developing Chasing Planets. However, many of the viable suggestions in this model were barely used, or not utilized at all, because of financial constraints.

We will now move from the description of the overall method used to develop Chasing Planets to how the behavior analytical perspective was used when developing specific exercises. A non-linear approach was used when creating tasks presented in a specific model, guide or fluency phase. "Non-linear" is a term coined by Goldiamond (1974; 1979; 1984) meaning that a target behavior is not only determined by the occasion and past consequences for the target behavior, but also on the occasion and past consequences of alternative behaviors. For instance, hearing the question "Is 5+2 an addition fact?" (occasion) and answering "Yes" (target behavior) may not only be a function of having previously been reinforced for this target behavior. It can also be a function of a learning history where hearing "Is 5-2 an addition fact?" (occasion) and answering "No" (alternative behavior) has been reinforced. Thus, the presently observed consequences may not even exert the greatest control over the target behavior, but instead the previous consequences of an alternative behavior may be the dominant influence on the target behavior. The instructions in the Headsprout reading programs build on these multiple sources of control over behavior, or non-linearity. A systematic analysis was not only conducted on the immediately or directly visible contingencies of the target behavior, but also on the contingences for matrices of alternative behaviors that were part of the same skill set (Twyman et al, 2004).

In the same way Chasing Planets aimed at building exercises using a non-linear analysis when teaching problem solving skills and the concept of number families. When teaching very basic word problem solving skills, this was done, for example, by reinforcing students for answering correctly to target behaviors (e.g., seeing "take away", and selecting "-" from an array with "+ - =") as well as reinforcing students for answering alternative behaviors correctly (e.g., seeing "get more", and selecting "+" from an array with "- = +"). It should be noted though, that while Headsprout very systematically investigated contingencies for target behaviors and matrices of contingencies of alternative behaviors, such analyses were done in a much more simplified way in the current context due to lesser resources.

Although this section set out to describe how Chasing Planets was developed, some theoretical reasoning was brought into the description. However, the next section will provide a more thorough account of the theoretical foundations for this thesis and how it relates math training.

Theoretical foundation

Mutual and combinatorial entailment

Behavior analysis and Relation Frame Theory (RFT) assume there are unique human capacities, related to math performance, that can be influenced by training of certain types and thus provide individual variability (Hayes, Barnes-Holmes, & Roche, 2001).

At the core of these capacities are the behaviors of mutual entailment and combinatorial entailment. Mutual entailment is manifested in infancy in typically developing children, and means that in a given context, if stimulus A is related in a characteristic way to B, as a result B is now related in another characteristic way to A (De Houwer, 2013). When a child is in a room (context), watching a lamp (stimulus A) and the parent says "Lamp" (stimulus B), mutual entailment is present when the child enters another room (context), hears the word "lamp" (stimulus B) and then looks at the lamp (stimulus A). Combinatorial entailment (sometimes called equivalence) is present in a given context, if A is related in a characteristic way to B, and A is related to C, as a result a relation between B and C is now entailed. For a typical developing child this would occur at about one year of age. A child who learned that a cat (stimulus A) is the same as a simple drawing of a cat (stimulus B), and that the drawing and the sound "cat" (stimulus C) are also equivalent, would derive that the cat (stimulus A) and the sound "cat" (stimulus C) are equivalent (De Houwer, 2013). A mathematical example of a derived stimulus relation (or a mutual and combinatorial entailment) is apparent when a child learns to use numbers. The child has learned that the sound "seven" and the symbol "7" are equivalent, and that the sound "seven" and seven objects are equivalent (i.e., the child can identify seven things regardless of type of object). Typically developing children will most likely derive, without direct training, that the symbol "7" and seven objects are equivalent. Deriving stimulus relation is a very effective way to learn language and numeracy skills since many relations do not need to be trained directly.

Mutual entailment and combinatorial entailment are, together with the term transformation of stimulus function (not explained here, please read Hayes, Barnes-Holmes, & Roche, 2001), defining features of a relational frame and thus being at the core of RFT.

Mutual and combinatorial entailment explain why language acquisition is so rapid among typically developing children, because children, after deriving a few initial stimulus relations, learn how to derive at faster a pace and the gradual accumulation of previous derivations further facilitates new derivations. Instead of using cognitive constructs, RFT uses this basic terminology to describe and understand the unique capacities of humans, although with several additional and important extensions which are beyond the scope of the current text (for in depth explanations, please read Hayes, Barnes-Holmes, & Roche, 2001; De Houwer, 2013). Even if mutual and combinatorial entailment can be found to some extent in other species, other species are far less successful in using these capacities.

In the math training we did not wait for students to derive, but tried to help them derive relations so that they could learn how to derive and create useful rules that would guide later behavior. The assumption was that by systematically, and over time, providing many opportunities of frequent responding and immediate feedback, the relations between stimuli and the derivation of stimulus relations would be strengthened.

In summation, according to the just referenced literature, the ability to derive stimulus relations is the basis for learning mathematical skills, and generalized operants altogether.

The importance of math fact from a RFT perspective

From a theoretical point of view one may critique the importance of training math fact, which was the primary target of the intervention in this thesis, as simply training rote memory. However, both from an empirical perspective and a RFT perspective training math fact would not be considered trivial. From an empirical point of view, fast retrieval of math facts is a crucial component skill for later, more complex skills (i.e., algebra), as stated in the bottleneck hypothesis, which has strong empirical support (e.g., Jordan et al., 2003; Vasilyeva et al., 2015). Learning math facts could possibly be compared to the ability of identifying and using phonemes in reading, which is a necessary component skill when learning to read. Yet, from a RFT perspective, fluently using math fact entails more complex learning than phonem identification, which is purely arbitrarily in its nature and dependent on context (i.e., meaning that when context/language changes, the same phoneme can have a completely different function). In contrast, when training math fact fluency behaviors differential reinforcement is provided for a large network of stimulus relations, both arbitrary and non-arbitrary.

To further explain what is meant by arbitrary and non-arbitrary an example will be provided. It is arbitrary that the sound "nine" is the same as the symbol "9", but it is not arbitrary, in any context or language, that yyyy (4) objects and yyyyy (5) objects put together constitutes yyyyyyyyy (9) objects because these stimuli can have physical counterparts. Moreover, the math fact training in the intervention study not only involved correctly answering math fact, but also training on prerequisite skills to math facts by explaining and reinforcing correct responding on basic relations between arabic symbols (e.g., 9), sounds (e.g., "nine"), text (e.g., nine), quantities (e.g., nine apples), and number lines (e.g., placing nine on a 0-10 number line). The relationships between arabic symbols, sounds and text are arbitrary, but when relations between symbols, sounds and text, on the one hand, and quantities and number line estimations, on the other hand, are introduced non-arbitrary stimulus relations are also entailed. The point is that the current math fact training not simply taught students arbitrary relationships such as seeing two stimuli and respond by selecting a third stimulus (and math fact teaching as a concept seldom does in other studies as well). Instead, math fact training involves strengthening arbitrary and non-arbitrary relational responding in the mathematics domain in particular, and derived stimulus relations in general.

The program also tried to vary the contexts, because RFT views the context as an essential influence on behavior. Exercises were created to make students practice mutual and combinatorial entailment in frames (or contexts) of coordination (e.g., the sound "five" is the same as 5, or the numbers 3, 5 and 8 are numbers that go together, i.e., a number family), discrimination (e.g., between - and +), comparison (e.g., 16 is bigger than 15), and temporal relations (e.g., word problems containing a story with temporal elements that are to be translated correctly into a math fact).

Arguably, on a daily basis, speaking, reading, and writing behaviors are more frequently reinforced than mathematics behaviors. The reason is that there is much more socially mediated reinforcement available for these behaviors, than for engaging in math behaviors. Furthermore, many contexts provide corrective cues for speaking, reading and writing. If a person is writing a sentence with the words in the wrong order, it is likely that the same person, when reading the sentence, will feel that something is not correct (i.e., a cue). This would most likely not be the case in mathematics, were corrective cues are much less common in everyday situations.

Because of this relative lack of reinforcement and corrective feedback from the environment for math behaviors, it is of particular importance to ensure that the basic relations, such as the relations entailed in math facts, are manifest in the behavioral repertoire. Given that the basic relations are in place, these relations will become cues for later mathematics learning where it is necessary to judge whether a solution to a problem is reasonable or not. For instance, if you know that the basic relations in a triangle you will know that two sides are always larger than the third side. If a solution suggests that the third side is indeed larger than the other two, this will be a cue to try out another solution. Accordingly, basic relations need to be automatized, especially in mathematics, because the environment usually does not provide sufficient help to correct inadequate behaviors.

This section has tried to describe the intervention evaluated in this thesis and the theoretical foundations for the program. The program is a tablet based application aimed primarily at teaching math facts, and secondarily at teaching prerequisite skills to math facts and word problem solving including math facts. The exercises are built on an instructional design model using recurrent model, guide and fluency phases. The program also comprises gamified elements. A behavior analytical framework served as the theoretical foundation in the process of building the overall structure and the specific exercises in the program. Derivation of stimulus relations, as described in Relational Frame Theory, was utilized to some extent when teaching basic number relations. The next section will briefly address how the assessment was conducted and validated on a tablet (Study I) in the intervention study (Study II and Study III).

Assessment on a tablet

To facilitate screening and assessment in the intervention study almost all tests were conducted on a tablet. However, because no existing mathematical tests had been standardized on tablet, this posed several challenges. Most importantly, this required an investigation of the comparability (International Test Commission, 2006) between a standardized paper-based tests and the same test delivered in a new format (i.e., tablet), which was investigated in Study I. The test of comparability between formats was conducted with the, on paper, standardized mathematics test Heidelberger Rechen Test 1-4 (Haffner, Baro, Parzer, & Resch, 2005). Some additional questions regarding testing on a tablet were also investigated in Study I: potential administrator effects (important when comparing results from teacher led screening, and assessment led by research assistants), and psychometric differences in terms of reliability for various populations and tests. The next section will briefly describe the objectives of this thesis.

Objectives of the thesis

The main objective and contribution of this thesis was to investigate the short- and long-term effects of a math intervention on a tablet aimed at helping students with low math performance to improve critical math skills (Study II). An aligned objective was to identify characteristics of students demonstrating a positive long-term response to intervention (Study III). Additional contributions were to provide knowledge on whether a validated mathematical paper tests can be translated into tablets without compromising psychometric properties, as well as to provide Swedish norms for math tests on tablet (Study I).

Study I - validating tests on a tablet

The first four hypotheses/questions investigated the comparability between the paper and the tablet version of HRT (International Test Commission, 2006):

1. The test-retest correlations for the tablet version of the Heidelberger Rechen Test (HRT)-scales Addition, Subtraction, Missing Term and Count Amount would be comparable to the test-retest correlations for the paper version.
2. The correlation between the paper and tablet format would be at the same level as the test-retest reliability for each format.
3. The correlations, or convergent validity, between the tablet version of the HRT-scales Addition, Subtraction and Missing Term and the tablet version of the four subtests on the Grade 3 Math Battery (Math Battery; Fuchs, Hamlet, & Powell, 2003), would be comparable to the correlations between the paper version of HRT and the DEMAT 4 test (Gölitz, Roick, & Hasselhorn, 2006).
4. Are the absolute measures of central tendencies (mean and standard deviation) for both formats comparable for HRT Addition, Subtraction, Missing Term and Count Amount?

The following research questions examined administrator effects for tests on a tablet, and psychometric differences in terms of reliability for various populations and tests:

5. Does the type of test administrator (i.e., test personnel or teacher) have an effect on the test-retest reliability for tablet tests?
6. Are the test-retest reliabilities on tablet tests (HRT and Math Battery) comparable between a sample of students with lower math performance and more representative samples?
7. What was the test-retest reliability for the Math Battery on tablet?

Study II - investigating effects of an intervention

The following hypotheses were tested to investigate the short- and long-term effects of training mathematics on a tablet, moderating effects, effects of a combination of mathematics and working memory training, and if improvement in mathematics performance had a positive effect on mental health:

8. Low performing children in second grade participating in math training improve mathematical skills compared to children in control and placebo conditions.
9. The gained effects will be maintained during the follow-up period.
10. Intelligence and socio-economic status moderate the effect of math training.
11. Working memory training in combination with math training lead to more superior outcome in terms of mathematical skills.
12. Improvements in mathematics achievement will show a positive effect on mental health.

Study III - predicting long-term response to intervention

Two research questions were examined in order to identify factors that predicted long-term response to intervention, and how these factors were interrelated:

13. Which factors significantly predict positive long-term outcomes in a mathematics tablet intervention for second graders with low math performance?
14. How are these factors related to each other?

Methods

Design

Study I

Test-retest reliability for HRT and the Math Battery on tablet was investigated by assessing three samples of students twice on tablet in various contexts. Correlation between paper and tablet was examined by testing a sample of children on both paper and tablet versions of HRT utilizing a counterbalanced design to avoid possible ordering effects. Convergent validity was investigated by analyzing the correlation between the Math Battery and HRT on a sample.

Study II

An experimental design with longitudinal data gathered at baseline, post assessment, 6 and 12 months follow-up was conducted. Participants were individually randomized to one of four conditions (independent variable): passive control, reading (placebo), mathematics, or mathematics combined with working memory training. Outcome (dependent variable) of the intervention was test scores on mathematics test.

