JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH
543

Jonna Salminen

Response to Computer-Assisted Intervention in Children Most at Risk for Mathematics Difficulties

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 543

Jonna Salminen

Response to Computer-Assisted Intervention in Children Most at Risk for Mathematics Difficulties

Esitetään Jyväskylän yliopiston kasvatustieteiden tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Ruusupuisto-rakennuksen Helena -salissa (RUU D104)
joulukuun 10. päivänä 2015 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education of the University of Jyväskylä,
in building Ruusupuisto, auditorium Helena (RUU D104), on Deceember 10, 2015 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

Response to
Computer-Assisted Intervention
in Children Most at Risk
for Mathematics Difficulties

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 543

Jonna Salminen

Response to Computer-Assisted Intervention in Children Most at Risk for Mathematics Difficulties

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2015

Editors
Markku Leskinen
Department of Education, University of Jyväskylä
Pekka Olsbo, Sini Tuikka
Publishing Unit, University Library of Jyväskylä

Cover photo by Jonna Salminen

URN:ISBN:978-951-39-6432-0
ISBN 978-951-39-6432-0 (PDF)

ISBN 978-951-39-6431-3 (nid.)
ISSN 0075-4625

Jyväskylä University Printing House, Jyväskylä 2015

ABSTRACT

Salminen, Jonna
Response to computer-assisted intervention in children most at risk for mathematics difficulties
Jyväskylä: University of Jyväskylä, 2015, 84 p.
(Jyväskylä Studies in Education, Psychology and Social Research
ISSN 0075-4625; 543)
ISBN 978-951-39-6431-3 (nid.)
ISBN 978-951-39-6432-0 (PDF)

Evidence of an association between poor early number skills and a high risk for later difficulties in mathematics manifests a need for both early identification of risk and effective prevention of difficulties. However, only a few theory-based assessment and progress monitoring tools, and intervention methods are available for this purpose in Finnish pre-primary education. This thesis aimed to examine to what extent a theory-based computer-assisted intervention method (GraphoGame Math, GGM) can support early number skills development in children (6–7-year-olds) most at risk for mathematics difficulties (MD; defined as performance below the 10th percentile). The specific aims were threefold: to examine, first, the immediate and delayed condition-specific effects of an intensified GGM intervention; second, the potential effects of sample characteristics on gain scores; and third, the effects of two theory-based intervention sequences on basic addition skills. The thesis is based on three intervention data sets collected during the years 2007–2011. The first data set comprised 17 (out of a total of 236 children), and the second data set 21 intervention participants (of approximately 350 children) identified as having poor early number skills. The third data set comprised 33 intervention participants identified as having a need for extra support for their early number skills (Examination 1), and 14 participants identified as having poor early number skills (Examination 2) (out of a total of 278 children). The results revealed: 1) positive, condition-specific, immediate and delayed intervention effects after intensified GGM practice on dot counting, verbal counting, composing, and basic addition skills; 2) individual differences in response to intervention among the children most at risk for MD, while the benefits for low-achieving children were more general and transferable; and 3) specific effects of number concept training on basic addition skills. The findings suggest that, theory-based computerized methods with game-log data analyses can be successfully used in systematic progress monitoring, remediation of initial gaps, and ongoing identification of difficulties. The data sets were collected as part of the LukiMat-project coordinated by the Niilo Mäki Institute and funded by the Finnish Ministry of Education and Culture since 2007.

Keywords: early number skills development; mathematics difficulties (MD); children most at risk for MD; computer-assisted intervention (CAI); GraphoGame Math (GGM); responsiveness to intervention

Author's address
Jonna Salminen
Department of Education
P.O. Box 35
FI-40014 University of Jyväskylä, Finland
jonna.salminen@jyu.fi

Supervisors
Professor Mikko Aro
Department of Education
University of Jyväskylä, Finland

University researcher Tuire Koponen
Department of Education
Niilo Mäki Institute
University of Jyväskylä, Finland

Professor Anna-Maija Poikkeus
Department of Teacher Education
University of Jyväskylä, Finland

Reviewers
Academy professor Erno Lehtinen
Department of Education
University of Turku, Finland

Associate professor Bert De Smedt
Faculty of Psychology and Educational Sciences
University of Leuven, Belgium

Opponents
Academy professor Erno Lehtinen
Department of Education
University of Turku, Finland

ACKNOWLEDGEMENTS

The data for this thesis have been collected during the LukiMat-project carried out by the Niilo Mäki Institute and funded by the Finnish Ministry of Education and Culture in 2007–2015. I wish to thank the entire LukiMat research group for supporting my work and assisting with the development of the assessment and intervention methods. I also thank all my co-workers and research assistants who helped with organizing the intervention studies, collecting the data, and participated in editing, programming, as well as testing the different versions of GraphoGame Math. I also wish to express my gratitude to the University of Jyväskylä and the Foundation for the Haukkala Child Psychiatry Institute for their financial support of this work. I especially want to thank the children and their parents, kindergarten teachers and special kindergarten teachers from 74 pre-primary education groups who voluntarily participated in the intervention studies discussed in this thesis. Finally, I wish to thank my colleagues in the discipline of special education for all the humor we have shared during this process. My deepest gratitude is owned to reviewers of this thesis Academy professor Erno Lehtinen and Associate professor Bert De Smedt and to my supervisors Professor Mikko Aro, Doctor Tuire Koponen, Professor Anna-Maija Poikkeus, and to my mentor Doctor Markku Leskinen. I also wish to thank my family and friends, who supported me in countless ways while finalizing this thesis.

Jyväskylä, December 2015,
Jonna Salminen

ORIGINAL PAPERS

This thesis is based on the following three publications, referred to as sub-studies I, II, and III in this compilation dissertation:

Article I	Salminen, J., Koponen, T., Räsänen, P., & Aro, M. (2015). Preventive support for kindergarteners most at-risk for mathematics difficulties: Computer-assisted intervention. *Mathematical Thinking and Learning, 17*(4), 273–295. doi: 10.1080/10986065.2015.1083837
Article II	Salminen, J., Koponen, T., Leskinen, M., Poikkeus, A-M., & Aro, M. (2015). Individual variance in responsiveness to early computerized mathematics intervention. *Learning and Individual Differences, 43*, 124–131. doi: 10.1016/j.lindif.2015.09.002
Article III	Salminen, J., Koponen, T., Sorvo, R., Peura, P., & Aro, M. Response to computerized early number intervention: The role of risk-level. Manuscript submitted for publication.

These articles are reprinted with permission of the publishers. Copies of the articles are appended to this report.

The author of this thesis is the first author of all the three research articles. She had a major role in further developing the GraphoGame Math method from its original version, planning the intervention studies, coordinating data collection, analyzing the data, interpreting the results, reviewing the literature, and preparing the manuscripts for all the individual articles.

FIGURES

TABLES

CONTENTS

1 INTRODUCTION

Recent changes in the Finnish basic education legislation require that effective early prevention shall be made available to children in need of extra support in learning (Basic Education Act, 642/2010, section 30). Sufficient support should be offered immediately when weaknesses or delays in learning are identified to reduce the potential learning difficulties and the need for long-term special education efforts. The focus on early support is in line with international recommendations on the provision of effective, evidence-based, secondary and tertiary prevention (Fuchs & Fuchs, 2009; Fuchs, Fuchs, & Compton, 2012; Fuchs, Mock, Morgan, & Young, 2003). In Finland, support for growth and learning is formalized at three levels: general support, intensified support, and special support (Finnish National Board of Education, 2010). This three-level model of support which emphasizes early identification and intervention is endorsed as a goal for educational settings but has not yet been implemented widely and systematically as proposed by the response-to-intervention model (RTI) (Björn, Aro, Koponen, Fuchs, & Fuchs, 2015; Fuchs & Fuchs, 2001; Fuchs & Fuchs, 2006; Fuchs et al., 2003) and recent research findings (e.g., Butterworth, Varma, & Laurillard, 2011). Therefore, there is an evident need for further research on the implementation of the three-level model of support (i.e., general, intensified, and specific support) in pre-primary educational settings in Finland.

Inter-individual differences in early number skills performance appears to be highly stable from kindergarten (called pre-primary education in the Finnish context) to later primary grades in both international (Jordan, Kaplan, Locuniak, & Ramineni, 2007; Kohli, Sullivan, Sadeh, & Zopluoglu, 2015; Morgan, Farkas, & Wu, 2009, 2011; Murphy, Mazzocco, Hanich, & Early, 2007) and national research (e.g., Aunio & Niemivirta, 2010; Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Koponen, Salmi, Eklund, & Aro, 2013; for review see also Mazzocco & Räsänen, 2013). Differences between low- and typically-performing children can already be seen in the pre-primary year (recently Aunio, Heiskari, van Luit, & Vuorio, 2015). Weaknesses in early number skills likely contribute to low-performing children's inability to benefit from instruction to the same extent as their age-peers (e.g., Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006). Similar findings have been documented among Finnish children in pre-primary education (e.g., Aunola et al., 2004). More research is needed in the Finnish context because relatively few theory-based assessment and progress moniroring tools, and intervention methods are available for implementation in kindergarten or daycare settings (cf. Aunio, 2006; Hannula, 2005; Mattinen, 2006; Mononen, 2014).

Evidence indicates that children with the most severe difficulties (those performing below the 10th percentile) differ from low-achieving children (performance between the 11th and 25th percentiles) by having even more persistent deficits in number skills (e.g., Morgan et al., 2009, 2011; Murphy et al., 2007). Therefore, research on typical and atypical early number skills development, and on early identification and support is highly important (see also Butterworth et al., 2011; Geary 2011b).

The origin of mathematical deficits, performance levels, and the needs for support in early number skills vary greatly among children having (or at risk for) mathematics difficulties (e.g., Geary, 2011b; Jordan, Hanich, & Kaplan, 2003; Rubinsten & Henik, 2009; von Aster & Shalev, 2007). Accordingly, early intervention programs should be adapted to take into account the individuality and the nature of children’s problems in terms of duration, intensity, and numerical as well as non-numerical cognitive training content (e.g., Dowker, 2001; Fuchs et al., 2012; Geary, 2011b). Due to inter-individual variation in responsiveness to intervention, ongoing identification and progress monitoring are needed to determine which children respond or do not respond to early support, and how to change or individualize the instruction as needed (cf. Fuchs et al., 2012). Computer-assisted intervention (CAI) is a promising method to identify responders and non-responders, and individualize early support through meaningful training content, adaptation, and feedback system (e.g., Butterworth et al., 2011; Cheung & Slavin, 2013; Slavin & Lake, 2008). Personal game-log data can be used for ongoing progress monitoring (cf. recently Obersteiner, Reiss, & Ufer, 2013). Moreover, targeted CAIs could help prioritize the use of teaching resources in both special and general education settings (Clements, 2002).

The purpose of this thesis was to examine the value of the theory-based, computer-assisted GraphoGame Math (GGM) game as a potential evidence-based method to support early number skills development in the Finnish kindergarteners most at risk for MD. Focusing on a freely downloadable CAI program, the study also aimed at increasing opportunities for equal education and the availability of cost-effective methods to all children across Finland regardless of their community and family resources. The terms “kindergarten” and ”kindergarteners” are used interchangeably with ”pre-primary education” and “children in pre-primary education” to describe the year of instruction provided to 6-year-old Finnish children before they enter primary school.

A method must fulfill several requirements before it can be addressed as an evidence-based intervention. Identification of evidence-based practices relies on four critical indicators: a qualified research design, methodological quality, quantity of research, and magnitude of effect (Cook & Cook, 2013; Cook, Tankersley, & Landrum, 2009). These criteria, along with the guidelines for quality indicators in experimental and quasi-experimental research in special education (Gersten, Fuchs et al., 2005), were taken into consideration when examining the intervention effects of GGM and discussing the reliability and validity of sub-studies I, II, and III. In this thesis, evidence-based intervention is used as a standard for assessing the strengths and weaknesses of GGM as a potential targeted intervention method for intensive and individualized practice.

The data of this thesis have been collected during the LukiMat-project coordinated by the Niilo Mäki Institute and funded by the Finnish Ministry of Education and Culture since 2007 (see www.lukimat.fi). The Finnish Advisory Board on Research Integrity's guidelines for ethical issues in research were followed throughout the study process.

1.1 Early number skills development

The theoretical models of early number skills development propose several independent sub-skills that form a hierarchy of developmental steps. Infants seem to have an inborn ability to determine numerical quantities using two preverbal mechanisms: an approximate number system (ANS), with respect to larger quantities (> 4), and an object tracking system (OTS), with respect to smaller quantities (≤ 4) (Agrillo, Piffer, Bisazza, & Butterworth, 2015; see also Feigenson, Dehaene, & Spelke, 2004). ANS appears to be dependent on the ratio between quantities allowing a child to quickly estimate which of the two different quantities is larger, provided that the difference between the sets of objects is large enough (Dehaene, 2011; von Aster, 2000; von Aster & Shalev, 2007). OTS is believed to keep track of elements and allow a child to discriminate precisely a small number of elements (e.g., Agrillo et al., 2015; Sella, Berteletti, Lucangeli, & Zorzi, 2015). However, because different measures are used, it is still unclear whether the ANS or OTS system is activated when discriminating small sets of quantities (Agrillo et al., 2015; Cordes & Brannon, 2008).

It has been proposed that the object tracking system allows a child to quickly and accurately discriminate small sets of quantities without counting (Sella et al., 2015). This phenomenon is called as subitizing (Dehaene, 2011; Geary, 2013; Starkey & Cooper, 1980; von Aster & Shalev, 2007). A subitizing range seems to be restricted to small quantities, such as 1–3, or 1–4 (Piazza, Mechelli, Butterworth, & Price, 2002; Reeve, Reynolds, Humberstone, & Butterworth, 2012). Another way to discriminate a relatively small set of objects just over the subitizing range is to constitute the whole quantity by forming small subsets of the total amount presented (conceptual subitizing) (Sarama & Clements, 2009). A child may, for example, perceive the quantity of five via two small sets (e.g., two and three), without counting all the objects one by one. No matter the mechanism underlying the quick discrimination of quantities, it is thought to help children in learning different numerical representations such as number words, quantities, and number symbols (Butterworth, 2005; Geary, 2013; von Aster & Shalev, 2007).

The ability to distinguish small sets of quantities primes children to learn small number words (i.e., one, two, and three) and their numerical values relatively early (by age 3.5) (Wynn, 1990). Although this skill is a marker of incipiently developing number concept skills, young children still usually recite number words without grasping the idea of meaningful counting (Krajewski & Schneider, 2009), signifying that the order of number words is not yet exact for them. Young children typically repeat, miss, or mix number words while trying to recite the verbal number list beyond one, two, and three. Children's understanding of the numerical value of the number words and their correct order in the number sequence gradually becomes more precise (by age 4.5) (Fuson, 1988, 2009). Around this time, children acquire the exact number word sequence, although they do not always connect number words to exact quantities (Krajewski & Schneider, 2009). Most 4-year-old children know the connection between the number words one, two, three, and four and their corresponding quantities (Dowker, 2008a; Krajewski & Schneider, 2009).

Small numbers are used more frequently in everyday situations, and children more quickly distinguish them from each other than larger numbers (Siegler & Booth, 2004). For example, young children perceive the number word 10 as big (representing much) (Krajewski & Schneider, 2009), and the number word 100 is usually the largest the child knows (representing very much) (Krajewski & Schneider, 2009). Among children around 4 years old, representation of numbers 0–10 has been found to be rather linear, whereas representation of larger numbers from 10 to 100 or 1000 tends to be logarithmic (see Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010). This means that number line representation is dependent on the familiarity of number words and relations between numbers. Representation of larger numbers tends to become more precise and linear throughout formal learning (Berteletti et al., 2010). At school age, children's number line skills develop further, enabling their mental use as a tool for solving basic arithmetic calculations (i.e., addition, subtraction, multiplication, division) (cf. von Aster & Shalev, 2007).

To discriminate the exact amount of quantities a child needs the ability to count objects (Dehaene, 2011). At this developmental step (by age 5), number words are not anymore isolated from quantities (Krajewski & Schneider, 2009). Exact object counting requires an understanding of five counting principles: the three procedural how-to-count principles, the abstraction, and the order-irrelevance principles (Gelman & Gallistel, 1978). The how-to-count principles comprise the following: 1) the one-to-one correspondence between counting words and the objects being counted, 2) the stable order of counting words, and 3) the cardinal principle (an understanding that the last counted number word represents the total amount of items, which allows a child to determine the cardinal number of elements in a set). An understanding of 4) the abstraction principle means that a child knows what can be counted, and whether the elements to be counted belong to the certain set that is supposed to be counted. The internalized 5) order-irrelevance principle allows a child to understand that the order in which the elements are counted is not relevant. It has been acknowledged that the spontaneous tendency to count elements in an environment provides practice for exact object counting procedures (Gelman & Gallistel, 1978). It is worth noting that a tendency to spontaneously focus on countable elements (i.e., spontaneous focusing on numerosities; SFON) has been shown to be associated with counting skills development (e.g., Hannula & Lehtinen, 2005; Hannula, Räsänen, & Lehtinen, 2007). Von Aster and Shalev (2007) include verbal counting skills and counting strategies in the same developmental step of acquiring the verbal number system. This step is thought to require attentional control and executive functioning (von Aster & Shalev, 2007).

With the help of the exact object counting skill, the link between the number words, quantities and symbols becomes precise (Geary, 2013; mapping; Krajewski & Schneider, 2009; number concept skill). This developmental step is needed to describe the exact amount of quantities and furthermore, assimilate the explicit number system (Geary, 2013) which involves understanding the relationships between numbers. Previously, a child knew that 4 is larger than 2, but in this stage, a child knows that 4 is 2 more than 2. This skill is also needed for composing and decomposing magnitudes and learning basic arithmetic (cf. Geary, 2013; Krajewski & Schneider, 2009).

The suggested model of early number skills development presented in Fig. 1 combines and summarizes well-known theoretical models. The aim of Fig. 1 is to depict the typical development of early number skills, the different processes required for each developmental step, and their approximate connection to chronological age.

Researchers still debate the causality of the developmental steps, which are known to be affected by many factors (Kaufmann et al., 2013; von Aster, 2000), such as specific attentional or cognitive skills (cf. Geary, 2013; LeFevre et al., 2010). For example, children's own spontaneous focusing on and natural interest in quantities and numerical information observable in daily situations contributes to early number skills development, arithmetic skills, and achievement in school mathematics (recently Hannula-Sormunen, Lehtinen, & Räsänen, 2015; Hannula, Lepola, & Lehtinen, 2010). Home numeracy (i.e., the early mathematical experiences and activities offered at home) also supports number skills development and predicts performance in early mathematics (e.g., Kleemans, Peeters, Segers, & Verhoeven, 2012; Skwarchuk, Sowinski, & LeFevre, 2014). Overall, deficits (as well as strengths) in early number skills seem to be linked to genetic, developmental, cognitive, and environmental factors. Thus, many factors need to be taken into account when evaluating individual's typical and atypical development, supporting skills and strategy development, and understanding the nature of their responsiveness to interventions (see also Rubinsten & Henik, 2009; von Aster & Shalev, 2007).

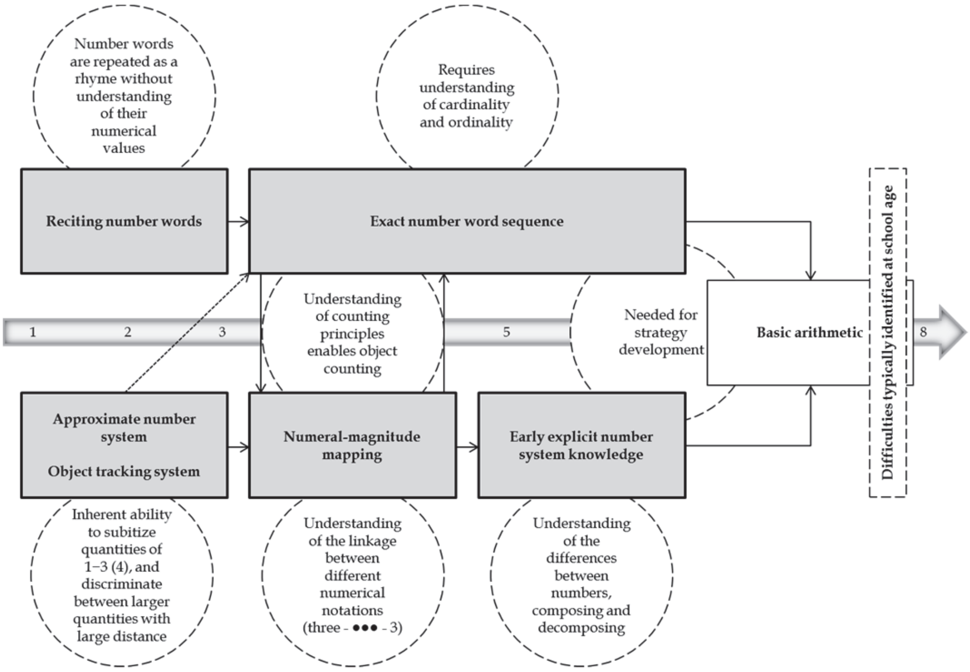

FIGURE 1 Model of early number skills development based on the literature (Dehaene, 2011; Geary, 2013; Krajewski & Schneider, 2009; von Aster, 2000; von Aster & Shalev, 2007). The developmental steps are shown in rectangles and the processes within or between the steps are shown in circles. The approximate chronological age is depicted on the background

1.2 Risk for mathematics difficulties

Various estimated prevalence ranges of mathematics learning disability (MLD) for school-age children have been presented: 3–7 % (Landerl & Moll, 2010), 5–7 % (Butterworth et al., 2011), and 5–8 % (Geary, 2011b, 2004). These children's performance-level is clearly discrepant from the level of their age-peers, and they might manifest different types of difficulties in mathematics (Geary, 2004).

Diagnosis of a specific learning disorder causing impairment in mathematics might be made if a child shows persistent difficulties in number sense, memorization of arithmetic facts, accurate or fluent calculation, or accurate mathematical reasoning (DSM-5 – APA, 2013). Difficulties manifest as low awareness or weak intuition of numbers, inaccurate or slow reasoning or recall of arithmetic facts which cannot be explained by developmental, neurological, sensory, or motor disorders (DSM-5 – APA, 2013). Diagnosis is based on test scores, teacher observations, documentation of children's response to academic interventions, and on the children's developmental, medical, educational, and family history (DSM-5 – APA, 2013). In the literature, "number sense" is defined in various ways: from more narrow definition referring to initial intuition of numbers to a wider definition of general capacity to fluently process the numbers and utilize the knowledge in mathematical activities (cf. Berch, 2005). In this thesis, early number skills development is described as separate numerical sub-skills and the term “number sense” is not used.