Study III

Predictors of long-term positive response for students in the math group and math combined with working memory training group were examined. Because these groups presented with similar results in study II, they were collapsed in both study II and III.

Participants

Study I

Five samples were investigated. Some additional details on demographic, math performance at the included schools, and attrition are reported in the papers on these studies.

The Test-retest sample 1 originally consisted of 152 participants in second grade recruited from five schools. The schools covered a spectrum of low and high levels of diverse ethnic background, and low, medium and high parental educational levels. Twenty-two children were excluded because they were absent on one of the assessments. Fifty-three of the remaining 130 children were female. The mean age was 8.7 years (*SD=0.*3), ranging from 8.0 years to 9.5 years.

The Test-retest sample 2: different administrators initially consisted of 164 participants from seven schools that represented low to high levels of diverse ethnic background, and a rather even split between number of students with low respectively high parental educational levels. Absence at assessment or technical problems reduced the sample with 27 children to a total of 137. Seventy-eight of the remaining 130 children were female. The average age was 8.8 years (*SD=0.*3), ranging from 8.0 years to 9.5 years.

The Test-retest sample 3: same administrators in small groups with low-performing students was the intervention sample (further described below under Participants in Study II).

The Different formats sample at first consisted of 124 children in second grade. Children were recruited from four schools covering low and high levels of diverse ethnic background respectively parent education. Ninety-four children (41 female) remained after excluding 28 children, the major reason being absence. The age was 9.0 years (*SD*=0.4) on average with a range from 7.4 to 10.2 years.

The Normative sample (testing convergent validity) was made up of 265 second-grade and 289 third-grade students coming from sevens schools located in privileged and socially disadvantaged areas in two smaller cities and one larger urban area in Sweden. Approximately 62 % of the children had parents with post high school background, compared to the national average of 56 %. The proportion of participants with diverse ethnic background was estimated to 41 %, compared to the national average of 21 %. A similar share (88 %) of the children in third grade passed on the four arithmetic scales on the Swedish National Tests in Mathematics, compared to the national mean (89 %).

Study II

The intervention study included 283 second-graders coming from two smaller cities, Landskrona and Helsingborg, (140 participants) and the capital (143 participants) of Sweden. The children were spread across 87 classes in 27 schools, situated in both privileged and socially disadvantaged areas. In general, there was a distinct divide between schools with relatively low shares of students with diverse ethnic background and high levels of parent education (privileged areas), and schools with high shares of students with diverse ethnic background and low parental educational levels (socially disadvantaged areas). Children in this sample had a similar parental educational level as the national average: 62 % of parents in the sample had a post high school education, compared to 56 % for the general population. Although no individual data on ethnic background was collected, a proxy for this could be primarily language spoken at home, where 20 % of the students did not primarily speak Swedish at home compared to the national average of 21 % of elementary school children having a diverse ethnic background. Detailed additional demographics, and some other variables, on this sample are described in the paper on study II. The average age was 8 years and 3 months (*SD*=4 months). Of the 283 children, 142 were female.

Study III

The analysis of predictors of positive long-term response to the intervention was conducted on children in the math group and the combined math and working-memory training group, which had been collapsed (as described under Design). Thus, the study included 153 children (81 female) spread across 73 classes in 26 schools. The age was on average 8 years and 3 months *(SD*= 4 months) for this sample.

Instrument

All tests, except for Ravens and the paper version of Heidelberger Rechen Test 1-4, were conducted on tablet. A higher score indicated a better performance on all measures. To keep this section reasonably short, full information on psychometric properties and some additional details on measures are not provided here, but in the papers for study I-III.

Heidelberger Rechen Test 1-4 (HRT) was used in study I-III. HRT comprises eleven subscales measuring basic math skills across grade one to four (Haffner, Baro, Parzer, & Resch, 2005). The scales Addition, Subtraction (both covering up to three-digit numbers), Missing Term (covering tasks such as 4+_=9 and _-32=9+50), Count Amount, and the control variable Speed were used in study I and II (where Addition and Subtraction were both used as screening and outcome measures). Addition, Subtraction, and Missing Term were utilized in study III. Haffner et al (2005) reported that criterion validity for the total scale (including all eleven subscales) of the original paper version was .67-.68 with math grades. Alpha for the intervention sample at baseline (study II) on the tablet version was .90 for Addition, .87 for Subtraction, and .60 for Missing Term.

The Grade 3 Math Battery (Math Battery*)* was used in study I-III. Four subtests, each with duration of one minute, were utilized to cover basic math facts: Addition 0-12, Addition 0-18, Subtraction 0-12, and Subtraction 0-18 (Fuchs, Hamlett, & Powell, 2003). Fuchs, Hamlett, and Powell reported that criterion validity for the Math Battery was .51-.53 with the total math score on Terra Nova for the original paper version. Alpha for the intervention sample at baseline was .83 for Addition; .75 for Addition 0-18; .70 for Subtraction, and .66 for Subtraction 0-18.

Arithmetic Composite (study III) was an index variable on arithmetic achievement computed by taking an average of z-values at pretest for seven arithmetic outcomes: the HRT-scales Addition, Subtraction, and Missing Term, and the four Math Battery scales.

Diamant AG1 consisted of twelve addition and twelve subtraction items and was used as part of the screening in study II (Swedish National Agency for Education, 2013).

Klurisen was an item response theory (IRT) scaled problem-solving measure constructed by the authors and used as outcome in study II. It comprised 15 items, starting from easy one-step (i.e., having to perform one calculation) addition-subtraction word problems, and incrementally moving to more difficult two-step problems (i.e., requiring two calculations in a specific order to solve the problem). Internal consistency was found to be .83 using Person Separation Index for a sample covering grades 1-5 (N=403). Test-retest reliability for students beginning second grade (N=243) was r=.72, and concurrent validity with the subscale Math Problem Solving of the Iowa Test of Basic Skills was r=.59-.67 (Hallström Elthammar, 2015).

Mathematical word problem-solving test was used in the screening (study II). The test includes 16 one-step and two-step word problems (Jitendra et al, 2007). The same authors reported that the original paper version of the test had an internal consistency of .84-.86, and a concurrent validity of .64-.71 with the SAT Mathematics: Problem Solving, and .51-.57 with the SAT-9 Mathematics: Procedures.

Ravens' Standard Progressive Matrices (Ravens) was used in study II to assess non-verbal intelligence. Criterion validity has been reported to be .98-1.0 in several studies, and predictive validity of academic achievement ranges up to .70 for non-English and English speaking children and adolescents (Raven, Raven, & Court, 2000).

Grid and Digit-Span backwards measured working memory and were used in study II to check manipulation of working memory training. Grid was based on Automated Working Memory Assessment (Alloway, 2007), and showed a grid of dots (4 x 4) that lit up in a specific order that the child had to repeat. Digit-span backwards was based on WISC-III (Wcchsler, 1991), and displayed a sequence of numbers that were then repeated by the child in reversed order. Test-retest correlation, with a 2 year interval, was r=.75 for Grid, and r=.74 for Digit-Span Backwards (Ullman, Almeida, & Klingberg, 2014).

Procedure

Study I

Teachers and test administrators in all samples followed test protocols containing explicit instructions to be read verbatim and instruction on principles to be followed during testing. Detailed information on procedures (e.g., intervals between test and retest, time point of testing) in the five samples is provided in the paper on study I.

Study II

Teachers administered the screening tests to their classes. Research assistants (university students) conducted all assessments (pretest to 12 months follow-up) in small groups with 1-5 students per group.

The Consort Flow diagram in the paper on study II presents the recruitment, randomization, and attrition rates. The attrition rate for the whole sample was 14.5 % from baseline to 12 months follow-up. The attrition for each condition from baseline to 12 months follow-up were: Control 5.8 %, Placebo 11.5 %, mathematics 15.8 %, and math and working memory training 22.1 %. Additional information on reasons for leaving the study, non-attendance during training, cohorts, test administration, description of conditions, and a detailed account of the implementation of the program and fidelity are provided in the paper on study II.

Students in the reading (placebo) group practiced primarily reading, and to a lesser extent motoric skills on tablet for 20 minutes each day during the intervention. Participants in the mathematics training condition practiced Chasing Planets for 20 minutes each day. Students in the mathematics combined with working memory training group trained working memory (Nemmi et al., 2016) 10 minutes each day, adjacent to the 20 minutes of math training, thus adding up to a total of 30 minutes combined training per day. The math and working memory training were integrated in the same system, requiring no additional login.

The working memory training consisted of four different exercises. Grid and Digit-Span Backwards were the same as the measures described earlier with the additional feature of feedback on each response, adaptive adjustment of the level according to the student's performance and systematic alteration between exercises. The third working memory task was a rotating circle consisting of dots that lit up in a specific sequence that were to be repeated by the child. The fourth task was a three-dimensional cube consisting of a 3X3 grid on each side. The cube was slightly rotated as different parts on various sides lit up, which were then to be repeated by the child.

Study III

The same procedure applied as described above for study II. A response to intervention was defined as a child performing above the average improvement for the collapsed control/placebo group on the Arithmetic Composite. To count as a response at 6 months follow-up, the child had to perform above the control/placebo group at posttest and at 6 months follow-up. To count as a response at 12 months follow-up, the child had to perform above the average at posttest, 6 and 12 months follow-ups.

Statistical Analysis

Analyses of similarity between test formats (paper versus tablet tests), test-retest reliability, and convergent validity was calculated with Pearson correlation. Intervention effects were estimated with multilevel methods (Heck, Thomas & Tabata, 2014).

Data mining and machine learning

Data mining, and in particular machine learning techniques, were used to identify students with positive long-term outcome to the math intervention. Data mining is usually not viewed as a set of specific methods, but rather a process of selecting, exploring and modeling large amounts of data in order to identify patterns that reveal meaningful information (Guidici & Figini, 2009). A number of different approaches gathered from statistics (e.g., regression, clustering, classification, and hypothesis testing) and computer science (e.g., machine learning, multidimensional data bases, and data visualization) are often used in data mining. Predictive data mining is an established field within clinical medicine where the goal is to derive models that can support clinical decision-making by analyzing large data sets containing patient information (Bellazzi & Zupan, 2008).

Machine learning is a frequently utilized technique within predictive data mining that has been valuable when predicting outcomes (e.g., Demichelis, Magni, Piergiorgi, Rubin, & Bellazzi, 2006; Lee et al, 2009). Machine learning utilizes different types of algorithms that have been categorized into five categories by Domingos (2015): symbolists (e.g., decision tree induction), evolutionaries, connectionist (e.g., neural networks), bayesian (e.g., naive bayes) and analogizers (e.g., k-nearest neighbor). Specific techniques are also used to identify relevant predictors in a large set of potential predictor variables, such as filtering techniques and data transformation.

Overall, the greatest threat to a machine learning analysis is the risk of overfitting the model. One crucial way of avoiding overfitting is to never test a model on the same data as the model was built on. Ideally, two separate data sets should be used where the test set is utilized to develop the model, and the validation set is used to confirm the model (Witten, Frank, Hall, & Pal, 2017). This is similar to how research findings are considered to be more robust when replicated in a separate study. However, data is often sparse and a common technique within machine learning is cross-validation where data is partitioned data into separate folds that are then used as separate test and validation sets (Witten, Frank, Hall, & Pal, 2017).

Following the guidelines suggested and used for data mining and predictive modeling (Bellazzi & Zupan, 2008; Dagliati et al., 2017) an analytical sequence, as it applied to Study III, was made up of three steps:

1. Selecting predictor domains, cleaning and reducing data. Based on knowledge about the subject matter and a literature review, relevant predictor domains were identified. Data mining filtering techniques were used to obtain a few relevant predictor variables within each domain.
2. Predictive model(s) construction. Classification methods (algorithms) and a strategy for handling preprocessing problems (such as class unbalance and missing data) were chosen.
3. Predictive model(s) validation. A validation strategy to measure performance of the planned classification methods was selected.

Empirical studies

Study I - Validating tests

The first study in this thesis focused on validating mathematical tests on tablet for children in second and third grade, as well as providing norms on tablet for these grades. The first four hypotheses/questions investigated the comparability between the paper and the tablet version of HRT, following the guidelines from the International Test Commission (2006):

1. The test-retest correlations for the tablet version of the Heidelberger Rechen Test (HRT)-scales Addition, Subtraction, Missing Term and Count Amount would be comparable to the test-retest correlations for the paper version.
2. The correlation between the paper and tablet format would be at the same level as the test-retest reliability for each format.
3. The correlations, or convergent validity, between the tablet version of the HRT-scales Addition, Subtraction and Missing Term and the tablet version of the four subtests on the Grade 3 Math Battery (Math Battery; Fuchs, Hamlet, & Powell, 2003), would be comparable to the correlations between the paper version of HRT and the DEMAT 4 test (Gölitz, Roick, & Hasselhorn, 2006).
4. Are the absolute measures of central tendencies (mean and standard deviation) for both formats comparable for HRT Addition, Subtraction, Missing Term and Count Amount?

The following research questions examined administrator effects for tests on a tablet, and psychometric differences in terms of reliability for various tests and populations:

5. Does the type of test administrator (i.e., test personnel or teacher) have an effect on the test-retest reliability for tablet tests?
6. Are the test-retest reliabilities on tablet tests (HRT and Math Battery) comparable between a sample of students with lower math performance and more representative samples?
7. What was the test-retest reliability for the Math Battery on tablet?

Table 1 presents an overview of the research questions, how they were investigated, central characteristics of the five samples (N=94-554), and what procedures were used in the validation study.

Table 1 *Overview of Analyses.*

Study	Participants*	Procedure	Hypothesis/question
Test-retest tablet, Same Administrators	n=128, Second grade	Whole class, Test administrator	First hypothesis[1]
Test-retest tablet, Different Administrators	n=114-134, Second grade	Whole class, Teacher respectively test administrator	First hypothesis[1]
Test-retest tablet, Before math training	n=264-265, Low performing, Second grade	Groups of 1-5 students with test administrator	Examination of performance effects on reliability
Test-retest tablet, After math training	n=236-238, Low performing, Second grade	Groups of 1-5 students with test administrator	Examination of performance effects on reliability
Paper vs Tablet	n=94, Second grade	Whole class, Test administrator	Second hypothesis[2]
Convergent Validity, HRT & Math Battery	n=554, Second and third grade	Whole class, Administrated by teacher	Third hypothesis[3]

* Variation in ”n” is due to missing data on some items in some of the analyses.
[1] First hypothesis: the test-retest correlations for the tablet version of the HRT-scales Addition, Subtraction, Missing Term and Count Amount would be comparable to the test-retest correlations for the paper version.
[2] Second hypothesis: the correlation between formats would be at the same level as the test-retest reliability for each format.
[3] Third hypothesis: the correlations, or convergent validity, between the tablet version of the HRT-scales Addition, Subtraction and Missing term and the tablet version of the four subtests on the Grade 3 Math Battery, would be comparable to the correlations between the paper version of HRT and the DEMAT 4 test.

Results showed that: (1) the test-retest reliability for the tablet version of the arithmetic HRT-scales (i.e., Addition, Subtraction, and Missing Term) was comparable to the test-retest reliability for the paper version, but the same level of comparability was not obtained for the pictorial scale HRT Count Amount, (2) the correlation between paper and tablet for the arithmetic HRT-scales were at the same level as the test-retest reliability for each format, but this was once again not true for HRT Count Amount, (3) correlations, or convergent validity, between the arithmetic HRT-scales and the Math Battery were comparable with the correlation between the paper version of HRT and another paper-based mathematical test, (4) scores were not interchangeable between paper and tablet, (5) no administrator effects were found, (6) test-retest reliability was lower among low-performing students (i.e., participants in the intervention study), and (7) test-retest reliability for Math Battery on tablet was acceptable for the more representative samples, but inferior for lower performing students. Although no previous studies were found that have compared tablet- and paper-based tests measuring academic skills, these results are in line with a meta-analytic comparison of paper- and computer-based tests were scores were significantly higher on computer (Kingston, 2009). Meanwhile, those findings and the present study depart from another meta-analysis comparing paper- and computer-based tests where no differences between formats were demonstrated (Shudong, Hong, Young, Brooks, & Olson, 2007).

In sum, evidence suggest that arithmetic scales (i.e., HRT-scales Addition, Subtraction, and Missing Term) can potentially be transferred from paper to tablet with comparable psychometric properties (questions 1-3), although complete comparability was not manifested because scores were not interchangeable between paper and tablet (question 4) which points to the need of separate norms for tablet tests. The same level of comparability was not found for a pictorial scale (i.e., HRT Count Amount) suggesting that further adjustments are needed when transferring such scales.

Study II - Effects of a math intervention

The second study investigated the effects of a math intervention on tablet, which was the main objective of the thesis. The validated tests in the first study were used to assess the outcome of the math training. The specific hypotheses were:

1. Low performing children in second grade participating in math training improve mathematical skills compared to children in control and placebo conditions.
2. The gained effects will be maintained during the follow-up period.
3. Intelligence and socio-economic status moderate the effect of math training.
4. Working memory training in combination with math training lead to more superior outcome in terms of mathematical skills.
5. Improvements in mathematics achievement will show a positive effect on mental health.

The effects of math training on tablet were investigated using a randomized placebo controlled design with direct and follow-up assessments at 6 and 12 months. Students were individually randomized to one of the following four groups: passive control, reading (placebo), mathematics, or mathematics combined with working memory training. Math training focused primarily on basic arithmetic, and secondarily on number knowledge (i.e., prerequisite skills to basic arithmetic) and word problem solving. Participants were 283 children in second grade with low mathematics achievement (at or below the 45th percentile on basic arithmetic) and representative characteristics. On average, students trained math close to 19 hours during approximately 20 weeks in both math conditions, and the combined math and working memory group trained working memory for an additional 9 hours. Students in the placebo group trained reading on average a little more than 17 hours during the same interval. Attrition was 14.5 % for the whole sample from pretest to 12 months follow-up, and it was highest (22.1 %) for the combined math and working memory training condition, followed by the math condition (15.8 %), the placebo (11.5 %), and the control (5.8 %).

The results demonstrated that both math conditions scored significantly higher than control and placebo on the post assessment of basic arithmetic, but not on arithmetic transfer and word problem solving. Given the virtually identical patterns (Figure 2), we collapsed the math groups, respectively the control and placebo groups. The collapsed math group displayed effects of medium size (d=0.53-0.67) on basic arithmetic at post assessment compared to the collapsed control/placebo condition (first hypothesis).

To begin with we expected the mathematics training to yield transfer effects in domains that were not trained directly, in particular the measures covering addition up to 18 and subtraction up to 18. Initially it appeared as if there was a transfer effect with significant effects for Subtraction 0-18, but not for Addition 0-18. But, this claim is weak given that only 10 of the 25 total items were in the transfer range (i.e., minuends 13 or larger) and the mean scores across all time points ranged from a minimum of 1.9 to a maximum of 7.23. Consequently, we cannot be certain that the significant improvements only depend on increased correct answers in the transfer domain.

The effect sizes faded out at 6 months follow-up (d=0.18-0.28), and vanished further at 12 months follow-up (d=0.03-0.13) (second hypothesis). Given the lower test-retest reliabilities reported on these measures for the intervention sample in study I, these effects should be interpreted with care. No effects were found on problem solving.

Non-verbal IQ was a significant moderator of immediate and follow-up effects, where students with lower non-verbal IQ benefitted more than higher non-verbal IQ students (Figure 3). Although socioeconomic status (SES) and diverse ethnic background were not significant moderators, the trajectories for these moderators showed interesting patterns that may indicate that students with lower SES and students at schools with higher diverse ethnic background benefitted more from training (Figures 4 and 5). Albeit speculative, one explanation for these patterns is that the study was underpowered in this respect and that the hypothesis would have been confirmed for SES and ethnicity as well given a larger sample (third hypothesis). Furthermore, no additive effects for working memory training were found (fourth hypothesis), although the working memory training was implemented successfully in the MA+WM group.

Figure 2 not published here for copyright reasons.

Please see this figure in the original publication of this article.

Figure 3 not published here for copyright reasons.

Please see this figure in the original publication of this article.

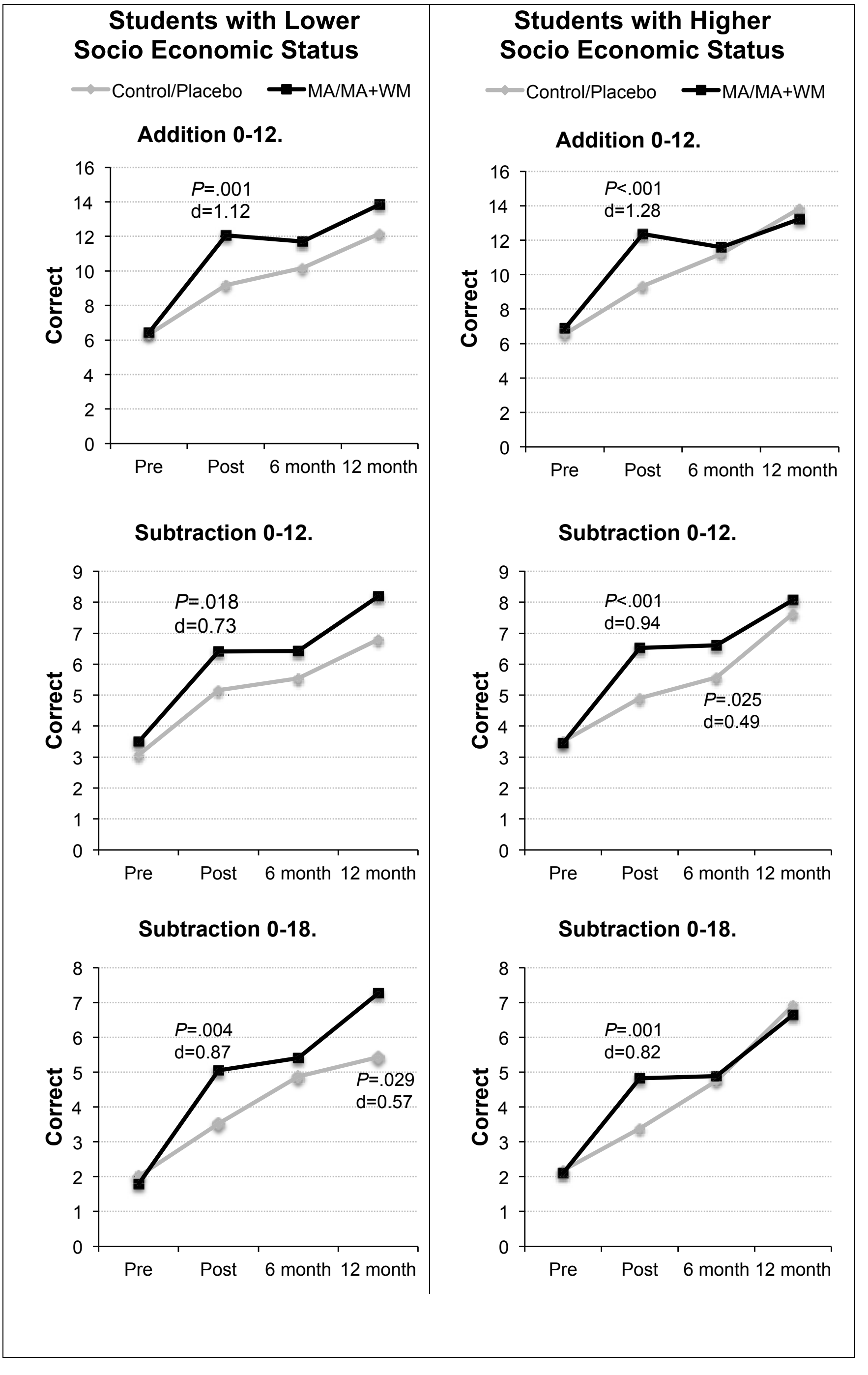


Figure 4. Socio Economic Status as moderator for significant primary outcome variabels on collapsed groups. *d*=Cohen's *d*.

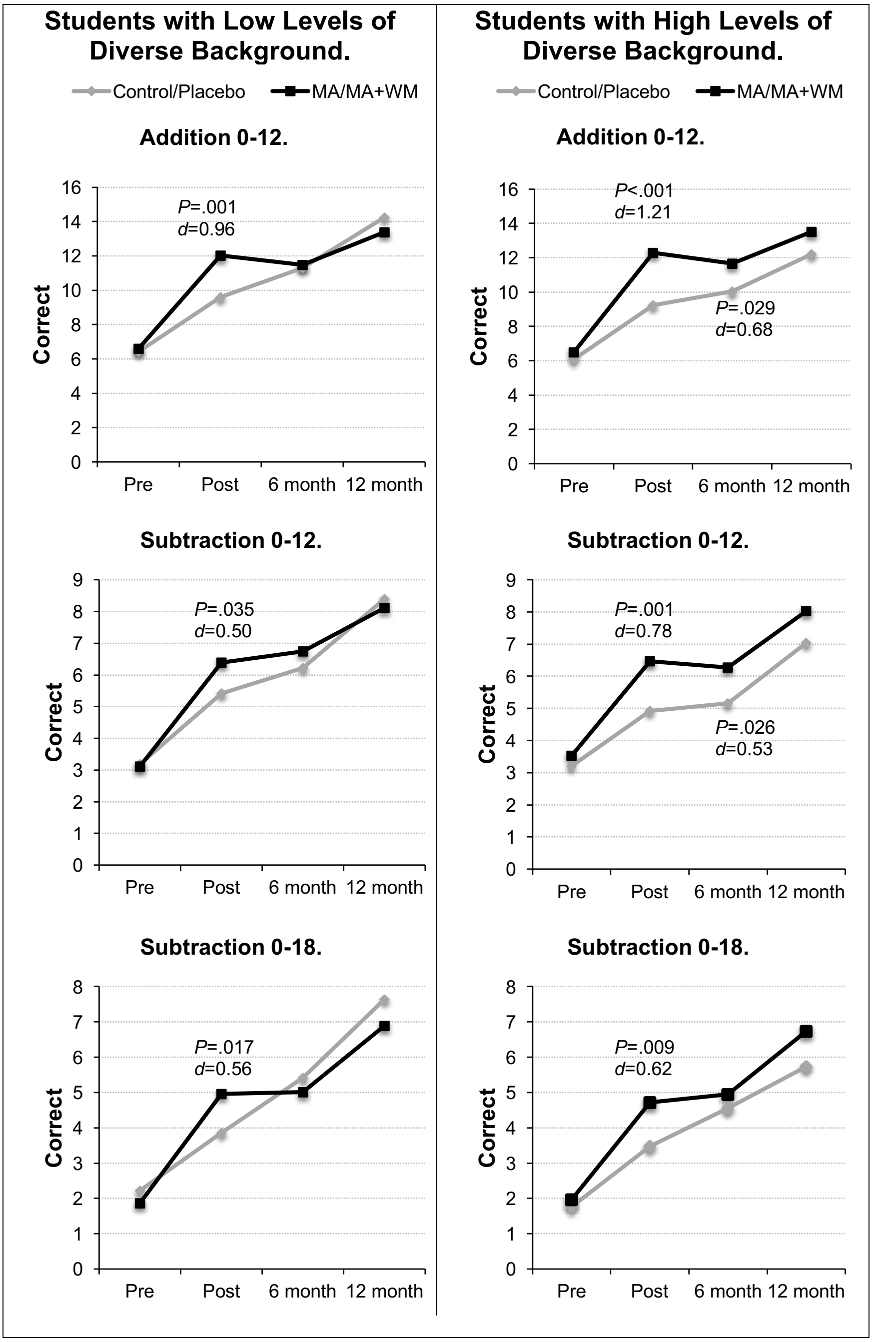


Figure 5. Diversity as moderator for significant primary outcome variabels on collapsed groups. *d*=Cohen's *d*.