According to the International Classification of Diseases (WHO, 2015, ICD-10) a specific arithmetic impairment is seen as deficits in computational skills. This impairment typically involves difficulties with addition, subtraction, multiplication, and division, rather than more complex mathematical skills (e.g., algebra, trigonometry, geometry, or calculus). This impairment cannot be explained by general cognitive delays, mental retardation or inadequate instruction (WHO, 2015, ICD-10, F 81.2.).

Research conducted in past decades has relied on different definitions and cut-off points for MD (Dowker, 2005; Geary, 2004; Gersten, Jordan, & Flojo, 2005). Recently, descriptions of MD have become more specific, likely due to increased understanding of the individuality in factors behind learning problems (e.g., Geary, 2011b). Researchers use, however, varied terminology to define MD. The most common terms in literature are mathematics (learning) difficulties (i.e., MD or MLD), mathematics disabilities or deficits, or developmental dyscalculia (DD). Dyscalculia has been used to refer more specifically to deficits in calculation skills.

This thesis focuses on the poorest-performing children at the pre-primary stage (i.e., 6–7-year-olds) who perform below the 10th percentile as compared to age-level. At that age, Finnish children do not receive diagnoses for either math or reading difficulties. However, according to several longitudinal studies, the differences in the early number skills between low- and typically-performing children is notable, and this gap tends to grow in the following years (the Matthew-effect) (e.g., Aunola et al., 2004; Desoete & Grégoire, 2006; Jordan et al., 2007; Kohli et al., 2015). Children with low early number skills at the beginning of kindergarten (pre-primary education in the Finnish context) do not seem to benefit from instruction in the same way as their age-peers (Aunola et al., 2004; Jordan et al., 2006). Two risk-groups have been identified based on developmental growth-rates in early mathematics achievement (e.g., Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Morgan et al., 2009, 2011; Murphy et al., 2007). First, low-achieving children (performance between the 11th and 25th percentiles) differ from their typically-achieving age-peers, and the gap between these performance-level groups tends to remain stable, and might not increase over time (see Morgan et al., 2009, 2011; Murphy et al., 2007). Second, the children with the most severe difficulties (the poorest-performing children; performance below the 10th percentile) seem to fall even further behind both low- and typically-achieving age-peers in later primary grades (see also Salaschek, Zeuch, & Souvignier, 2014).

In this thesis, the terminology used to describe the target participants' high risk for MD is based on their poor performance in early number skills contrasted to age-level performance. Hence, the following definitions specify that the participants are not yet diagnosed as having MD. A distinction is drawn between the target children (below the 10th percentile) and low-performing children (between 11th and 25th percentiles). The former group of children is referred to using the following interchangeable terms: children most at risk for MD, children with the most severe difficulties, and the poorest-performing children (see Fig. 2). The latter group of children is referred to using the following interchangeable terms: children at risk, low-performing, or low-achieving children (see Fig. 2).

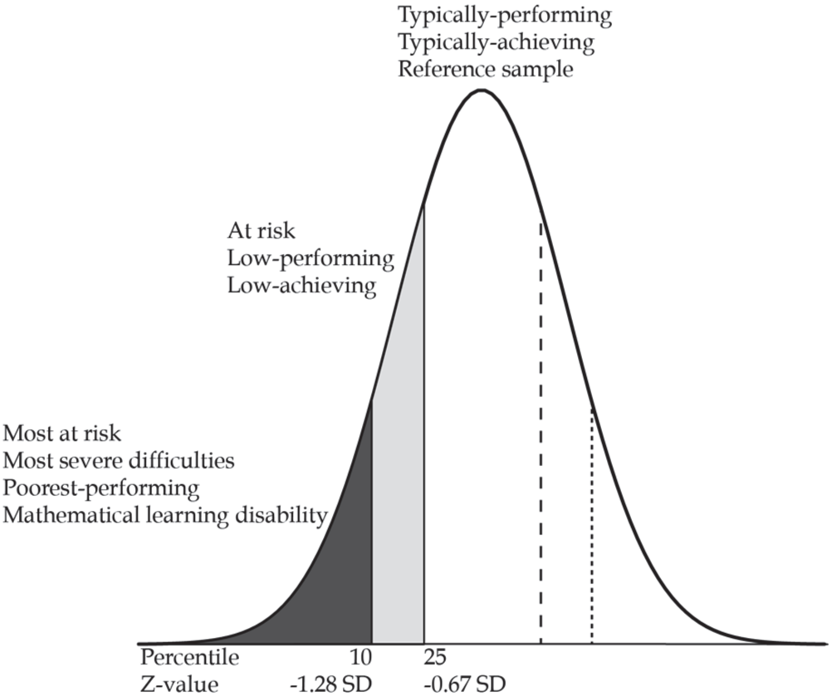

FIGURE 2 Terminology describing risk levels for mathematics difficulties

Prior studies have established that children with the most severe difficulties in mathematics differ from their low-achieving and typically-achieving age-peers by having more serious deficits in processing numbers, learning arithmetic procedures, retrieving arithmetic facts, and working memory capacity (Geary, 2011b). Number processing skill refers to the ability to map between different numerical representations, especially between representations of non-symbolic and symbolic magnitude. Two explanations for impairments in understanding and processing numerical magnitudes have been proposed. First, the children with serious MD have deficits in the innate ability to process quantities, which could be associated with difficulties learning number symbols and basic arithmetic (the defective number module hypothesis; e.g., Butterworth, 2005; see also Piazza et al., 2010). Second, the children with MD have deficits in "accessing" numerical meaning from number symbols (the access deficit hypothesis; De Smedt & Gilmore, 2011; Rousselle & Noël, 2007; for a review, see De Smedt, Noël, Gilmore, & Ansari, 2013). Evidence also indicates that children with serious MD have deficits in both non-symbolic and symbolic processing (e.g., Landerl, Fussenegger, Moll, & Willburger, 2009; Wong, Ho, & Tang, 2015).

Arithmetic procedures demanding conceptual and procedural knowledge are used during the response process to find the correct solution for an arithmetic problem (cf. Dowker, 2009a). Retrieving arithmetic facts requires the ability to quickly access a restored correct solution from long-term memory, or to work it out based on associations with similar type of problems (e.g., Geary, 2004). Deficits in language and visuospatial skills (e.g., Geary, 2004), working memory and other domains of executive functioning (cf. Best, Miller, & Naglieri, 2011; Bull & Scerif, 2001; Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013) might also contribute to difficulties using, controlling, or changing sufficient counting and arithmetic strategies during the problem solving process (e.g., Gelman & Gallistel, 1978; von Aster & Shalev, 2007). However, these skill domains and arithmetic strategy development were not the focus of this thesis, and the literature on these issues is not reviewed further. In Fig. 3, however, the request for working memory capacity is shown as a growing triangle. Evidence indicates that the basic and arithmetic skills do not easily become automatized in children with (or at risk for) MD, and therefore, working memory requirements most likely increase when arithmetic procedures becomes more demanding (cf. von Aster & Shalev, 2007).

It has been proposed that different MD sub-types with difficulties in certain domain-general (i.e., non-numerical cognitive skills) and domain-specific (i.e., numerical skills) skills exist (e.g., Bartelet, Ansari, Vaessen, & Blomert, 2014; see also Skagerlund & Träff, 2014). The findings of these sub-types raise the need for specific, sensitive assessment and targeted intervention methods (Butterworth et al., 2011; De Smedt & Gilmore, 2011; Dowker & Sigley, 2010; Geary, 2011b; Pennington, 2006; Rousselle & Noël, 2007; Willcutt et al., 2013) for clinical, special and general educational settings. More evidence, however, is needed due to contradictory results and a lack of replications. For example, some recent studies have documented an association between approximate non-symbolic and approximate symbolic comparison skills (e.g., Toll, van Viersen, Kroesbergen, & van Luit, 2015), whereas other studies have found that symbolic numerical processing skills, not non-symbolic processing skills, predict mathematical competence (for a review, see Schneider et al., in press) or more specifically, the level of arithmetic proficiency (e.g., Bartelet, Vaessen, Blomert, & Ansari, 2014). These contradictory findings might be explained by the task types or test-batteries used (cf. Schneider at al., in press; Wong et al., 2015), sampling issues, sample characteristics, children's age (cf. De Smedt et al., 2013), or other unknown factors (i.e., other moderator and/or mediator effects).

In sum, the deficits in the number skills of the poorest-performing children seem to be highly persistent (e.g., Chong & Siegel, 2008; Geary, Hoard, & Bailey, 2012; Geary, Hoard, Nugent, & Byrd-Craven, 2008; Morgan et al., 2009, 2011; Murphy et al., 2007). Proficiency-levels in critical early number skills distinguish the poorest-performing children from low-performing ones (Geary, 2011b; Kaufmann et al., 2013). Additionally, the children with (or at risk for) MD are likely to form a heterogeneous group in terms of their neurocognitive deficits (e.g., Ansari, Holloway, Price, & van Eimeren, 2008; Butterworth et al., 2011; Dehaene, 2011; Geary, 2004; Fletcher, Lyons, Fuchs, & Barnes, 2007; Rubinsten & Henik, 2009; von Aster & Shalev, 2007), domain-general and domain-specific deficits (e.g., Kaufmann et al., 2013; von Aster, 2000), potential comorbidity problems (e.g., Jordan et al., 2003; Kleemans, Segers, & Verhoeven, 2011; Landerl & Moll, 2010; Moll, Kunze, Neuhoff, Bruder, & Schulte-Körne, 2014; Pennington, 2006), and responsiveness to intervention (Fuchs et al., 2015; McMaster, Fuchs, Fuchs, & Compton, 2005). Due to the variation in these factors underlying deficits and the diverse nature of MD themselves, identification should be multifaceted to uncover the specificity of the child's difficulty (see Fig. 3).

Determination of risk should be based on the three following criteria: 1) children have identified deficits in early number skills (e.g., Geary, 2004), 2) their performance-level remain stable throughout repeated assessments as compared to age-peers (Geary, 2011b, 2013), and 3) they show poor or non-existing intervention-related growth rate (Fletcher et al., 2007; Fuchs & Fuchs, 1998; McMaster et al., 2005). At an early stage, it should also pay attention to spontaneous focusing on numerosities (Batchelor, Inglis, & Gilmore, 2015; Hannula-Sormunen et al., 2015) and the home numeracy environment (Kleemans et al., 2012; Skwarchuk et al., 2014), both of which have been found to influence number skills development. All these dimensions could and should be detected during the pre-primary education (see Fig. 3).