The motivation for investigating if improved mathematics performance had an effect on mental health (fifth hypothesis) were the findings in a systematic review by Gustafsson et al. (2010) that demonstrated a reciprocal relationship between school performance and mental health. This implies that low school performance can lead to psychosocial problems, and conversely, that poor psychosocial conditions can lead to poor school outcome. Although the relationship is considered reciprocal, the randomized design of the current study provided an interesting opportunity to test the causality in one of the suggested directions, which has found support in other studies (e.g., Halonen, Aunola, Ahonen, & Nurmi, 2006; Morgan, Farkas, Tufis, & Sperling, 2008).

At all four assessment points (pre, post, and follow-ups at 6 and 12 months) questionnaires on students' mental health were completed by students, parents, and teachers.

Students were assessed by test personnel following a test protocol. Test administrators read questions and answer options verbatim to students and noted the answer on the questionnaire. Instruments using children as informants were Multidimensional Anxiety Scale for Children (MASC; March, Parker, Sullivan, Stallings, & Conners, 1997), Short Mood and Feelings Questionnaire (SMFQ; Angold, Costello, Messer, & Pickles, 1995) and Kidscreen containing the scales Physical Well-Being, Psychological Well-Being, Autonomy and Parents, Peers and Social Support, and School Environment (Ravens-Sieberer, et al., 2007).

The parent version of the Strength and Difficulties Questionnaire (SDQ; Goodman, 2001) was answered by parents at home and sent to the research group by mail. The subscales Emotion, ADHD, Peer, Conduct, and Prosocial, as well as the composite scales Externalizing Behavior, Internalizing Behavior, and the Total scale, were utilized in the analyses.

Teachers completed the teacher version SDQ (analyzing the same scales as above) and the Behavior Rating Inventory of Executive Function (BRIEF; Giola, Guy, Isquith, & Kenworthy, 1995) at school or at home.

Multilevel analysis with the four conditions demonstrated only one significant main interaction effect between time and group when controlling family wise for multiple comparisons (e.g.., student scales being one family). Although the Kidscreen Peer scale showed a significant interaction, no systematic pattern was found using post hoc pair wise comparison.

Consequently, the hypothesis stating that improvements in mathematics achievement would show a positive effect on mental health was not supported (fifth hypothesis).

Study III - Predicting positive long-term response

The third study explored predictors of positive long-term response to the math intervention. The specific hypotheses were:

1. Which factors significantly predict positive long-term outcomes in a mathematics tablet intervention for second graders with low math performance?
2. How are these factors related to each other?

Participants were 153 students who had completed math training in the mathematics or mathematics combined with working memory training conditions. Several predictor variables on math competencies, cognitive factors, students´ characteristics, and mental health were collected before training, and predictor variables from training (e.g., to what extent students had completed the program) were also gathered. A composite measure of arithmetic performance, assessed at 6, and 12 months after math training had terminated was used as outcome measure. Relevant predictor variables were identified after using data mining filtering techniques. An inductive decision tree algorithm and a logistic regression model were then applied to create models predicting long-term response to training. Lastly, the models were validated using 10-fold cross validation.

Model accuracies at 6 months follow-up (69.4 % for the decision tree model; X % for logistic regression) were significantly better than the baseline model (54.4 %). The predictive models were not significantly better than the baseline model at 12 months follow-up. The decision tree model showed that long-term responders at 6 months follow-up were students who had completed > 90 % of the training program, in combination with having at least one employed parent, shared custody, and training with default settings of program difficulty. Meanwhile, students having very low initial arithmetic performance and meeting these conditions, except for receiving easier (not default settings) training, were predicted to be long-term responders as well (first and second hypothesis).

To sum up, students coming from a relatively stable socioeconomic background who have completed at least 90 % of the training at default level are more likely to make long term gains. Data mining techniques can potentially inform school professionals on which students need additional support to make long-term gains when using educational technology applications. However, initial analyses like the current study should be interpreted tentatively since further validation with new data sets is needed to draw firm conclusions.

Discussion

Summary of principle findings

The empirical investigations in this thesis suggest that additional math training on tablet can improve basic arithmetic skills among low performing second graders compared to placebo and control conditions. Effects were medium sized at post assessment. The effects faded out at 6 months follow-up, where small effects were demonstrated, and vanished at 12 months follow-up. Students with lower non-verbal IQ seemed to benefit significantly more at post and follow-up assessments than higher non-verbal IQ students (study II). Students displaying a maintained positive response at 6 months follow-up were characterized by having completed at least 90 % of the math program at the default level, in combination with having a relatively more advantaged socioeconomic background (study III). Combining math training with working memory training did not demonstrate additional effects. Neither did improvements in mathematics performance impact mental health status among children. Lastly, the measurement investigations showed that arithmetic scales, but not a pictorial scale, on paper could be transferred to tablet with comparable psychometric properties, although separate norms are needed for tablet because scores were not interchangeable between the formats (study I).

Explaining the outcome

Interpretation of the magnitude of the intervention effect

The most relevant comparison of the results found in the current intervention would arguably be the effects demonstrated by small group instruction within the responsiveness-to-intervention (RTI) framework, because they target approximately the same population (i.e., at-risk students in RTI and low-performing in the current study) using similar design (i.e., comparing the intervention condition to a control group not receiving additional training). As mentioned in the introduction, effect sizes are typically reported to be medium sized (d=0.50-0.79) for RTI group instruction studies (Fuchs, Fuchs, & Compton, 2012), which is comparable to the results reported for the direct effects (d=0.53 and 0.67) observed in this thesis. The most notable difference between small group instruction and the current tablet intervention was that small groups are led by a tutor while the tablet training was conducted independently by the children.

Typically, effects are significantly larger in studies with high level of implementation compared to studies with lower levels of implementation (Cheung & Slavin, 2013). However, this conclusion should be viewed somewhat carefully because researchers who knew that there would be no difference between the control group and the intervention group may have described poor implementation as the cause. Still, this finding corresponds to the current intervention that should be considered to have a high level of implementation, as argued in detail in the paper on the second study.

The meta-analysis by Cheung and Slavin (2013) reporting on effects by ed-tech interventions does not apply to this context, because the studies did not use the same design (i.e., studies in the meta-analysis compared the intervention condition to a control condition that received a standard method or an alternative program resulting in substantially smaller effects).

Why was there a fade-out effect?

The fade out of effects at follow-ups replicates the findings in earlier non-technological math intervention studies on slightly younger children (Clements et al., 2013; Smith et al., 2013). The principal reason for fadeout effects is arguably that students did not reach mastery, or the necessary fluency criterion to maintain effects. Plausibly, a mastery or fluency criteria, defined as number corrects per time unit with no or few errors, need to be established since retention is supposed to depend heavily on rate of performance Johnson & Layng, 1992). Empirical support for the importance of fluency and a strong mastery goal orientation has been provided by several studies (e.g., Geary et al., 1996; Lin, Hung, Lin, Lin, & Lin, 2009). If a student masters an empirically derived criterion for fluency, chances are high that this skill will be retained even at longer follow-ups according to this line of research. Data on mastery criteria for math facts and problem solving would be most useful to improve the current intervention. The concepts of frustrational and instructional range, introduced by Deno and Mirkin (1977), refined by Shapiro (2004), and empirically investigated for mathematics in a meta-analysis by Burns, Codding, Boice, and Lukito (2010), could be used in conjunction with mastery criteria. Instructional range is the rate (usually corrects per minute) where students have not yet reached the mastery criterion and are susceptible for fluency building exercises without continuous guidance. Consequently, frustrational range is the rate immediately below instructional range where students require continuous modeling and scaffolding to not become overly frustrated. If frustrational and instructional range was even more closely defined for math facts and problem solving than existing research suggests, the current program could have been adjusted to fit different entry repertoires (i.e., frustrational or instructional) among students instead of having a one-size-fits all approach. Or alternatively, only students meeting the instructional criteria would have been included using a downsized version of the program, while students in the frustrational range would receive the full version. The standard approach would be to use linear models for predicting criteria for mastery and instructional range (frustrational range follows once instructional range is known). But, now other sophisticated analytical tools are available that could improve such estimates. Machine learning could likely be utilized to make non-linear predictions about the same criteria.

Why do children with lower non-verbal IQ seem to gain more?

The seemingly differential gains for children with lower non-verbal IQ are not surprising given that children who are resource weak (e.g., lower non-verbal IQ) are predicted to benefit more from increased resources, for instance in the form of increased quantity and quality of teaching, compared to students who are rich in resources (Gustafsson et al., 2010). This prediction comes from the Conservation of Resources Theory (COR theory) by Hobfol (1989; 2001), and the same pattern has also emerged from the research literature reported by Gustafsson (2003).

Potential moderating effects

As argued in the results section, even though moderator analyses of individual SES and diverse ethnicity revealed non-significant main interaction effects, the patterns depicted in Figure 4 and Figure 5 may indicate that the study was underpowered in this respect. Speculatively, the moderator analyses may have been significant had the sample been larger. As suggested above, it could be predicted that children with lower non-verbal IQ would benefit more from increased quantity and quality of teaching than children with higher non-verbal IQ. This could also apply to children with lower SES and children attending schools with a more diverse ethnic background, both of which could be considered less resourceful children. Lastly, the similar patterns emerging between individual SES and diverse ethnicity at the school level contribute to the coherence of these speculative findings since ethnicity has been regarded as a proxy for SES, or at least highly confounded with SES (Byrnes & Wasik, 2009).

Comparing the variable- and person-oriented analyses

The moderator analysis is considered a variable-oriented approach, in essence meaning that the person is viewed as a summation of variables over time. The predictive analysis, using the decision-tree model, on the other hand, takes a dynamic and holistic view of the person as an integrated totality over time. Although variable- and person-oriented analyses are often combined, they are used to study the person from two different worlds (Bergman & Trost, 2006). Consequently, the tentative suggestions that children with lower SES and children attending schools with a more diverse ethnic background benefit more from the math training at 6 and 12 months follow-ups (variable-oriented analyses) are not contradictory to the results in the predictive, or person-oriented, analysis where children with unemployed parents or only living with one parent were predicted to not show a response to intervention 6 months after training terminated. The variable-oriented analyses demonstrate that the magnitude of the suggested effect of math training for children with lower non-verbal IQ is larger than for children with higher non-verbal IQ, and speculatively that the effects are larger for students with lower SES and students at schools with higher ethnic diversity. Viewed from a holistic standpoint over time, the person-oriented analyses show that students with a combination of certain characteristics (i.e., having completed 90 % of the program at default setting, living with both parents, and having at least one employed parent) were predicted to present a prolonged response at 6 months follow-up.

Regarding the person-oriented analysis, the risk of model overfit implies that the model should be viewed with caution. Furthermore, the model partially replicated earlier findings where beneficial background has been predictive of better school performance (OECD, 2016). This particular model contributes with new, albeit tentative, knowledge about training variables that affect long-term outcome for this particular program. It is uncertain if this particular model would generalize to other math programs or topics. However, the non-linear combination of training variables, SES-variables and initial math performance can provide valuable recommendations on which predictor variables should be included in future studies using machine learning.

Absence of effects on problem solving

The main reason for the lack of effects on the problem solving measure Klurisen was probably that students had too little time to practice on the complex problem solving skills. However, Klurisen may also have been insensitive to incremental improvements because this item response scaled measure was developed across grades 2-5, while the intervention study assessed children in 2-3 grade. The absence of effects on problem solving, albeit the improvement in math facts, is not surprising given that a previous study by Fuchs et al. (2013) demonstrated that math facts improvement do not transfer to word problems when controlling for the role of number combination skill in word problems.

Lack of effects on mental health

The pervading absence of effects on mental health outcome, regardless of type of informant (i.e., child, parent, or teacher), is interesting because the research presented in the introduction suggests a reciprocal relation between academic achievement and psychological well-being (Gustafsson et al., 2010). There are several conceivable reasons for this: the impact of the intervention might have been to small, the domain of improvement was too narrow, or it could simply be the case that improved math achievement does not impact mental health.