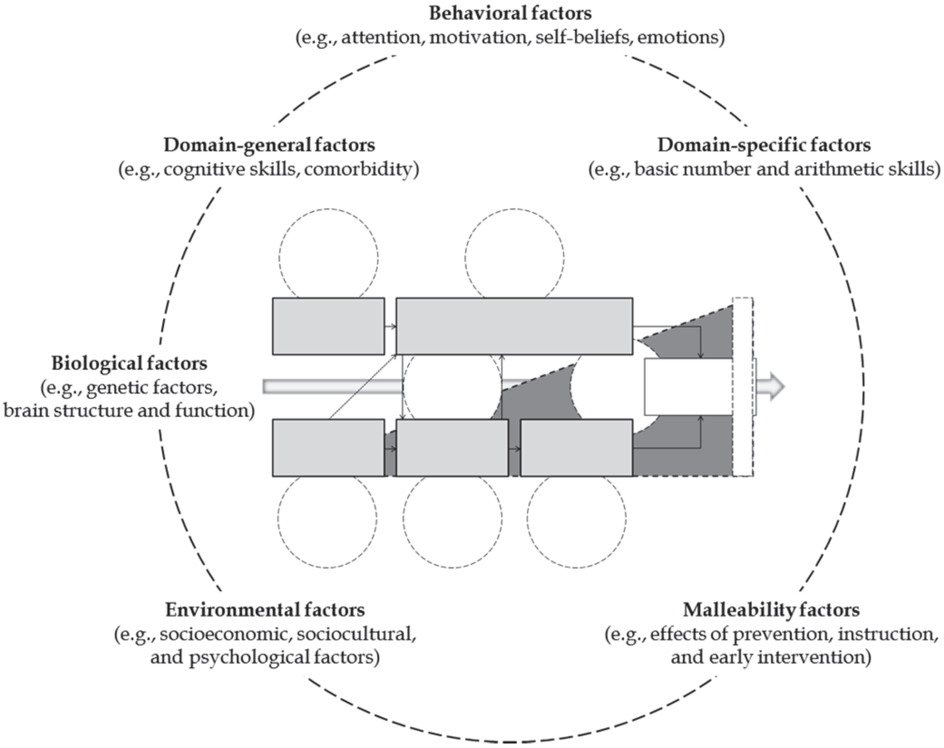

FIGURE 3 Factors affecting early number skills development

Evidence indicates that children with low socio-economic status (SES) exhibit lower spontaneous development during the kindergarten year (Jordan et al., 2006) and that the growth rate of their math skills decreases in first grade (Jordan et al., 2007) and later grades (McClelland, Acock, & Morrison, 2006). Children from low SES families tend to be over-represented among those diagnosed as having MD (Jordan & Levine, 2009). However, in some studies, early performance level has been shown to predict later achievement regardless of initial SES level (Duncan et al., 2007). In the Finnish context, the association between parents' SES and children's learning outcomes tends to be small (Sahlberg, 2007), and the documented association between SES and performance in early (Aunio & Niemivirta, 2010) and later number skills (cf. Hirvonen, Tolvanen, Aunola, & Nurmi, 2012) is also small.

Maturation and chronological age might influence children's response to pre-educational instruction. Older children tend to show more improvement than younger children (DiPerna, Lei, & Reid, 2007; Jordan et al., 2007, Jordan et al., 2006) and perform better in some numerical sub-skills (e.g., verbal counting skills; Aunio & Niemivirta, 2010) and domain-general skills associated with later mathematics achievement (cf. Kurdek & Singlair, 2001).

The results concerning gender-differences in performance level in early number skills are contradictory. Although there seem to be no evidence of clear differences in girls' and boys' early number skills performance, there tends to be more variance among boys (e.g., Aunola et al., 2004). Boys are often over-represented in low-achieving children at the beginning of kindergarten year but not in primary grades (Lachance & Mazzocco, 2006). However, some studies have shown no gender effect on early number skills performance (e.g., Aunio & Niemivirta, 2010), and no gender differences in predictions to later achievement level (Duncan et al., 2007; Jordan et al., 2006, 2007; Lepola, Niemi, Kuikka, & Hannula, 2005; Mazzocco & Thompson, 2005). Finally, the variation in findings concerning gender differences seems to be dependent on the analyses methods used (Devine, Soltész, Nobes, Goswami, & Szücs, 2013). However, the frequency of girls and boys defined as having DD or MD does not suggest under- or over-representation (e.g., Devine et al., 2013; von Aster, 2000).

1.3 Computer-assisted intervention

Computers have been utilized in learning settings for several decades. Since the 1960s, computer-assisted instruction has been seen as an effective method to evaluate skills, individualize learning, and monitor achievement level in mathematics (e.g., Suppes, 1968; Suppes & Jerman, 1969; Suppes & Morningstar, 1969). The number of methods and their level of specificity have increased. The rationale for using computers in education varies form broad approaches, such as computer-based education and instruction, to computer-managed or computer-enriched instruction, to more specific computer-assisted instruction (Cotton, 1991). Computer-assisted instruction refers to the use of tutorials, drill and practice, and programs offering direct instruction (Li & Ma, 2010). In this thesis, the term computer-assisted intervention (CAI) is used instead of computer-assisted instruction to refer to brief, intensive training using a specific computerized method based on early number skills development and adaptive educational intervention theories and incorporating a more pedagogical feedback system than usual in computer-assisted instruction.

Meta-analyses and review studies have found evidence of positive intervention effects on mathematics learning from using short and intensive computerized practice with children with special educational needs (Kulik & Kulik, 1991; Kroesbergen & Van Luit 2003; Li & Ma, 2010; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009; Schmidt, Weinstein, Niemic, & Walberg, 1985–86). In general, the opportunity to concentrate intensely on specific training content (Kroesbergen & Van Luit 2003) with extensive repetition and continuous feedback (Hasselbring, 1986) offered by assistive tutorials (Kulik, Kulik, & Bangert-Drowns, 1985) has been seen as a benefit of CAI. In addition, CAI offers a motivating (Becker, 1992; Kulik & Kulik, 1991), activating (Chambers & Sprecher, 1980), and individually supportive environment for learning (Li & Ma, 2010; Slavin & Lake, 2008).

However, the variation in study designs, sample sizes, participants' characteristics, and intervention density in terms of duration or intensity, assessment tools (standardized or unstandardized), intervention contents, as well as variation in reporting descriptive statistics causes several methodological problems in meta-analyzing and reviewing the effectiveness of CAI in children having (or at risk for) MD (e.g., Cheung & Slavin, 2013; Leskinen & Salminen, 2015; Räsänen et al., 2009; Seo & Bryant, 2009; Slavin & Lake, 2008). These problems increase a need to pay attention to the quality of research methods and publications and interpretations of the effects. In addition, it seems that the average annual gain in learning mathematics varies between grade levels and performance levels (Bloom, Hill, Black, & Lipsey, 2008). This might explain why the strongest effects of CAI are found in children at primary school (Kulik et al., 1985) and kindergarten (Fletcher-Flinn & Gravatt, 1995). If the expected gain depends on the sample characteristics, the effect size estimations should be interpreted in relation to the target group characteristics (Bloom et al., 2008; Hill, Bloom, Black, & Lipsey, 2008; see also Grissom & Kim, 2012). At the moment, more evidence is needed in order to do so.

Intervention research on computerized support for early number skills development among low-achieving children is at an early stage, and no previous studies have focused on kindergarteners most at risk for MD. Among low-achieving children, by matching quantities to number symbols and vice versa, positive intervention effects have been found for number recognition (McCollister, Burts, Wright, & Hildreth, 1986) and dot counting skills (Ortega-Tudela & Gómez-Ariza, 2006). After approximate comparison and exact object counting training, transfer effects were seen in arithmetic skills (Praet & Desoete, 2014). Improvement in symbolic number comparison has been reported after practicing approximate and exact magnitude comparison with training of number neighbors and basic arithmetic (Räsänen et al., 2009), and practicing approximate magnitude comparison skills with concrete dots, number symbols, and basic (symbolic) arithmetic (Wilson, Dehaene, Dubois, & Fayol, 2009). In addition, the practice of basic addition, addition and subtraction, addition and multiplication facts, as well as mixed practice of different arithmetic contents has been found to enhance arithmetic skills (Christensen & Gerber, 1990; Fuchs et al., 2006; Okolo, 1992; Mevarech & Rich, 1985).

Recently, CAI applications have become more precise in terms of targeting the training on the theory-based deficits behind MD. Positive improvements in arithmetic have been found by linking subitizing, dot counting, number recognition, transcoding between different numerical representations and basic arithmetic (Baroody, Eiland, Purpura, & Reid, 2013; see also Baroody et al., 2012). Both former intervention studies were carried out in educational context in children at risk for MD (performance below the 25th percentile), and were aimed at supporting the hierarchy in early number skills development.

Based on the theory of deficits in approximate magnitude comparison within children having (or at risk for) MD (cf. Butterworth, 2005; Dehaene, 2011; von Aster & Shalev, 2007), a specific Number Race (NR) game has been developed (Wilson, Dehaene et al., 2006). This game supports and automatizes number processing and mental number line skills. It also aims at enhancing counting, basic addition, and subtraction skills. Intensive practice with NR has resulted in gains in basic subtraction skills (Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006).

Basic arithmetic skills have also been enhanced by training with the specific Rescue Calcularis program (Kucian et al., 2011). It requires placing the magnitudes of dots and digits or sums and differences on their representative places in a number line (Kucian et al., 2011; von Aster & Shalev, 2007). The goal of this program is to improve the access to and flexible use of the mental number line needed for arithmetic thinking (cf. von Aster & Shalev, 2007). The program is based on the triple-code model of number processing (Dehaene, 1992, p. 31, Fig. 5); it strengthens the links between verbal (number words), Arabic (Arabic number symbols), and analogue modules (magnitude representations as depicted on an internal number line). The Calcularis program has the same elements as the Rescue Calcularis but provides more comprehensive training with flexible adaptation based on the learner's progress and the difficulty level of the content (cf. Käser et al., 2011; Käser et al., 2013). There is also evidence of the effects of Calcularis on arithmetic skills (Käser et al., 2013). The three latter specific intervention methods, and the aforementioned experimental studies on them, have been conducted among 7–11-year-old children diagnosed as having MD or DD. These studies vary with respect to methodology, identification of MD, and intervention procedures (e.g., some of the interventions have been carried out at home), thus, more studies are needed to replicate the positive results and to provide more evidence for the development of theory-based CAI programs. Evidently, there is also a call for knowledge on specific assessment and applicable CAI methods targeted for children most at risk for MD.

1.4 GraphoGame Math

GraphoGame Math (GGM) has originally been designed as part of the GraphoGame project at the University of Jyväskylä in Finland (see Richardson & Lyytinen, 2014). The purpose of this project is to develop an effective method for learning to read. To study the intervention effects of the GraphoGame reading game, GraphoGame Math (GGM) was programmed as a control game. Due to this purpose, the first version of GGM was analogous to the reading game, so its content was somewhat limited. The layout, adaptation, game logic, and feedback system were similar to those in the original reading software.

The further development of GGM has been one of the main purposes of the LukiMat-project coordinated by Niilo Mäki Institute and funded by the Finnish Ministry of Education and Culture since 2007 (www.lukimat.fi). GGM is intended to be a theory-based intervention method for supporting early number skills. The development of GGM is based on the literature and intervention studies presented in this thesis. The purpose of sub-study I (conducted in 2007) was to examine, whether positive intervention effects on early number skills could be seen in the target group of children most at risk for MD after intense training with the original version of GGM or Number Race (NR; Wilson, Dehaene et al., 2006; Wilson, Revkin et al., 2006; see description in section 3.2.3). Afterwards, the layout, numerical content, conceptual knowledge, explicit instructions, and general game logic of GGM were modified (see the following sections). Since the original version of GGM did not produce significant intervention effects on basic arithmetic skills, a stronger conceptual basis was drawn for the second version of GGM before letting a child proceed to procedurally oriented contents. The effectiveness of training with the second version of GGM was assessed in sub-study II (conducted in 2008–2009). Then, the tutorial component (including conceptualization and practice trials), assessment component (addressing the suitable number area for an individual child), content-based adaptation, and feedback system of GGM were further developed. Sub-study III (conducted in 2011) assessed the effects of the third version of GGM. The descriptions of different GGM versions are reported below.