Limitations

The differential attrition rates between conditions in the intervention study are a limitation in this thesis. Even though attrition rates were not significantly different between conditions from pre assessment to 12 months follow-up, the frequency of dropout for each condition varied: Control 5.8 %, Placebo 11.5 %, mathematics (MA) 15.8 % and mathematics combined with working memory intervention (MA+WM) 22.1 %. The attrition for the whole sample was 14.5 %, which indicates a larger share of dropout in the math conditions, especially in the MA+WM condition. In the paper (study II) a detailed account of reasons for terminating the study is presented in the Consort flow diagram. One reason for terminating the study was that students experienced the training as boring, which was reported by some children, while others reported migration (i.e., moving to another school, city, or country), having a teacher on sick leave, disappointment about the randomization, or difficulties with the working memory training as reasons for ending training. However, several children did not give reasons for leaving and the attrition was therefore not considered systematic, although the pattern of dropout was of concern.

The attrition rates in the psychometric study (study I) varied between 12 and 24 %, and the lack of individual data limited the possibility to determine if the attrition was systematic. However, a clear majority of the attrition was due to absence at one of the two testing occasions or technical problems (i.e., failing internet connection), which seemed less likely to be systematic.

As just mentioned, the lack of individual data, except for Test-retest sample 3, was another limitation in the psychometric study. However, the coverage of broad spectrums of parent education level and diverse ethnic background allowed for good generalizability. Furthermore, the effects of administrator and achievement effects were analyzed in depth with the three test-retest samples.

Another limitation concerns the low test-retest reliabilities for the low-performing students in the intervention sample. This limitation warrants a cautious interpretation of the reported effects of the intervention. The lack of psychometric stability found for HRT Missing Term, HRT Count Amount, and HRT Speed on both low-performing and more typically performing children should also be taken into account. Higher test-retest reliabilities for these measures would have been preferable. The possibility that the problem solving measure Klurisen may also have been insensitive to incremental improvements because it was an item response scaled measure developed across grades 2-5, while the intervention study assessed children in 2-3 grade, should also be considered a limitation.

The lack of comparison between the tablet version of HRT and an external criterion (i.e., grades) is also a restriction in the psychometric study. Comparison with an external criterion would have provided additional evidence of comparability between formats.

In the intervention study, the math training covered numbers up to 12 (e.g., 8+4, or 12-3). These tasks were measured with Math Battery Addition 0-12 and Math Battery Subtraction 0-12. In lack of more adequate validated measures, the Math Battery Addition 0-18 and Math Battery Subtraction 0-18 measures selected to assess near transfer effects (i.e., tasks containing numbers between 13 and 18), but not precisely in the desired way. These measures did not only cover the transfer domain, but also the directly trained domain, and were hence considered impure transfer measures. This short coming demonstrates the need of developing psychometrically valid and fine-grained assessment in this domain.

The problem of aligning the three waves so that they to be conducted during the same semesters presents another limitation in the intervention study. Although it could be argued that this increases the ecological validity, it also creates noise in the data.

The multiple comparisons in the intervention study potentially introduce a risk of mass significance. The four primary outcome measures (the Math Battery scales Addition 0-12: Addition 0-18; Subtraction 0-12; Subtraction 0-18) and the five secondary outcome measures (HRT Addition 0-999; HRT Subtraction 0-999; HRT Missing Term 0-99; HRT Count Amount; and the problem solving measure Klurisen) were utilized to answer specific hypotheses concerning the direct and transfer effects of training. Family wise adjustment of significance levels was used to account for multiple comparisons between the analyses of the primary and secondary outcome. The Sidak test was used to control for multiple comparisons within each analysis. A multivariate analysis including seven of the arithmetic measures in the study (i.e., HRT Addition, Subtraction, and Missing Term, and the four Math Battery scales) basically showed the same results as the original analyses, but since we beforehand had taken an interest in the direct and transfer effects this focus was maintained. The considerable cost, time, and efforts that go into an intervention study contributes to an inclination to make a thorough examination of the processes and outcomes. Hope is that the focus on a transparent presentation of effect size and confidence intervals will provide valuable knowledge for designing fewer, but more robust and specific measures in future studies.

The risk of overfitting the model in the predictive analysis (third study) should also be considered a limitation. The model ought to be interpreted with caution until tested on new data, which is always favored in data mining (Bellazzi, Ferrazzi, & Sacchi, 2011). Additional studies need to confirm the results before findings can be generalized.

Lastly, the particular pool of predictors utilized in the study may impact the performance of some predictors in the model. Had the same set of predictors been combined with other variables, the same predictors may have performed the same, worse or better. The cross validation technique used may compensate for this to some extent, but less so for a fairly small sample as in the present case.

Implications

This thesis demonstrates that a combination of technology and instructional design elements from behavior analytical theory can potentially effectively help low performing students decrease the achievement gap in mathematics. So far, the promise of increased effectiveness of educational interventions due to technological advancement has been unmet (Cheung & Slavin, 2013), but hopefully this thesis can provide new hope about the improvement of ed-tech interventions when combined with an effective theory of instruction. Interestingly, contemporary intervention research tends to investigate technological interventions less often than non-technological interventions. Cheung and Slavin (2013) concluded that even though technological innovations in reading and mathematics can supplement or support changes in teaching, they do not have an important impact on learning in themselves, whereby the authors instead stressed the importance of improving teacher education.

However, it is still possible that new technological innovations, with increased adaptivity and more tailored to support teachers efficiently, could challenge the above claim by Slavin to some extent. While the math program created within this thesis has some adaptivity, there are already more adaptive ed-tech platforms (e.g., Knewton, Dreambox, Alex) available that are likely to improve the effectiveness of teaching. Those platforms utilize increasingly sophisticated methods from the artificial intelligence field to recommend exercises tailored to the individual student's need, although their effectiveness remains to be documented scientifically. Regarding decision support, for instance, improvements in person-oriented analysis, like the machine learning model presented in this thesis, can be put into practical use as a decision support tool for teachers. An individual prediction along with recommendations could be presented to the teacher once the necessary information about the student has been entered into the program. In the current case, the program could encourage the teacher to ensure that students with low SES, in particular, complete the program and that students performance is monitored after training has terminated, where a performance drop could inform the teacher that additional training is needed. Another aspect of finding important predictors of learning outcomes would be to investigate student characteristics that are easier to change, such as home learning environment (Niklas & Schneider, 2017), instead of traditional SES-variables that are harder to influence.

To sum up, although evidence suggest that ed-tech interventions have limited impact on learning, the current thesis and contemporary developments within the ed-tech field could, hopefully, be interpreted as signs of a more substantial contribution by ed-tech programs in the future.

Directions for future research

Given the possibility to collect big data and the emerging methods (e.g., new machine learning algorithms) to analyze such data, this should be a promising field within educational intervention research. Big data collection during training can already, to some extent, enable individualized automatized feedback and tailor the training to each specific student's need. Furthermore, if data collected by national institutes, schools, and by digital learning tools in the classrooms were securely collected and combined in major data bases, then they could eventually provide a very informative basis for analyses, and decision support for everyday teaching, for instance on what kind of support would be most efficacious for a group of students with certain characteristics. Additionally, in a not too distant future, well-designed studies assessing students genome could eventually give rise to instructional interventions tailored to students with certain characteristics, genetic predictors included, whereby the differential effects of genes and environmental interactions could be taken into account. This type of research is already being conducted at the experimental but not yet applied level, as described in a meta-analysis by Bakermans-Kranenburg and van Ijzendoorn (2015) on Gene X Environment interactions.

Another domain in the mathematics intervention field, which has received far too little attention, is the development of effective instruction. My conviction is that intervention research is too focused on evaluation of programs, at the expense of investigations on how to develop very effective programs. The literature on program construction cited in the introduction and the development of very thoroughly iterated and successful programs, where Headsprout serves as an excellent example or model, deserves more attention from the scientific field because the ultimate goal of intervention research is to improve essential skills among children, not just to prove a particular scientific hypothesis.

Conclusion

The findings in this thesis point to the potential of improving critical mathematical skills by providing standardized and adaptive instruction on tablet with a minimum requirement of teacher work. The research also presents how testing of mathematical skills on tablet can facilitate assessment among young children, while at the same time bringing up the need for additional validation of mathematical measurement on tablets (e.g., to increase reliability among low performing students). The explorative analysis on differential effects of non-verbal IQ and the machine learning analysis on predictors of positive long-term response both suggest for what audiences and settings this type of math program may be most effective, as well as for which students and environments additional support, or extensions of the program, is needed.

The major learning in conducting this research over several years is that although plentiful of resources where used to create a program, vastly more resources are needed to make ed-tech interventions with lasting and profound effects on children´s math performance across elementary school grades. Because of this, there ought to be considerably more funding on developing scientifically based interventions, not just funding on evaluations on existing programs. Given the current evidence that an ed-tech intervention can be effective, and the empirical findings presented in the introduction, increased spenditure on intervention programs ought to be a priority for each nation that wants to improve or maintain economic growth, and to provide citizens with necessary skills to function in society.

Acknowledgements

Ata, thank you for an overwhelming, extremely instructive, and outstanding journey in the scientific world, and the world of loyalty and friendship. Honestly, I still can not believe we never had a real argument (or perhaps we had, but I never realized it). Instead you taught me how to use ";)" symbols in mail in stressful situations, and how to manage many situations that needed less seriousness. The way you have modeled scientific thinking has had a profound effect on me (at least I hope;). You have showed how a true scientist behaves: searching for solid answers to relevant questions using the scientific method. Not chasing questionable results that proves ones favorite point, or gives you special attention. After seven years of working together closely, although with a geographic distance between us, I am happy that we still enjoy (maybe even more) talking and laughing with each other. I promise to mail you a little less, just a litte, in the future, on the other hand: we should celebrate this with a real trip.

My thanks also go to the research coordinators (especially Hawi Merdasa and Emma Malmgren), and assistants whose efforts made this study possible. Special thanks go to Felix Funck for all hard work and for your loyalty.

I would like to thank Anders Jansson at Uppsala University for enabling me to take courses on-line, and Annika Landgren for helping out with administrative tasks. I am also grateful to the PhD-students Matilda Frick, Tommie Frick, and Olof Hjorth for helping me out during courses.

I am thankful to all schools and municipalities in Sollentuna, Landskrona (especially Lena Andersson at Dammhagskolan), and Helsingborg who participated, and especially to all children, parents, teachers, and school leaders for their hard work.

I would like to express my gratitude to Dr. Kent Johnson, Seattle: thank you for your highly appreciated contribution in designing the intervention and for teaching me how to do instructional design. I thank Joe Layng and Martha Leon at Headsprout for inspiration and guidance.

Kris Melroe in Seattle, thank you for being such a lovely person, always hosting me, or inviting me for dinner, and for sharing your knowledge with me.

Anna Aldenius, my favourite teacher, thank you for our collaboration thru the years, and for all that you have learned me. You are a role model.

Thank you, Anders Sjöberg at Stockholm University, for help in creating the problem solving test Klurisen using Item Response Theory.

I would like to thank Torkel Klingberg at Karolinska for our collaboration throughout several years. I would also like to thank Tomas Furmark, Uppsala University, for being my co-supervisor, and Gustaf Gredebäck for comments on my thesis.

Thanks to all persons at Kjell & Company, at Lund University, for being my first work place in Lund. I also want to thank the Department of Psychology at Lund University for providing me a place to work subsequently. Special thanks to Per Johnson and Robert Holmberg. Thank you, Daiva Daukantaite, for our scientific discussions. Thank you, Martin Bäckström, for helping me learn statistics.

People at Linköping University:
Rickard Östergren, thank you for always being cheerful, for our interesting and inspiring conversations, and for believing that I will not end up working at a health care center. I would also like to thank Ulf Träff for giving me advice at several times, for taking an interest in my work, and for a good collaboration. Jörgen Öberg, thank you for an epic trip to Seattle. Stefan Gustafsson, Linköping University, thank you for your valuable comments on my thesis on how to improve interventions.

Thank you, Peter Waldemarsson (former principal), for giving me the first chance in 2008 to try out mathematics teaching with your students.

Bosse Vinnerljung at Stockholm University, thank you for putting me in the right direction be giving with early support, and valuable recommendations on articles (Duncan et al.;)

Thanks to Doktorandinnebandylaget, especially to Per Davidsson and Arvid Erlandsson for taking the moral lead. Fikagruppen at M-huset, Lund: Thank you for pretty decent fika during the years.

Thank you, Petra and Martin Karlberg, for food, shelter and nice conversation during visits to Uppsala. And thank you, Totte T and Helena Tegenmark for food and shelter, including a runaway cat.

Thank you, Martin Forster, for the single best advice over the last 8 years: you told me to ask Ata to be my supervisor, because he "replied to mail, and held deadlines". It was true:). I also appreciate your recommendations on learning multilevel analysis, and to not write a systematic review.

Anders Wiberg, thank you for being my friend, and for using your professional expertise to help me cope last semester.

Thank you, Niklas Gustavsson, my oldest friend, for making me laugh at jokes already told in 1982.

Svante Schrieber, thank you for your intermittent obnoxious sms, and for encouraging my scientific efforts by saying "du gör ju iallafall något".