The original version of GGM. The adaptation of the original version of GGM, used in sub-study I (2007) was based on three issues. First, the game content was designed to become gradually more challenging. The training started with simple non-symbolic comparisons (e.g., bigger / smaller, the biggest / the smallest, more / less, the most / the least) and dot counting. It progressed to mapping (i.e., linking number word-quantity-number symbol correspondence) and basic arithmetic training. The number range was set to expand from 1–3, through 1–6, 1–10, and 10–20, to 20–30. Second, the amount of response alternatives depended on the child's success. The better the child performed during practice, the more response alternatives were presented on the screen, limited to a maximum of 6. Third, the algorithm guiding the presentation of trials was preset to keep the individual accuracy rate at approximately 85 %. A description of all the numerical content and mathematical concepts in the first version of GGM can be found in Appendix 1.

During GGM practice, the child was asked to select the correct stimulus corresponding to an auditory cue. One correct stimulus was presented among incorrect response alternatives. Both the correct and incorrect options were presented as dots, number symbols, or basic arithmetic facts on balls descending from the top to the bottom of the screen. For a successful individual trial, the child immediately received positive auditory and visual signals of giving a correct response. Conversely, the child received immediate negative signals when selecting an incorrect response. In addition to this immediate continuous feedback, the child received delayed feedback through different colored butterflies indicating the child's success rate in a game level. The child's total success rate was shown as a progressive bar at the bottom of the screen. Total playing time was also shown as a bar after the child completed each game level.

*The second version of GGM***.** The content of the second version of GGM, used in sub-study II (2008–2009), was based on the findings of previous studies. First, longitudinal studies on the predictive value of early number skills for school-age mathematics achievement were analyzed to determine the importance of core early number skills. Second, studies on effective intervention principles and components were applied to adapt GGM's manuscript and game elements.

Third, the development of GGM took into account research on the layout and usability of educational software designed for the children having (or at risk for) MD. The studies used in modifications of the second version of GGM in 2008 are referenced in this section. The literature that was analyzed specifically for this compilation thesis (i.e., more recent evidence) is indicated by inclusion of the word "recently" in parenthesis with the citation.

As known, understanding of cardinality and counting principles is critical in early number skills development. For example, object counting fluency seems to remain rather stable between different performance-level groups (recently Reeve et al., 2012). In several longitudinal studies, this skill has been evaluated by measuring the accuracy of object counting that requires understanding of one-to-one correspondence, stable-order principle, and other counting principles (Desoete & Grégoire, 2006; Geary, Hamson, & Hoard, 2000; Jordan et al., 2007; Jordan et al., 2006; Mazzocco & Thompson, 2005). Understanding of ordinal numbers and the ability to order sets of objects predict later math achievement (Duncan et al., 2007) and, more specifically, counting fluency (Koponen, Aunola, Ahonen, & Nurmi, 2007), so ordering tasks were added to the second version of GGM. Finally, early numerical competence (i.e., a sense of quantities and numbers) predicts calculation fluency (Locuniak & Jordan, 2008) and applied problem solving skills (recently Jordan, Kaplan, Ramineni, & Locuniak, 2009). For this reason, various types of comparison tasks were included in GGM (e.g., non-symbolic and symbolic approximate and exact comparison tasks, such as more / less, the most / the least, one more / one less, two more / two less).

Understanding the correspondence between number words, quantities and number symbols and the relationship between different quantities and numerals seem to be critical skills in early number skills development (recently Geary, 2013; Krajewski & Schneider, 2009). Several longitudinal studies found that these skills predict mathematical achievement in primary school (Desoete & Grégoire, 2006; Jordan et al., 2006; Mazzocco & Thompson, 2005; McClelland et al., 2006) and are associated with school-age arithmetic skills (Koponen et al., 2007). As mentioned, GGM included several types of comparison tasks which required understanding of correspondences between different numerical representations. Explicit mapping tasks (i.e., number word-quantity-number symbol correspondence tasks) were also designed to support this skill.

Verbal counting has commonly been observed to be a core early number skill and a strong predictor of later mathematics skills (e.g., recently Aunio & Niemivirta, 2010; Aunola et al., 2004; Kurdek & Sinclair, 2001). The association of early verbal counting skills with primary school-age mental (Desoete & Grégoire, 2006; Lepola et al., 2005) and written calculation skills (Koponen et al., 2007; recently Koponen et al., 2013) has also been recognized. Implementing pure verbal counting in GGM was not possible, so this skill was intended to be supported by practicing number sequence knowledge. The child, for example, was asked to place numerals according to their numerical values along a number line with only the starting and ending points indicated. The number ranges used in this training in GGM (e.g., 0–5; 6–10) were more concise than those in, for example, number line tests with the greater distance between the starting and ending points (0–100, 0–1000) (see Siegler & Booth, 2004; Siegler & Opfer, 2003). Number sequence skills also predict later mathematics achievement (more recently Geary, 2011a; Jordan et al, 2007; Jordan et al., 2006).

Finally, performance in early arithmetic skills seems to uniquely predict later mathematics achievement (recently Geary, 2011a; Jordan et al, 2007; Jordan et al., 2006; Mazzocco & Thompson, 2005) and, more specifically, mental (Lepola et al., 2005) and written calculation skills (Desoete & Grégoire, 2006). Due to the current study's target age-group of children, only basic addition training was implemented in GGM. The task types included composing, decomposing, verbal story problems, and mental and written addition. A description of all the numerical content and concepts included in the second version of GGM can be found in Appendix 2.

Previous findings of pedagogically relevant and effective intervention (non-computerized) programs were analyzed to identify meaningful ways to address the numerical content supposed to be learned. The "Young Children, with Special Educational Needs Count, Too" program (van Luit & Schopman 2000) helped children (5–7-years-old) learn to count and transition from early to basic mathematics instruction in interactive, small-group learning situations. The essential feature of this program was the repetitive enhancement of different number representations, starting with concrete (e.g., 4 objects) and proceeding to semi-concrete (e.g., 4 dots) and abstract representations (e.g., the number word "four" and number symbol 4). The Number Worlds Program (Griffin, 2004, 2007) was developed for children from preschool through grade 6. The main purpose of this program was to ensure that children achieve the conceptual understanding of mathematical content expected at their age and grade level. The program was intended to help children broaden and deepen their understanding of the relationships between different numerical worlds, such as the worlds of counting numbers, quantities, and formal symbols (Griffin, 2004, 2007). Both these programs highlighted the importance of different number representations and therefore, GGM was also developed to support this link.

The Mathematics Recovery Programme (Wright, 2003; Wright, Martland, & Stafford, 2008) serves as an intensive, individualized intervention for children (6–8-years-old) with weak early number skills. The purpose of this program is to continuously assess children's skills and intervention progress profiles and based on these, to support individual gaps. This program was mostly based on similar principles as the aforementioned programs (cf. Griffin, 2004, 2007; van Luit & Schopman 2000), but to strengthen the number sequence skills; number neighbors were reinforced by training for knowledge of "numbers after" and "numbers before" on specific target numbers (Wright, 2003). Similar training contents were also added to GGM. Additionally, "base 5 strategies" (children are taught to utilize the amount of 5 as a base to form sets of objects, or to compose and decompose quantities or numerals larger than 5) were strengthened in

GGM to increase the meaningfulness and power of 5 (cf. Wright, 2003; Wright et al., 2008).

The Numeracy Recovery intervention program (Dowker, 2001, 2008b) concentrates on counting procedures and principles, arithmetic estimation and fact retrieval, word problem solving skills, and supports understanding of place value. The approach of word problem solving was embedded into GGM to develop conceptual understanding of basic addition, followed by verbal story problem solving and single-digit addition skills. In this approach, a simple word problem and the correct response are presented orally, and immediately afterward, the child is asked to pick the correct response (or stimulus) which is presented among incorrect ones. Next, a verbal story problem and the correct response are presented, and then, the child is required to remember one of two addends as a correct alternative. Finally, a problem is given without the resulting sum, and the child is asked to select the correct response.

The principles and components of effective numerical intervention for children having (or at risk for) MD recommend the use of explicit instructions and multifaceted, repetitive, massive practice of basic concepts with immediate, continuous, corrective, and rewarded feedback (Baker, Gersten, & Lee, 2002; Fuchs, Fuchs, Powell et al., 2008; recently Gersten et al, 2009; Moreno, 2004). Along with these principles, step-by-step progression from easy to more complex content (Clements, 2004; Dowker, 2001; Sarama & Clements, 2009; Seo & Woo, 2010), from concrete to semi-concrete and finally, to abstract numerical representations (cf. Griffin, 2004; van Luit & Schopman, 2000) were embedded into GGM. Similarly, the basic addition content progressed from the easiest single-digit problem types towards more complex ones (Riley, Greeno, & Heller, 1986). Before actual training, GGM offered six practice trials to familiarize the individuals with the task requirements (cf. Baker et al., 2002). GGM's game-like environment was thought to support motivation for practice and maintain the task-orientation in the target group of children most at risk for MD (cf. Baker et al., 2002; Fuchs, Fuchs, Powell et al., 2008; recently Gersten et al, 2009). To integrate different early number skills, instead of addressing them separately (recently Fuchs et al., 2012), the GGM content was organized hierarchically so that the child was systematically told what had been practiced previously and what would be practiced next (e.g., "We have practiced the number names. Now we will learn how the numbers can be presented as quantities and written number symbols.") These instructions were given when the child passed a certain subskill area and was starting the next one.

The literature guidelines for instructions, user interface design, and layout were followed when writing the manuscript of the second version of GGM (e.g., Carnine, 1997; Okolo, 1991; Swanson & Hoskyn, 1998; recently Seo & Woo, 2010). Adhering to these guidelines, visual representations and graphics were designed to attract the child's attention and ease understanding. The screen view was designed to be simple: the relevant stimuli were big, clear, and centered on the screen. There were no moving stimuli or distracting activities or music. The fonts and colors used in the tasks were clear, and the background images were faded. The child was given only one instruction a time, presented aurally because most kindergarteners cannot read yet. Finally, the relevant information was stressed to aid orientation toward task demands (e.g., "…*equal amount*..."; see Fig. 4).