Per Davidsson, post-doc at HARVARD, thank you for your inexhaustible verbosity, and for, at the same time, being fun.

Anna Lindqvist, and Per again, thank you for our conversations, and for our thrilling, timeless podcast-session, which are a wonderful distraction for me.

Björn Sjödén, thank you for our ongoing inspiring conversations.

Thank you, Henrik Rosvall, for creating the fine cover for the thesis. I am very happy that I have found you, and I am looking forward to the rest of our journey.

Thank you my sister Hilda for your unconditional cheering and support. To my sister Britta - thank you for being there during the closing seminar. Thank you, my brothers Henrik and Arne for being so knipvänliga.

Special thanks to my parents-in-law, Kajsa and Pelle Hallstedt! For all your endless support with taking care of the children, Karin, and me (Gruppen). For your janitorial support of our summer house on Gullholmen. For taking our car to vehicle inspection two weeks ago. For providing food (köttbullar i långpanna etc), when we were "out of food". For all you positive encouragement via words and sms. You are wonderful.

To my mother and father, Birgit and Leo, for being my parents for 44 years. I love you. My thoughts are also with my grandfather Henrik Lindström.

I would also like to thank the following funders of my research: The National Board of Health and Welfare (Socialstyrelsen), Solstickan (especially Per-Anders Rydelius), Jerringfonden, and Sparbanksstiftelsen Öresund and Bo Lundgren. Thank you very much, Mats Ekstrand, for help with finding financing for the last mile of this thesis.

Lastly, my wife Karin, thank you for staying with me until the completion of this endless effort. I (almost) promise to never say these things again: "Det tar en halv dag, max.", and "Det är bara lite kvar nu.". Without you, and our children, all the work would have taken much less time, however I would never have lasted until the end. Most likely, I would have qualified 100 % (instead of 34 % as of now) for the A-syndrome and ended up somewhere awkward. I am fortunate to have you as my friend, and my love.

References

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From First Grade Forward: Early Foundations of High School Dropout. *Sociology of Education, 70*(2), 87-107. doi:10.2307/2673158

Adams, G. L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 Years beyond DISTAR*. Educational Achievement Systems, 319 Nickerson Street, Suite 112, Seattle, WA 98109.

Alloway, T. P. (2007). *Automated working memory assessment manual*. Retrieved from Oxford

Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International journal of methods in psychiatric research*.

Archer, A. L., & Hughes, C. A. (2011). *Explicit Instruction: Effective and efficient teaching*. New York, NY: Guilford Press.

Bailey, D. H., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness*, 1-33. doi:10.1080/19345747.2016.1232459

Bailey, D. H., Nguyen, T., Jenkins, J. M., & Domina, T. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental psychology, 52*(9), 1457-1469. doi:10.1037/dev0000188

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Dev Sci, 17*(5), 775-785. doi:10.1111/desc.12155

Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2015). The Hidden Efficacy of Interventions: Gene×Environment Experiments from a Differential Susceptibility Perspective. *Annual Review of Psychology, 66*(1), 381-409. doi:10.1146/annurev-psych-010814-015407

Bellazzi, R., Ferrazzi, F., & Sacchi, L. (2011). Predictive data mining in clinical medicine: a focus on selected methods and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(5), 416-430. doi:10.1002/widm.23

Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics, 77*(2), 81-97. doi:http://dx.doi.org/10.1016/j.ijmedinf.2006.11.006

Bergman, L. R., & Trost, K. (2006). The Person-Oriented Versus the Variable-Oriented Approach: Are They Complementary, Opposites, or Exploring Different Worlds? *Merrill-Palmer Quarterly, 52*(3), 601-632.

Bishop, J. (1992). The impact of academic competencies on wages, unemployment, and job performance. *Carnegie-Rochester Conference Series on Public Policy, 37*, 127-194. doi:http://dx.doi.org/10.1016/0167-2231(92)90006-5

Bjorkenstam, C., Weitoft, G. R., Hjern, A., Nordstrom, P., Hallqvist, J., & Ljung, R. (2011). School grades, parental education and suicide--a national register-based cohort study. *J Epidemiol Community Health, 65*(11), 993-998. doi:10.1136/jech.2010.117226

Björn, P. M., Aunola, K., & Nurmi, J.-E. (2016). Primary school text comprehension predicts mathematical word problem-solving skills in secondary school. *Educational Psychology, 36*(2), 362-377. doi:10.1080/01443410.2014.992392

Brounstein, P. J., Zweig, J. M., & Gardner, S. E. (2001). *Science-Based Substance Abuse Prevention: A Guide*. Retrieved from Rockville:

Burns, M., Matthew, K. B., Rebecca, K., & Megan, D. (2012). Effect of a Computer-Delivered Math Fact Intervention as a Supplemental Intervention for Math in Third and Fourth Grades. *Remedial and special education, 33*(3), 184-191.

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-Analysis of Acquisition and Fluency Math Interventions With Instructional and Frustration Level Skills: Evidence for a Skill-by-Treatment Interaction. *School Psychology Review, 39*(1), 69-83.

Byrnes, J. P., & Wasik, B. A. (2009). Factors predictive of mathematics achievement in kindergarten, first and third grades: An opportunity–propensity analysis. *Contemporary Educational Psychology, 34*(2), 167-183. doi:https://doi.org/10.1016/j.cedpsych.2009.01.002

Carr, J. M. (2012). Does Math Achievement h'APP'en when iPads and Game-Based Learning are Incorporated into Fifth-Grade Mathematics Instruction?

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113. doi:10.1016/j.edurev.2013.01.001

Cheung, A. C. K., & Slavin, R. E. (2015). How Methodological Features Affect Effect Sizes in Education. *Educational Researcher, 45*(5), 283-292.

Chodura, S., Kuhn, J.-T., & Holling, H. (2015). Interventions for Children With Mathematical Difficulties. *Zeitschrift für Psychologie, 223*(2), 129-144. doi:10.1027/2151-2604/a000211

Christmann, E. P., & Badgett, J. L. (2003). A Meta-Analytic Comparison of the Effects of Computer-Assisted Instruction on Elementary Students' Academic Achievement. *Information Technology in Childhood Education.*, 91-104.

Clarke, B., Nelson, N., & Shanley, L. (2016). Mathematics Fluency—More than the Weekly Timed Test *The Fluency Construct* (pp. 67-89). New York: Springer.

Clearinghouse, W. W. (2014). WWC procedures and standards handbook (Version 3.0). Washington, DC: Institute of Education Sciences.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal Evaluation of a Scale-Up Model for Teaching Mathematics With Trajectories and Technologies. *American Educational Research Journal, 50*(4), 812-850. doi:10.3102/0002831212469270

Codding, R. S., Chan-Iannetta, L., Palmer, M., & Lukito, G. (2009). Examining a classwide application of cover-copy-compare with and without goal setting to enhance mathematics fluency. *School Psychology Quarterly, 24*(3), 173-185. doi:10.1037/a0017192

Codding, R. S., Shiyko, M., Russo, M., Birch, S., Fanning, E., & Jaspen, D. (2007). Comparing mathematics interventions: Does initial level of fluency predict intervention effectiveness? *Journal of School Psychology, 45*(6), 603-617. doi:10.1016/j.jsp.2007.06.005

Christmann, E. P., & Badgett, J. L. (2003). A meta-analytic comparison of the effects of computer-assisted instruction on elementary students' academic achievement. *Information Technology in Childhood Education Annual*, *2003*(1), 91- 104.

Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., Tibollo, V., Bellazzi, R. (2017).

Machine Learning Methods to Predict Diabetes Complications. *Journal of Diabetes Science and Technology*, 1932296817706375. doi:10.1177/1932296817706375
Davis, O. S., Kovas, Y., Harlaar, N., Busfield, P., McMillan, A., Frances, J., . . . Plomin, R. (2008). Generalist genes and the Internet generation: etiology of learning abilities by web testing at age 10. *Genes Brain Behav, 7*(4), 455-462. doi:10.1111/j.1601-183X.2007.00370.x
De Houwer, J. (2013). *Advances in relational frame theory: Research and application*. New Harbinger Publications
Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21. doi:10.1016/j.intell.2006.02.001
Decker, S. L., & Roberts, A. M. (2015). Specific cognitive predictors of early math problem solving. *Psychology in the Schools, 52*(5), 477-488. doi:10.1002/pits.21837
Demichelis, F., Magni, P., Piergiorgi, P., Rubin, M. A., & Bellazzi, R. (2006). A hierarchical Naïve Bayes Model for handling sample heterogeneity in classification problems: an application to tissue microarrays. *BMC Bioinformatics, 7*(1), 514. doi:10.1186/1471-2105-7-514
Deno, S. L. M., P. K. . (1977). *Data-based program modification: A manual*.
Domingos, P. (2015). *The master algorithm: how the quest for the ultimate learning machine will remake our world*. New York: Basic Books.
Dowker, A. (2017). Interventions for Primary School Children With Difficulties in Mathematics. *Advances in Child Development and Behavior*. doi:http://dx.doi.org/10.1016/bs.acdb.2017.04.004
Duhon, G. J., House, S., Hastings, K., Poncy, B., & Solomon, B. (2015). Adding Immediate Feedback to Explicit Timing: An Option for Enhancing Treatment Intensity to Improve Mathematics Fluency. *Journal of Behavioral Education, 24*(1), 74-87. doi:10.1007/s10864-014-9203-y
Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental psychology, 43*(6), 1428-1446. doi:10.1037/0012-1649.43.6.1428
Dymond, S., & Roche, B. (Eds.). (2013). *Advances in Relational Frame Theory: Research and Application*. Oakland: New Harbinger Publications, Inc.
Farrington, D. P. (2015). Prospective longitudinal research on the development of offending. *Australian & New Zealand Journal of Criminology, 48*(3), 314-335. doi:10.1177/0004865815590461
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences, 8*(7), 307-314. doi:10.1016/j.tics.2004.05.002
Fergusson, D., Swain-Campbell, N., & Horwood, J. (2004). How does childhood economic disadvantage lead to crime? *J Child Psychol Psychiatry, 45*(5), 956-966. doi:10.1111/j.1469-7610.2004.t01-1-00288.x
Fleishner, J. E., Garnett, K., & Shepherd, M. J. (1982). Proficiency in arithmetic basic fact computation of learning disabled and nondisabled children. *Focus on Learning Problems in Math- ematics, 4*, 47-56.
Freeland, J. T., & Noell, G. H. (1999). Maintaining accurate math responses in elementary school students: the effects of delayed intermittent reinforcement and programming common stimuli. *Journal of Applied Behavior Analysis, 32*(2), 211-215. doi:10.1901/jaba.1999.32-211
Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math dif ficulty. *Journal of Educational Psychology*, *97*(3), 493.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: its power and limitations. *J Learn Disabil, 45*(3), 257-269. doi:10.1177/0022219412442167

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., .. . Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29-43. doi:10.1037/0022-0663.98.1.29

Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., Hamlett, C. L., Schatsch neider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? Journal of Educational Psychology, 100, 491–509.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., . . . Changas, P. (2013). Effects of First-Grade Number Knowledge Tutoring With Contrasting Forms of Practice. *Journal of Educational Psychology*. doi:10.1037/a0030127.supp

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). *Grade 3 Math Battery*. Department of Special Education.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., . . . Zumeta, R. O. (2009). Remediating Number Combination and Word Problem Deficits Among Students With Mathematics Difficulties: A Randomized Control Trial. *Journal of Educational Psychology, 101*(3), 561-576. doi:10.1037/a0014701

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental psychology, 47*(6), 1539-1552. doi:10.1037/a0025510

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L. (2009). First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive Development, 24*(4), 411-429. doi:http://dx.doi.org/10.1016/j.cogdev.2009.10.001

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of Arithmetical Competencies in Chinese and American Children: Influence of Age, Language, and Schooling.

Geary, D. C., Bow-Thomas, C. C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology, 54*(3), 372-391. doi:http://doi.org/10.1016/0022-0965(92)90026-3

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' Functional Numeracy Is Predicted by Their School Entry Number System Knowledge. *PLOS ONE, 8*(1), e54651. doi:10.1371/journal.pone.0054651

Geary, D. C., Widaman, K. F., Little, T. D., & Cormier, P. (1987). Cognitive addition: Comparison of learning disabled and academically normal elementary school children. *Cognitive Development, 2(3)*, 249-269.

Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early Identification and Interventions for Students With Mathematics Difficulties. *Journal of Learning Disabilities, 38*(4), 293-304. doi:10.1177/00222194050380040301

Gersten, R. M., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., & Witzel, B. (2009). *Assisting Students Struggling with Mathematics: Response to Intervention (RtI) for Elementary and Middle Schools*. Retrieved from

Gersten, R. M., & Chard, D. (1999). Number Sense: Rethinking Arithmetic Instruction for Students with Mathematical Disabilities. *The Journal of Special Education, 33*(1), 18-28. doi:10.1177/002246699903300102

Gioia, G. A., Guy, S. C., Isquith, P. K., & Kenworthy, L. (1996). *Behavior rating inventory of executive function*: Psychological assessment resources.

Giudici, P., & Figini, S. (2009). *Applied data mining for business and industry* (2. ed.). Hoboken, N.J: Wiley.