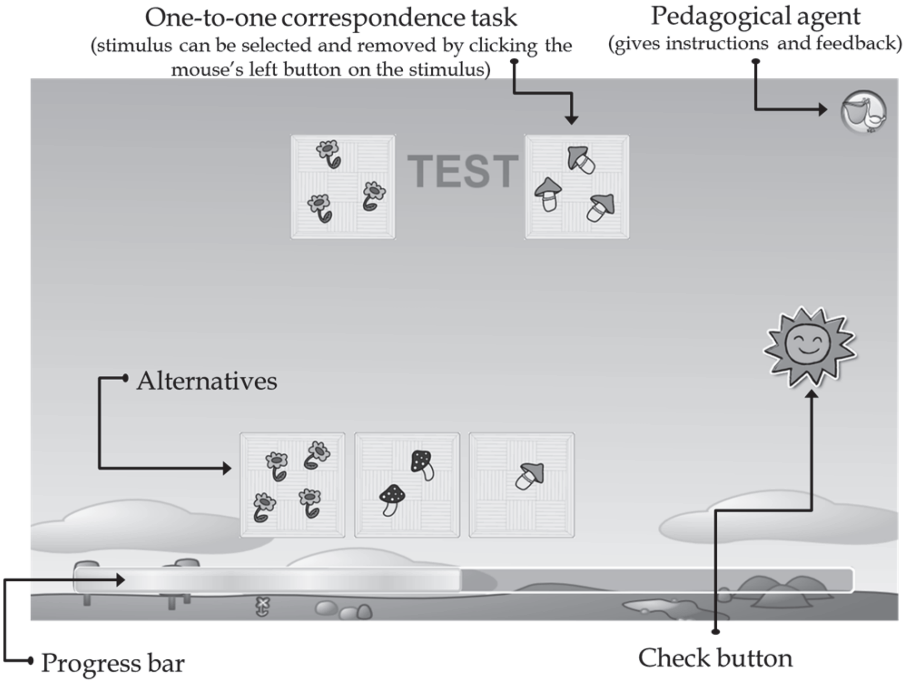

FIGURE 4 Layout example of the second version of GraphoGame Math

The third version of GGM. The third version of GGM, used in sub-study III (2011), differed from the second version in its adaptation logic and feedback system. The content was also divided into three main components. With reference to the model of early number skills development presented in this thesis, the three skill-components relate to Fig. 1 (in section 1.1) as follows: 1) understanding of exact number word sequence is supported by number sequence component, 2) OTS / ANS and numeral-magnitude mapping with early explicit number system knowledge (here exact relations between numbers) are supported by number concept component, and 3) early explicit number system knowledge (here composing and decomposing) and basic arithmetic are supported by basic addition component (cf. Dehaene, 2011; Geary, 2013; Krajewski & Schneider, 2009; von Aster, 2000; von Aster & Shalev, 2007). The studies used in modifications of the third version of GGM are described in this section. The literature that was analyzed specifically for this compilation thesis (i.e., more recent evidence) is indicated by inclusion of the word "recently" in parenthesis with the citation. A description of all the numerical content and concepts included in the third version of GGM can be found in Appendix 3.

To support conceptual knowledge, GGM started with six demonstrative tutorials to explain the main concepts expected to be learned and used in the upcoming training components: number concepts, numbers sequence, and basic addition skills (cf. Dowker, 2009a). The concepts were presented first as automatized tutorials and then included in practice trials while explained aurally by a pedagogical agent (a pelican). Finally, the child solved the practice trials and was instructed how the game logic and checking process work (i.e., how to select a stimulus or change the selected stimulus to another stimulus and how to check a trial and proceed to the next one).

After tutorials in all these components, the training started with content-specific assessment trials (maximum of 32) which determined in which number area the child started training. Eight assessment trials were presented for each of the 4 main number ranges (0–5, 3–7, 6–10, 10–20). If the child solved all 8 trials in the number area of 0–5, the child continued to the next evaluation trials for the number area of 3–7. If the child made an error in the number area of 3–7, the ongoing assessment trial was terminated, and the child started the training again in that particular number area (in this example, 3–7). After successful training in a number area, GGM ensured the child's accuracy level by reassessing the trained content before allowing the child to proceed to the next number area.

The adaptation of the third version of GGM was based on the numerical distance between the correct and incorrect response alternatives (distance effect) (recently Dehaene, 2011). The distance was first set to be large and then gradually reduced. At the same time, the number of required correct responses per individual trial increased from one to three. With this change, the amount of incorrect alternatives decreased, and the child was asked to find one to three correct responses corresponding to the auditory cue. The purpose of these changes was to continuously strengthen, for example, the relationship between the different numerical representations within a single trial, not in separate, consecutive trials as in the second version of GGM. The other principles of adaptation logic were the same as in the previous GGM version (numerical notation, number area, and adaptation algorithm for the success rate).

Based on experiences with earlier versions of GGM, the feedback system was modified to better support the child in proceeding through GGM practice. Specifically, a scaffolding, answer-until-correct feedback system was implemented. Previous GGM versions used immediate, continuous, and corrective feedback (Baker et al., 2002; Gersten et al., 2009; Hasselbring, 1986; Moreno, 2004) with delayed rewards (Fuchs, Fuchs, Powell et al., 2008). For the third version, the dimension of scaffolding was added (Finn & Metcalfe, 2010). The feedback was modified to alter the complexity of the content, and to ease the requirement of cognitive effort by offering clear, specific, and task-focused feedback with cues to facilitate performance and support response process instead of using just correct/incorrect feedback (recently Hunt, Valentine, Bryant, Pfannenstiel, & Bryant, 2015; Fuchs et al., 2008; Shute, 2008). For example: the child was instructed to complete the number sequence of 5, __, __, __, 9 presented at the top of the screen. If the child gave an incorrect response, such as 5, 7, 8, 6, and 9, GGM returned the stimuli to the group of alternatives at the bottom of the screen (immediate feedback) and moved the first correct stimulus to its correct place (in this case, the number symbol 6). This was thought to be a cue that could help the child to find the correct solution (corrective, scaffolding feedback). If the child could not continue the number sequence with this hint and responded incorrectly again, the incorrect stimuli were returned to the alternatives (immediate feedback), and the next correct stimulus (here, number symbol 7) was moved to its correct place (corrective, scaffolding feedback). The child then continued the task. If the child made a third error, the third and final correct stimulus (here, number symbol 8) was moved to the correct place (corrective feedback). Finally, to offer a model for verbal counting, GGM aurally repeated the existing number sequence while highlighting the target number symbol in a sequence (restorative feedback). The aim of this scaffolding feedback was to strengthen knowledge of number sequences, so the child could learn them and proceed to more complex tasks.

2 AIMS

The main purpose of the present thesis was to examine to what extent a computer-assisted intervention (CAI) method (GraphoGame Math, GGM) can support early number skills development in Finnish kindergarteners identified as most at risk for MD. The effects of GGM intervention were studied with respect to theory-based numerical content, domain-general, domain-specific, and malleability factors (Fig. 5).

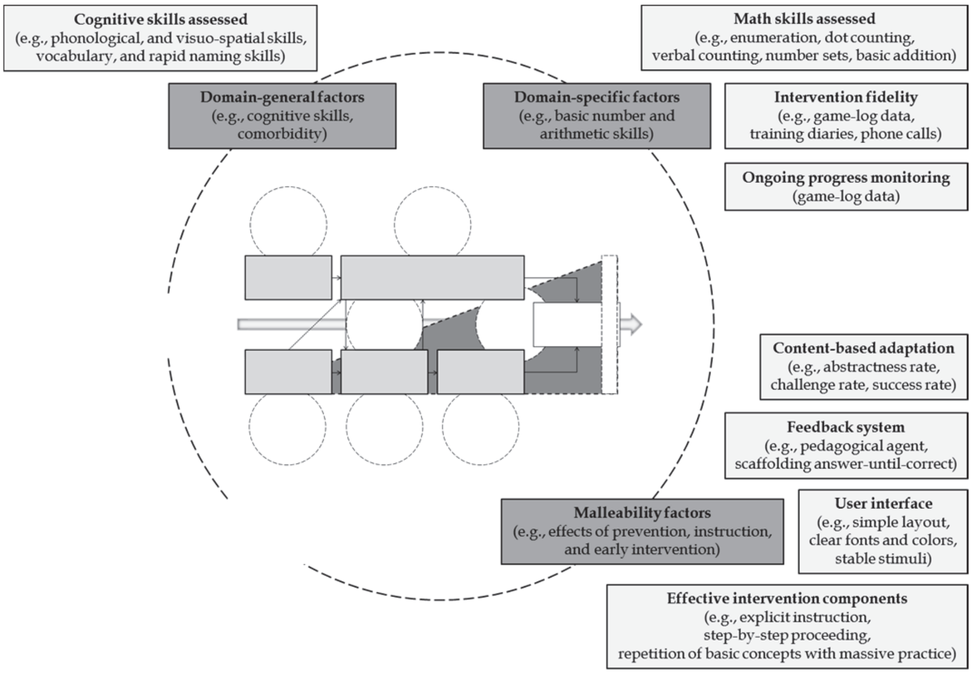

FIGURE 5 Core components of the GraphoGame Math intervention studies

In prior studies, an intensive (e.g., Cheung & Slavin, 2013), targeted (e.g., Dowker & Sigley, 2010; Geary, 2011b), and individually adaptive training

(Fuchs et al., 2012; Slavin & Lake, 2008) has been shown to be efficient in supporting early number skills in children having (or at risk for) MD. In the present study, theory-based content (e.g., Geary 2013; Krajewski & Schneider, 2009), recommended intervention principles and components (e.g., Baker et al., 2002, Fuchs, Fuchs, Powell et al., 2008; Gersten et al., 2009, Li & Ma, 2010), and a meaningful feedback system (e.g., Finn & Metcalfe, 2010; Fuchs, Fuchs, Powell et al., 2008) were implemented in the CAI context.

This thesis concentrated on examining the effects of GGM. Positive intervention improvements were expected from intensifying and individualizing massive, explicit, theory-based practice for the children most at risk for MD whose response to early intervention is unknown. The effects of Number Race game (NR) are not discussed in detail, although the game was used as a control condition in two of the three sub-studies. The specific research questions of the three sub-studies are presented in Table 1.

TABLE 1 Research questions of the three sub-studies

Study	Specific questions
Sub-study I	• How does short and intensive practice with GraphoGame Math or Number Race support the early number skills (verbal counting, object counting or basic arithmetic) in kindergarteners most at risk for MD? • What between-condition differences in early number skills are found in potential intervention gain scores? • Are potential intervention gains associated with total intervention exposure per condition?
Sub-study II	• Does a short and intensive intervention with GraphoGame Math program produce immediate group-level effects in kindergarteners with poor early addition skills? • Does the potential intervention effect remain stable over the nine week follow-up period? • To what extent are there individual differences in the target group's responsiveness to GGM intervention based on game-log data?
Sub-study III	• Can immediate condition specific effects on poor early number skills in kindergarteners most at risk for difficulties in mathematics be found after a short and intensive practice? (The study included two examinations evaluating the effects of different sample characteristics on early number skills.) • Do the effects of two intervention conditions on basic addition skills differ? (The study included two examinations evaluating the effects of different sample characteristics on basic addition skills.)

3 METHOD

3.1 Participants

In the three sub-studies, kindergarten teachers and special kindergarten teachers nominated children in need of acute, extra support in early number skills. The nominations were based on teachers' observations of kindergarteners' skills and low interest in numerical activities. The number of daycare centers which voluntarily participated in data collection was 12, 24, and 14 in sub-studies I, II, and III, respectively. The pool of kindergarteners from which the nominations were requested, ranged from 236 to 350 in the sub-studies.