Goldiamond, I. (1974). Toward a Constructional Approach to Social Problems: Ethical and Constitutional Issues Raised by Applied Behavior Analysis. *Behaviorism, 2*(1), 1-84.

Goldiamond, I. (1979). Behavioral approaches and liaison psychiatry. *Psychiatric Clinics of North America, 2*, 379-401.

Goldiamond, I. (1984). Training parent trainers and ethicists in nonlinear analysis of behavior. *Parent training: Foundations of research and practice*, 504-546.

Goldman, S. R., Mertz, D. L., & Pellegrino, J. W. (1988). Extended practice of addition facts: Strategy changes in learning-disabled students. *Cognition and Instruction, 5*, 223-265.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American academy of child & adolescent psychiatry, 40*(11), 1337-1345.

Government Offices of Sweden, 2017. *Nationell digitaliseringsstrategi för skol väsendet.* Retrieved from http://www.regeringen.se/4a9d9a/contentassets/00b3d9118b0144f6bb95302f3e08d11c/nationell-digitaliseringsstrategi-for-skolvasendet.pdf

Greven, C. U., Kovas, Y., Willcutt, E. G., Petrill, S. A., & Plomin, R. (2014). Evidence for shared genetic risk between ADHD symptoms and reduced mathematics ability: a twin study. *J Child Psychol Psychiatry, 55*(1), 39-48. doi:10.1111/jcpp.12090

Griffin, K. W., Botvin, G. J., Doyle, M. M., Diaz, T., & Epstein, J. A. (1999). A six-year follow-up study of determinants of heavy cigarette smoking among high-school seniors. *Journal of Behavioral Medicine, 22*(3), 271-284. doi:Doi 10.1023/A:1018772524258

Gross, J., Hudson, C., & Price, D. (2009). The Long Term Costs of Numeracy Difficulties. London: Every Child a Chance Trust and KPMG.

Gross, T. J., Duhon, G. J., Hansen, B., Rowland, J. E., Schutte, G., & Williams, J. (2014). The Effect of Goal-Line Presentation and Goal Selection on First-Grader Subtraction Fluency. *The Journal of Experimental Education, 82*(4), 555-571. doi:10.1080/00220973.2013.813369

Gustafsson, J.-E., Nilsen, T., & Hansen, K. Y. (2016). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*. doi:10.1016/j.stueduc.2016.09.004

Gustafsson, J. E., Allodi Westling, M., Alin Åkerman, B., Eriksson, C., Eriksson, L., Fischbein, S., . . . Persson, R. S. (2010). *School, learning and mental health: a systematic review.* Stockholm: The Royal Swedish Academy of Sciences, The Health Committee

Gölitz, D., Roick, T., & Hasselhorn, M. (2006). *DEMAT 4: Deutscher Mathematiktest für vierte Klassen*. Retrieved from

Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *HRT 1-4 Heidelberger Rechentest. "Erfassung mathematischer Basiskompetenzen im Grundschulalter"*. Göttingen: Hogrefe.

Hallström Elthammar, O. (2015). *Klurisen: ett nytt sätt att mäta matematisk förmåga. Valideringsstudie av ett instrument för vetenskaplig testning av akademiska färdigheter hos barn i lågstadiet.*, Lund University, Lund. Retrieved from http://www.lu.se/lup/publication/5046683

Halonen, A., Aunola, K., Ahonen, T., & Nurmi, J. E. (2006). The role of learning to read in the development of problem behaviour: A cross-lagged longitudinal study. *British Journal of Educational Psychology*, *76*(3), 517-534.

Hanushek, E. A., & Woessman, L. (2015). *The knowledge capital of nations: Education and the economics of growth*: MIT Press.

Hao, L., Hu, A., & Lo, J. (2014). Two Aspects of the Rural-Urban Divide and Educational Stratification in China: A Trajectory Analysis. *Comparative Education Review, 58*(3), 509-536. doi:10.1086/676828

Haßler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: a critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning, 32*(2), 139-156. doi:10.1111/jcal.12123

Hassler Hallstedt, S. M., & Ghaderi, A. (submitted). Tablets Instead of Paper-based Tests for Young Children? Equivalence, Test-retest Reliability and Convergent Validity of the Mathematical Heidelberger Rechen Test 1-4.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. Springer Science & Business Media.

Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American psychologist, 44*(3), 513-524. doi:10.1037/0003-066X.44.3.513

Hobfoll, S. E. (2001). The Influence of Culture, Community, and the Nested-Self in the Stress Process: Advancing Conservation of Resources Theory. *Applied Psychology, 50*(3), 337-421. doi:10.1111/1464-0597.00062

Hotard, S. R., & Cortez, M. J. (1983). Computer assisted instruction as an enhancer of remediation. (ERIC Document Reproduction Service No. ED 139655).

Huffstetter, M., King, J. R., Onwuegbuzie, A. J., Schneider, J. J., & Powell-Smith, K. A. (2010). Effects of a computer-based early reading program on the early reading and oral language skills of at-risk preschool children. *Journal of Education for Students Placed at Risk*, *15*(4), 279-298.

Houten, R. V., & Thompson, C. (1976). The effects of explicit timing on math performance. *Journal of Applied Behavior Analysis, 9*(2), 227-230.

International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, *6*(2), 143-171.

Jitendra, A. K., Griffin, C. C., Haria, P., Leh, J., Adams, A., & Kaduvettoor, A. (2007). A comparison of single and multiple strategy instruction on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 99*(1), 115-127. doi:10.1037/0022-0663.99.1.115

Jitendra, A. K., Harwell, M. R., Dupuis, D. N., Karl, S. R., Lein, A. E., Simonson, G., & Slater, S. C. (2015). Effects of a research-based intervention to improve seventh-grade students' proportional problem solving: A cluster randomized trial. *Journal of Educational Psychology, 107*(4), 1019-1034. doi:10.1037/edu0000039

Johnson, K. (2008). Morningside mathematics fluency: Math facts (Vols. 1–6; curriculum program): Seattle, WA: Morningside Press.

Johnson, K. R., & Layng, T. J. (1992). Breaking the structuralist barrier. *American psychologist, 47*(11), 1475-1490.

Johnson, K. R., & Layng, T. V. (1994). *The Morningside Model of Generative Instruction*. Seattle, WA: Morningside Press.

Johnson, K., & Street, E. M. (2012). *Response to intervention and precision teaching: Creating synergy in the classroom*. Guilford Press.
Jordan, N. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*(2), 103-119. doi:10.1016/s0022-0965(03)00032-8
Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Developmental psychology, 45*(3), 850-867. doi:10.1037/a0014939
Joseph, L. M., Konrad, M., Cates, G., Vajcner, T., Eveleigh, E., & Fishley, K. M. (2012). A meta-analytic review of the cover-copy-compare and variations of this self-management procedure. *Psychology in the Schools, 49*(2), 122-136. doi:10.1002/pits.20622
Kingston, N. M. (2009). Comparability of Computer- and Paper-Administered Multiple-Choice Tests for K–12 Populations: A Synthesis. *Applied Measurement in Education, 22*(1), 22-37. doi:10.1080/08957340802558326
Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S. B., Dale, P. S., & Plomin, R. (2013). Literacy and numeracy are more heritable than intelligence in primary school. *Psychological science, 24*(10), 2048-2056. doi:10.1177/0956797613486982
Kroesbergen, E. H., & van Dijk, M. (2015). Working memory and number sense as predictors of mathematical (dis-) ability. *Zeitschrift für Psychologie*.
Kulik, J. A., Kulik, C. L. C., & Bangert-Browns, R. L. (1985). Effectiveness of computer-based education in elementary schools. *Computer in Human Behavior, I*, 59-74.
Lee, E.-H. (2009). Computerized Measurement for Asthma-Specific Quality of Life: Comparison with a Conventional Paper-and-Pencil Questionnaire. *J Korean Acad Nurs, 39*(6), 781-787.
Leung, F. K. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics, 47*, 35–51.
Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*(3), 215-243.
Li, Y., & Geary, D. C. (2013). Developmental Gains in Visuospatial Memory Predict Gains in Mathematics Achievement. *PLOS ONE, 8*(7), e70160. doi:10.1371/journal.pone.0070160
Li, Y., & Geary, D. C. (2017). Children's visuospatial memory predicts mathematics achievement through early adolescence. *PLOS ONE, 12*(2), e0172046. doi:10.1371/journal.pone.0172046
Liao, Y.-K. C. (1998). Effects of hypermedia versus traditional instruction on students' achievement: A meta-analysis. *Journal of research on computing in education, 30*(4), 341-359.
Lin, C.-J., Hung, P.-H., Lin, S.-W., Lin, B.-H., & Lin, F.-L. (2009). THE POWER OF LEARNING GOAL ORIENTATION IN PREDICTING STUDENT MATHEMATICS ACHIEVEMENT. *International Journal of Science and Mathematics Education, 7*(3), 551-573. doi:10.1007/s10763-008-9132-0
Malofeeva, E., Day, J., Saco, X., Young, L., & Ciancio, D. (2004). Construction and Evaluation of a Number Sense Test With Head Start Children. *Journal of Educational Psychology, 96*(4), 648-659. doi:10.1037/0022-0663.96.4.648
March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): factor structure, reliability, and validity. *Journal of the American academy of child & adolescent psychiatry, 36*(4), 554-565.

McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science*, *15*(11), 776-781.
McTiernan, A., Holloway, J., Healy, O., & Hogan, M. (2016). A Randomized Controlled Trial of the Morningside Math Facts Curriculum on Fluency, Stability, Endurance and Application Outcomes. *Journal of Behavioral Education, 25*(1), 49-68. doi:10.1007/s10864-015-9227-y
Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a Unidimensional Multilevel Construct: Validation, Sources of Influence, and Implications in a Longitudinal Study in Grades 1 to 4. *Scientific Studies of Reading, 9*(2), 85-116. doi:10.1207/s1532799xssr0902_1
Mercer, C. D. (1996). Constructivistic Math Instruction for Diverse Learners. *Learning Disabilities Research and Practice*, *11*(3), 147-56.
Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other?. *Journal of learning disabilities*, *41*(5), 417-436.
Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers' Social and Emotional Competence. Executive Summary. OPRE Report 2014-44. *MDRC*.
Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Herengracht 487, Amsterdam, 1017 BT, The Netherlands: International Association for the Evaluation of Educational Achievement.
Murnane, R. J., Willett, J. B., & Levy, F. (1995). *The growing importance of cognitive skills in wage determination*. Retrieved from
National Board of Health and Welfare (2010). Social Report 2010. Retrieved from http://www.socialstyrelsen.se/publikationer2010/2010-3-11.
National Council of Teachers of Mathematics. (2011, April 23). Principles and standards for mathematics education. Retrieved from http://www.nctm.org/standards/content.aspx?id=26792.
Nemmi, F., Helander, E., Helenius, O., Almeida, R., Hassler, M., Räsänen, P., & Klingberg, T. (2016). Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. *Developmental cognitive neuroscience*, *20*, 43-51.
Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly, 36*, 550-560. doi:http://dx.doi.org/10.1016/j.ecresq.2016.02.003
Niklas, F., & Schneider, W. (2017). Home learning environment and development of child competencies from kindergarten until the end of elementary school. *Contemporary Educational Psychology*, *49*, 263-274.
OECD. (2010). *The High Cost of Low Educational Performance: The Long-run Economic Impact of Improving PISA Outcomes*. Paris: OECD Publishing.
OECD. (2012). PISA 2012 Results in Focus. What 15-year-olds know and what they can do with what they know. Retrieved from http:////www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf
OECD. (2016). *Low-Performing Students: Why They Fall Behind and How to Help Them Succeed*. Retrieved from Paris: http://dx.doi.org/10.1787/9789264250246-en
Ozdamli, F. (2012). Pedagogical framework of m-learning. *Procedia-Social and Behavioral Sciences*, *31*, 927-931.

Pagani, L. S., Fitzpatrick, C., Archambault, I., & Janosz, M. (2010). School readiness and later achievement: a French Canadian replication and extension. *Developmental psychology, 46*(5), 984-994. doi:10.1037/a0018881

Panel, N. M. A. (2008). *Foundations for Success: The Final Report on the National Mathematics Advisory Panel*. Washington, DC.

Parsons, B. F. (2005). Measuring basic skills for longitudinal study: the design and development of instruments for use with cohort members in the age 34 follow-up in the 1970 British Cohort Study (BCS70).

Pitchford, N. J. (2015). Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. *Frontiers in psychology, 6*, 485. doi:10.3389/fpsyg.2015.00485

Pitkanen, T., Kokko, K., Lyyra, A. L., & Pulkkinen, L. (2008). A developmental approach to alcohol drinking behaviour in adulthood: a follow-up study from age 8 to age 42. *Addiction, 103 Suppl 1*, 48-68. doi:10.1111/j.1360-0443.2008.02176.x

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard Progressive Matrices. Including the Parallel and Plus Versions. With Norms for the SPM Plus and formulae for Calculating Change Score*: Harcourt.

Ravens-Sieberer, U., Auquier, P., Erhart, M., Gosch, A., Rajmil, L., Bruil, J., . . . Czemy, L. (2007). The KIDSCREEN-27 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. *Quality of Life Research, 16*(8), 1347-1356.

Rhymer, K. N., Henington, C., Skinner, C. H., & Looby, E. J. (1999). The effects of explicit timing on mathematics performance in second-grade Caucasian and African American students. *School Psychology Quarterly, 14*(4), 397-407. doi:10.1037/h0089016

Rhymer, K. N., & Morgan, S. K. (2005). Comparison of the explicit timing and interspersal interventions: Analysis of problem completion rates, student preference, and teacher acceptability. *Journal of Behavioral Education, 14*(4), 283-303.

Rhymer, K. N., Skinner, C. H., Henington, C., D'Reaux, R. A., & Sims, S. (1998). Effects of explicit timing on mathematics problem completion rates in african-american third-grade elemenetary students. *Journal of Applied Behavior Analysis, 31*(4), 673-677. doi:10.1901/jaba.1998.31-673

Riconscente, M. M. (2013). Results From a Controlled Study of the iPad Fractions Game Motion Math. *Games and Culture, 8*(4), 186-214. doi:10.1177/1555412013496894

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: replication and extension using a nationwide Canadian survey. *Developmental psychology, 46*(5), 995-1007. doi:10.1037/a0018880

Schacter, J., & Jo, B. (2017). Improving preschoolers' mathematics achievement with tablets: a randomized controlled trial. *Mathematics education research journal*. doi:10.1007/s13394-017-0203-9

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). NY: Guilford Press.

Shudong Wang, Hong Jiao, Young, M. J., Brooks, T., & Olson, J. (2007). A Meta-Analysis of Testing Mode Effects in Grade K-12 Mathematics Tests. *Educational and Psychological Measurement, 67*(2), 219-238. doi:10.1177/0013164406288166

Siegler, R. S., & Booth, J. L. (2004). Development of Numerical Estimation in Young Children. *Child Development, 75*(2), 428-444. doi:10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., . . . Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological science, 23*(7), 691-697. doi:10.1177/0956797612440101

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*(4), 273-296. doi:https://doi.org/10.1016/j.cogpsych.2011.03.001

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research, 75*(3), 417-453.

Slavin, R. E., & C., L. (2008). Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis.

Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating Math Recovery. *American Educational Research Journal, 50*(2), 397-428. doi:10.3102/0002831212469045

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*(5), 401-426. doi:10.1016/j.intell.2006.09.004

Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology, 98*, 265–281. doi:10.1037/0022-0663.98.2.265

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The Relationship Between Working Memory and Mathematical Problem Solving in Children at Risk and Not at Risk for Serious Math Difficulties. *Journal of Educational Psychology, 96*(3), 471-491. doi:10.1037/0022-0663.96.3.471

Swedish National Agency for Education. (2013). Diamant - ett diagnosmaterial i matematik. Retrieved from Diamant - ett diagnosmaterial i matematik

Tiemann, P. W., & Markle, S. M. (1991). *Analyzing instructional content*. Champagne, IL: Stipes Publishing Company.

Torgerson, C. J., Wiggins, A., Torgerson, D. J., Ainsworth, H., Barmby, P., Hewitt, C., & al., e. (2011). *Every child counts: The independent evaluation executive summary*. Retrieved from London:

Twyman, J. S., Layng, T. J., Stikeleather, G., & Hobbins, K. A. (2004). A non-linear approach to curriculum design: The role of behavior analysis in building an effective reading program. *Focus on behavior analysis in education, 3*, 55-68.

Ullman, H., Almeida, R., & Klingberg, T. (2014). Structural Maturation and Brain Activity Predict Future Working Memory Capacity during Childhood Development. *The Journal of Neuroscience, 34*(5), 1592.

Houten, R. V., & Thompson, C. (1976). The effects of explicit timing on math performance. *Journal of Applied Behavior Analysis*, *9*(2), 227-230.

Vasilyeva, M., Laski, E. V., & Shen, C. (2015). Computational fluency and strategy choice predict individual and cross-national differences in complex arithmetic. *Developmental psychology, 51*(10), 1489-1500. doi:10.1037/dev0000045

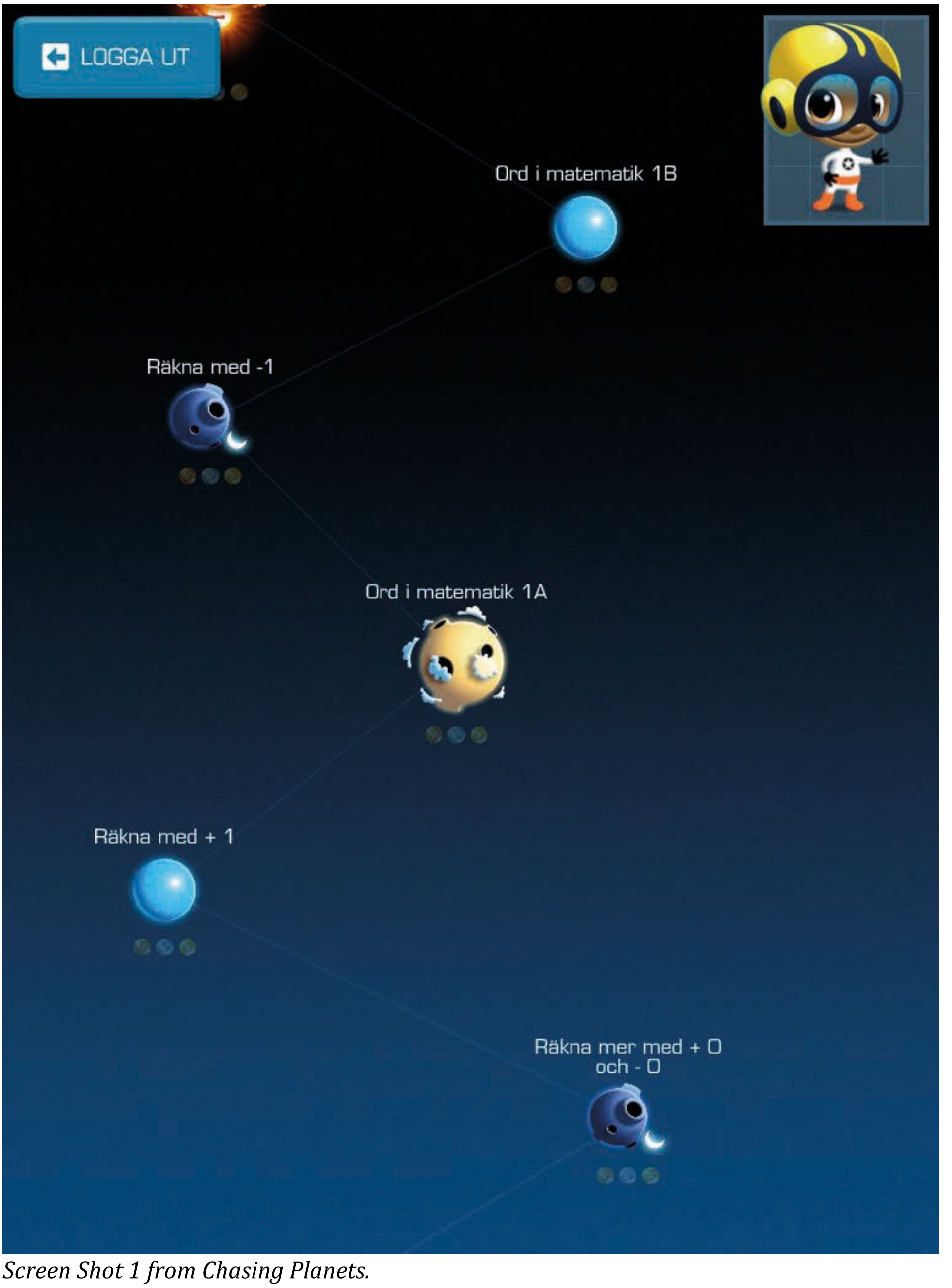
Verkijika, S. F. (2015). *Assessing the use of a Brain-Computer Interface (BCI) in mathematics education: the case of a cognitive game.* University of the Free State.

Viterbori, P., Usai, M. C., Traverso, L., & De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: A longitudinal study. *Journal of Experimental Child Psychology, 140*, 38-55. doi:http://dx.doi.org/10.1016/j.jecp.2015.06.014

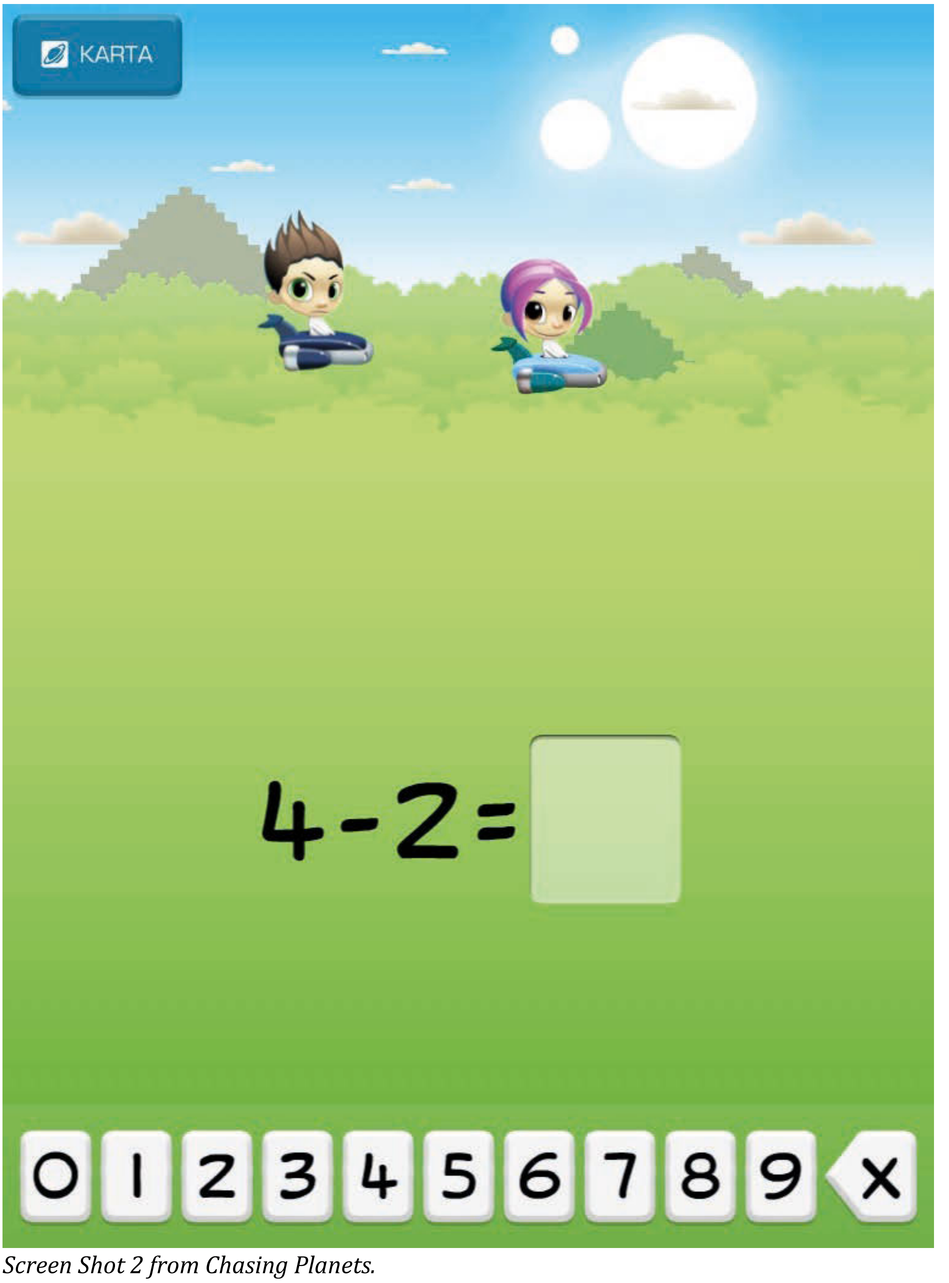
Wechsler, D. (1991). *Manual for the Wechsler intelligence scale for children - (WISC-III)*. Retrieved from San Antonio, TX:

Witten, I. H., Eibe, F., Hall, M. A., & Pal, C. J. (2017). *Data mining: practical machine learning tools and techniques* (4 ed.). Cambridge, MA: Morgan Kaufmann.

Wright, R. J., Martland, J., & Stafford, A. (2006). *Early numeracy: Assessment for teaching and intervention* (2nd ed.). London: Paul Chapman.
Wynn, K. (1995). Infants possess a system of numerical knowledge. *Current Directions in Psychological Science, 4*(6), 172-177.



Screen Shot 1 from Chasing Planets.
The space map containing the 261 planets/exercises. The student taps on the planet and starts practicing (as shown on Screen Shot 2). A new planet is available, or unlocked, when the student has reached at set criterion on the preceeding planet, although the criterion is adjustable if the student has difficulties.



Screen Shot 2 from Chasing Planets.
Fluency building training on math facts. This is a timed fluency phase where the student's avatar (to the right) races against a computer avatar. The race ends when one avatar reaches the right part of the screen. Preceding exercises focus on modeling and guided practice, withou timing, on number families containing math facts.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 150

Editor: The Dean of the Faculty of Social Sciences

A doctoral dissertation from the Faculty of Social Sciences, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences".)

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

Distribution: publications.uu.se
urn:nbn:se:uu:diva-336694