The proportions of nominated participants who were confirmed to have poor early number skills and were included in the analyses were 7 % in sub-study I (17 of 236 participants), 6 % in sub-study II (21 of approximately 350 participants), 12 % in Examination 1 in sub-study III (33 of 278 participants), and 5 % in Examination 2 in sub-study III (14 of 278 participants). The proportions of the poorest-performing children were comparable to the estimated prevalence of MD, ranging from 3–7 % to 5–7 % to 5–8 % (cf. Butterworth et al., 2011; Geary, 2011b, 2004; Landerl & Moll, 2010). In sub-study II, the total number of children in participating kindergarten groups was unknown. Therefore, the approximation related to the original sample size (350 children) was based on the number of daycare centers (24) and a cautious approximation of the mean of the kindergarten group sizes in Finland (14–15 children).

The studies were conducted in eastern Finland (sub-study I), in southern Finland (II), and in central Finland (III). All participants in the three sub-studies were native speakers of Finnish. None had severe visual, hearing, motor, or intellectual impairments.

All parents whose children were candidates for the three sub-studies were informed of the studies. The parents of participants in individual assessments and CAIs provided written, informed consent. For ethical reasons, all the nominated children were allowed to participate in the individual assessments and intervention cycles, although the skill assessments indicated that some did not perform poorly in early number skills. These children were not included in the analyses. Also, for ethical reasons, all the children at the participating daycare centers were allowed access to the intervention programs after the study if their parents gave permission for that. The parents and kindergarten teachers were also informed of the study results.

In sub-study I, the analyses involved 17 intervention children who performed below the 10th percentile of the reference sample in verbal counting skills. The reference sample was a normative sample of Finnish kindergarteners (*n* = 502). The normative data were collected for a nationally normed assessment test of early number skills (Polet & Koponen, 2011). For example, the majority of intervention children could not count correctly up to 20, count backward from a given number, or skip count by 2s. The final group sizes for the two intervention conditions were 9 for the GraphoGame Math (GGM) condition (7 boys and 2 girls; mean age = 80.1 months, SD = 4.5), and 8 for the Number Race (NR) condition (4 boys and 4 girls; mean age = 78.4 months, SD = 4.1).

In sub-study II, the sample consisted of 21 intervention children who performed below the 7th percentile as compared to the age-level in basic addition (below 1.5 SD; 6.68 %). The level of performance indicated that they failed to solve even the simplest addition problems (e.g., 2 + 1, 1 + 3, 3 + 2). The final group sizes for the two intervention conditions were 13 for the GGM condition (4 boys and 9 girls; mean age = 78.6 months, SD = 5.4), and 8 for the NR condition (5 boys and 3 girls; mean age = 76.6 months, SD = 4.3).

In sub-study III, the total intervention sample consisted of 33 children (6–7-years-old). The first examination was administered to all nominated participants (n = 33), and the second examination to the poorest-performing participants based on individual assessments (n = 14). Inclusion criteria for the second examination required that participant performed below the 10th percentile in at least 2 of the 3 early number skills: verbal counting, dot counting fluency, and basic addition. In Examination 1, the final group sizes for the two intervention conditions were: 17 for the NC-condition in which number concept intervention was administered in cycle 1, and basic addition in cycle 2 (NC-group: 6 boys and 11 girls), and 16 for the NS-condition in which number sequence intervention was administered in cycle 1, and basic addition in cycle 2 (NS-group: 5 boys and 11 girls). In Examination 2, the final group sizes were 8 for number concept intervention in cycle 1, and basic addition in cycle 2 (NC-group: 1 boy and 7 girls), and 6 for number sequence intervention in cycle 1, and basic addition in cycle 2 (NS-group: 2 boys and 4 girls).

Children participating in GGM interventions were comparable across the three sub-studies in terms of their basic number skills, age, pre-primary settings, and socio-economic backgrounds. The majority of the children could not count correctly up to 20, and they failed to count forward and backward from a given number. The children also had difficulties in composing non-symbolic and symbolic quantities and they failed to solve basic symbolic addition tasks (2 + 1).

3.2 Measures and materials

3.2.1 Domain-general measures

In sub-studies I and III, domain-general measures were used to assess the comparability of participants' characteristics in the two intervention conditions (cf. Gersten, Fuchs et al., 2005). In sub-study I, visuo-spatial (Corsi blocks tapping task; Corsi, 1972; Milner, 1971) and phonological working memory (Nonword repetition; Korkman, Kirk, & Kemp, 2008), as well as Rapid Automatized Naming of colors (RAN) (Denckla & Rudel, 1974; standardized Finnish version by Ahonen, Tuovinen, & Leppäsaari, 2006) were used to assess the initial level of children's domain-general skills. In the first pre-test of sub-study III, a shortened version (30 tasks) of the Peabody Picture Vocabulary Test (PPVT – Revised) (Dunn & Dunn, 1981) was administered to control for the initial level of children's vocabulary skills. RAN of objects was used to assess initial level processing speed across intervention conditions. No domain-general measures were included in sub-study II, because the kindergarten teachers administered the data collection, and using these assessment materials would have required specific training.

Spearman's correlation coefficients for test-retest reliability in the domain-general tasks can be found from Table 2. Domain-general measures were used to evaluate the comparability of the two intervention groups. RAN task was also used to assess the specificity of the intervention effects in sub-study I (i.e., if the GGM or NR interventions effect on the general fluency-level in task-performance, improvements would also be expected to be seen in RAN task).

3.2.2 Domain-specific measures

Enumeration task. In the enumeration task, the child was asked to pick up, and put on a platter the number of beads requested (3, 6, 8, 10, 13, and 17) (Salminen, Räsänen, Koponen, & Aunio, 2008c). This task was included in sub-studies II and III. The maximum score for this task was 6.

Verbal counting. Verbal counting skills were assessed with three sub-tests adapted for sub-study I from the Early Numeracy Test (counting forward, counting backward, and skip counting by 2s) (Van Luit, Van de Rijt, & Aunio, 2006) and for the sub-studies II and III from Diagnostic Tests 3 (counting forward, counting forward from a given number, and counting backward from a given number) (Salonen et al., 1994). The maximum score for this task was 12.

Subitizing and dot counting fluency. Subitizing and dot counting fluency were assessed by tasks in which 1–5 (sub-study III; 40 items) or 1–6 (sub-study I; 18 items) black dots were presented in a random arrangement, and the child was asked to give as quickly and accurately as possible the number of dots in each individual item. In sub-study I, a computerized test (E-Prime) was used. In sub-study III, the child gave the number of dots presented in a matrix on a laminated piece of paper (Salminen & Koponen, 2010). In both studies, the number of (in)correctly nominated sets of dots was measured. The time used for either the items (sub-study I) or the task (sub-study III) was also measured. In sub-study I, dot counting fluency was evaluated by the mean of median reaction times to correctly recognize groups of 4–6 dots. In sub-study III, dot counting fluency was assessed by dividing the total time used for the task by the amount of correct responses.

The Number Sets Test. The original Number Sets Test (Geary, Bailey, & Hoard, 2009) was translated into Finnish and used in sub-studies II and III. In this test, the child was required to determine as quickly and accurately as possible whether pairs or trios of object sets and/or Arabic numbers match the given standard numbers (here 5 and 9). The test included four parts: parts A and B for standard number 5, and for standard number 9. Each part consisted of 18 correct and 18 incorrect stimuli presented in random order. Part A involved only objects, while in part B, objects and numerals were mixed. This test measured composing skills and understanding of relations between numbers. The maximum score for this test was 18. Due to the intervention participants' weak performance in this task, only the part A with target number of 5 was used to assess their composing skills.

Story problems. In sub-study III, non-symbolic story problem solving skills were assessed with 3 addition and 3 subtraction tasks. The tester moved, added, or removed the beads under a scarf so that the child could see them. The tester simultaneously explained what she was doing. After each individual task, the child was required to tell how many beads there were under the scarf (Salminen, Räsänen, Koponen, & Aunio, 2008b). The test was used to assess understanding of basic concepts, such as adding and taking away. The maximum score for this task was 6.

Basic arithmetic. In sub-study I, basic arithmetic skills were assessed with a speeded, 3-minute, paper-and-pencil task which included both symbolic addition and subtraction calculations (2 + 1 = __, 4 – 1 = __, 7 + __ = 14, 15 – __ = 9, 3 + 4 + 6 = __, __ – 3 = 10, 16 = 9 + __) (Aunola & Räsänen, 2007). The child was asked to write the missing value for each equation as quickly and accurately as possible within a given time limit.

In sub-studies II and III, basic arithmetic was assessed only with basic addition tasks (45 items). The child was asked to respond orally to vertically presented addition problems as quickly and accurately as possible within a 3-minute time limit (Salminen, Räsänen, Koponen, & Aunio, 2008a). Pseudo-randomly ordered single-digit problems (a + b = __) with a sum of 5 or less were presented first. These were followed by problems with a sum of 10 or less. Highly similar problems did not follow each other (e.g., 2 + 2 = __, 2 + 3 = __), and problems involving adding 0 were not included.

Spearman's correlation coefficients for test-retest reliability in the domain-specific tasks can be found from Table 2. Domain-specific variables with low test-retest reliability were not included in the main analyses.

TABLE 2 Spearman's correlation coefficients for test-retest reliability in the tasks administered in the three sub-studies

	Test–retest correlation		
Variable	Sub-study I	Sub-study II[a]	Sub-study III
Domain-general			
Corsi blocks	NA		
Nonwords	NA		
PPVT			NA
Rapid naming	.83***		.74***
Domain-specific			
Enumeration		.72***	(.54**)
Verbal counting	.73***	.80***	.91***
Subitizing fluency	.76***		.70***
Dot counting fluency	.66***		
Number Sets Test		.82***	(.57**)
Story problems			(.66***)
Basic addition		.90***	.94***
Basic addition and subtraction	.84***		

Note. NA = not available in the current data. Domain-general skills were mostly assessed only in the first pre-assessment. Domain-specific variables with low test-retest reliability (shown in parenthesis) were not included in the main analyses. However, they were used to describe the initial skill level of the two intervention groups to assess their comparability.
[a]Test-retest reliability was tested in a pilot study conducted in the same living area, and similar daycare conditions before the actual study.
*** $p < .001$.

3.2.3 Intervention programs

GraphoGame Math (GGM). The purpose of this thesis was to further develop the original version of GGM and to assess the extent to which it can be utilized for supporting early number skills development in kindergarteners most at risk for MD. Consequently, GGM is described in more detail than the control method Number Race (NR). The development process and different version of GGM are described in Section 1.4 and Appendices 1–3. The second and the third versions of GGM can be downloaded for free by registered users at https://ekapeli.lukimat.fi.

Number Race (NR). The NR computer game was used as another intervention method to examine the effects of CAIs within the target group of children most at risk for MD. The intent of NR is to remediate dyscalculia by supporting quantity representation (Wilson, Dehaene et al., 2006; Wilson et al., 2006).

NR is targeted primarily at children ages 5–8. The specific aims are to enhance and automatize number processing, mental number line, counting, basic addition, and subtraction skills. NR is an adaptive computer game in which the numerical distance, response time, and notation are based on the child's performance level. In the NR version used in the Studies I and II, the number range was 1–9. The game's item-selection algorithm is designed to keep the accuracy rate above 75 %. The updated NR version (open source for multiple languages) can be downloaded for free (http://thenumberrace.com/nr/home.php).

In NR, the child first chooses between two visual game contexts: an underwater world or a jungle. The game logic is identical in both contexts, but the graphics differ. Within the game, the child's first task is to choose the larger of two quantities presented with concrete objects (coins or coconuts), symbols, or basic addition or subtraction calculations (see also Wilson et al., 2009). After making this selection, the child moves game characters on a racetrack according to the selected quantity. The child's character advances the quantity selected, and the character of the opponent advances the quantity left available after the child's selection. The game counts out loud how many steps the characters advance on the track. The child should choose the larger quantities to win (i.e., reach the end of the track first).

NR gives immediate and continuous feedback. Every time the child manages to choose the larger quantity, the child hears the sound of applause, and if the child chooses the smaller quantity, the child hears only a short signal indicating incorrect selection. NR also gives rewarded feedback. Every time the child wins a single track, the child can unlock an animal (fish or butterfly), and after winning seven tracks, the child can unlock new character with which to play.

3.3 Procedures

In sub-study I, the participants were individually assessed by two research assistants at each of the three time points: February, February–March, and April. Each assessment session was held in a quiet, separate room in the daycare center and lasted approximately 20–30 minutes. Assessment tasks were administered in the following order: Corsi blocks, Nonword repetition, verbal counting, subitizing and dot counting fluency, basic arithmetic, and RAN. The aim was to examine whether intervention effects on early number skills could be observed in the group of children most at risk for MD. After two pre-tests, the children were randomly assigned to two intervention conditions: practicing with either GGM or NR. The kindergarten teachers organized the intervention sessions and helped the children log in and out of the intervention games. The children were instructed to play individually with the headphones 12 to 15 times over the 3-week intervention. The intervention was followed immediately by a post-assessment. Each intervention session was held during kindergarten hours and lasted 10–15 minutes. The target practice time was 120 minutes, which was achieved in both intervention conditions. To assess intervention fidelity, the teachers also reported the number of sessions and duration of each session in a practice diary. The study design is described in Table 3.

In sub-study II, kindergarten teachers trained in the assessment procedure assessed the children individually at four time points: October–November, December, January, and February. The individual assessment sessions were held in a quiet room at children's daycare centers and lasted approximately 20 minutes. The assessment tasks were presented in the following order: enumeration, verbal counting skills, Number Sets Test, and basic addition. A waiting-list control design was employed to examine the specificity of the immediate intervention effects of GGM, particularly their stability. After the pre-test, the children were randomly divided into two intervention groups. One group received GGM intervention for three weeks, while the other group did not receive extra training. After this period, the children were individually tested for early number skills. Before and immediately after the second intervention period, the children were again assessed for early number skills. During the second intervention period, the wait-list control group underwent an NR intervention for three weeks, while the GGM group received no extra training. The kindergarten teachers organized the intervention sessions and helped the children log in and out of the intervention games. As in sub-study I, teachers were asked to arrange 12–15 individual practice sessions during normal kindergarten hours. Each session was designed to last 10–15 minutes. Training diaries and individual game-log data on intervention fidelity and performance during practice were collected throughout the study. The study design is described in Table 3.

In sub-study III, participants were assessed individually at five time points (January, January, February, March, and May) by two research assistants trained in the assessment procedure. All assessment sessions took place in quiet, separate room at children's kindergartens and lasted for approximately 20 minutes. The tasks were presented in a fixed order: PPVT (background variable, measured only in the first pre-test), verbal counting skills (outcome variable), the Number Sets Test (numerical background variable; Geary et al., 2009), verbal story problems (numerical background variable), basic addition (outcome variable), dot counting fluency (outcome variable), and RAN of objects (background variable). After two pre-tests, the children were randomly assigned to two intervention conditions in which they practiced either number concept or number sequence knowledge skills with GGM. This was followed by an individual assessment. Next, both groups received basic addition training for three weeks with GGM. The kindergarten teachers organized the intervention sessions, and helped the children log in and out of the intervention games. The kindergarten teachers were asked to arrange 4–5 individual practice sessions weekly (12–15 sessions in total) for each participating child during normal kindergarten hours in both 3-week intervention cycles. Each session was designed to last approximately 10–15 minutes. Intervention fidelity was measured in three ways: 1) teachers reported the number of sessions and their duration in a practice diary, 2) the kindergarten teachers received phone calls to ensure that they followed the intervention procedure and to offer consultation in case of problems, and 3) the total practice performance and actual playing times were analyzed by game-log data. The study design is described in Table 3.

TABLE 3 Intervention designs in the three sub-studies

Study		T	T	Intervention	T	T	Intervention	T	T
I	Condition 1	X	X	GGM	X				
	Condition 2	X	X	NR	X				
II	Condition 1	X		GGM	X	X		X	
	Condition 2	X			X	X	NR	X	
III	Condition 1	X	X	GGM-NC	X		GGM-BA	X	X
	Condition 2	X	X	GGM-NS	X		GGM-BA	X	X

Note. T = Time points for assessment tasks; X = assessment administered; GGM = GraphoGame Math; NR = Number Race; GGM-NC = GGM-Number Concept; GGM-NS = GGM-Number Sequence; GGM-BA = GGM-Basic Addition. Children were randomly assigned to two intervention conditions before the first intervention started. At the initial level, the two conditions did not differ in age, domain-general (sub-studies I, III), and domain-specific skills. In sub-study II, however, the wait-list control group showed significantly better verbal counting skills than the GGM-group at the first assessment point. Different versions of GGM were used in sub-studies I, II, and III.

In all three sub-studies, teachers were asked to not help the children solve the tasks. Instead, teachers were encouraged to focus on the classroom situation because GGM was treated as a cost-effective, relatively teacher-independent method for individualizing early number skills support.

3.4 Data analysis

Due to small intervention samples and the participants' low performance in tasks assessing early number skills (positively skewed distributions) nonparametric tests were used in the analyses in all of the three sub-studies (with SPSS 20–22). The Wilcoxon signed-rank test was used to analyze the within-group effects. To calculate the within-group effect sizes for the Wilcoxon signed-rank test results, the following formula was used: $ES\ (r) = Wilcoxon\ Z/\sqrt{N}$, where N is the number of observations (Field, 2013).

The Mann-Whitney U -test was used to analyze between-group differences before the intervention period, and to compare between-group differences in gain scores. To calculate the between-group effect sizes for the Mann-Whitney U -test results, the same formula as for the Wilcoxon signed-rank test results was used. The within-group results and the between-group results in gain scores were interpreted with exact, one-tailed p values. The initial level group comparisons were interpreted with exact, two-tailed p values.

In sub-study II, individuality in responsiveness to intervention was also investigated by analyzing performance during practice using game-log data. The performance during the practice was contrasted to intervention gain scores.

4 OVERVIEW OF SUB-STUDIES

4.1 Study I

The aim of sub-study I was to examine whether short, intensive practice with GGM or NR supports the early number skills development in children with poor skills compared to the age-level (performance below the 10th percentile). Based on the two pre-assessments, the GGM and NR groups did not differ in age, kindergarten group, domain-general (Corsi blocks, Nonwords, and RAN) and domain-specific skills (verbal counting, subitizing fluency, dot counting fluency, and basic arithmetic). After daily practice for three weeks, immediate within-group improvements and between-group differences in intervention gain scores were tested.

After an intensive intervention period, the participants in the GGM group improved their accuracy in verbal counting skills, which was measured by counting forward, backward, and skip counting tasks. This positive change in verbal counting skills can be interpreted to reflect better understanding of the concept of cardinality and relations between numbers. In GGM, cardinality, ordinality, and relations between numbers were supported by training object counting, number neighbors (i.e., number before or after a target number), and exact number comparison (see Appendix 1).

After the intervention, the children in the GGM group could more quickly discriminate sets of four to six dots while maintaining the same accuracy rate. The first version of GGM consisted of practice requiring mapping between different numerical representations (number words, quantities, and symbols; see Appendix 1). In most time-limited training trials, the children were required to count or recognize different amounts of quantities or written number symbols in order to select the correct alternative among three to five incorrect options. The children needed to quickly determine or count at least several or all of the amounts in each alternative to select the correct response. GGM might have supported the use of more efficient strategies to determine the exact amount of objects than counting all the dots, one by one (i.e., conceptual subitizing).

A positive intervention effect on basic arithmetic was found in the NR group. In sum, the two groups which received different types of practice improved different numerical skills after a short training period, even though the intervention intensity and curriculum setting were similar. This finding suggests that a specific program could produce specific effects in kindergarteners most at risk for MD.

4.2 Study II

In sub-study II, the first goal was to find out to what extent a short, intensive intervention with the second version of GGM program can support early number skills, particularly basic addition skills, in children with poor basic addition skills as compared to the age-level (performance below 1.5 SD; below the 7th percentile). Secondly, the stability of the potential effects was assessed over the 9-week follow-up period. The third purpose was to examine individual differences in responsiveness to GGM intervention using the game-log data collected during the practice.

In the pre-assessment, the two intervention groups did not differ in age, kindergarten group, basic addition skills (main outcome), enumeration skills, and Number Sets Test results (Geary et al., 2009). However, the wait-list control group performed better in verbal counting skills than the GGM group. After the first 3-week intervention period, within-group improvements and between-group differences in intervention gain scores were tested. This was followed by the third assessment. After that, the wait-list control group received Number Race training for 3 weeks, and finally, the fourth assessment was administered. The two intervention groups received different training, so between-group differences were not tested after the second intervention period. However, this design enabled examining the stability of the potential intervention improvements, which has been rarely done in earlier CAI studies (cf. Fuchs et al., 2006; Wilson et al., 2009).

After the first intensive 3-week intervention period, significant within-group improvement in accuracy was seen in basic addition, verbal counting, and composing skills. These positive changes were stable over the 9-week follow-up period. However, there was no significant difference in gain scores between the intervention group and the wait-list control group, and the between-group difference in raw scores of basic addition (main outcome variable) was only marginally significant immediately after the first intervention period ($U = 31.0$, $Z = -1.53$, $p = .067$, $r = .33$).

Raw and gain scores and game-log data showed that some intervention participants benefited from GGM training by improving their basic addition skills. However, some intervention participants did not seem to progress much during or after the practice. This seemed to be related to the adaptation algorithm of GGM which allowed children to proceed in the training contents only after satisfactorily passing each individual level (i.e., after a certain amount of correct responses).

4.3 Study III

Sub-study III was aimed at examining the effects of two theory-based intervention sequences on basic addition skills. The effects of the third version of GGM were tested within low-performing kindergarteners (identified by special kindergarten teachers) and the poorest-performing kindergarteners (identified by special kindergarten teachers and individual assessments). The latter group of participants performed poorly in at least two of the following skills: dot counting fluency, verbal counting, and basic addition skills (below the 10th percentile as contrasted to the age-level performance). In both examinations, the intervention groups did not initially differ in age, curriculum conditions, domain-general (PPVT, and RAN), and domain-specific skills (main outcome variables: dot counting fluency, verbal counting, and basic addition skills; background variables: enumeration, composing, and story problem solving skills).

The first condition was based on the access deficit hypothesis according to which children with MD have deficits in processing numerical meaning from number symbols and, consequently, difficulties in learning arithmetic (cf. De Smedt & Gilmore, 2011; Rousselle & Noël, 2007). Based on this hypothesis, interventions focusing on number symbol-quantity mapping skills might be beneficial for children having (or at risk for) MD. The second condition was based on findings that early verbal counting skills strongly predict calculation fluency (e.g., Koponen et al., 2013) and are required for strategy development and, moreover, flexible and sufficient strategy use during arithmetic problem solving (Baroody, Bajva, & Eiland, 2009; Dowker, 2009a; Wylie, Jordan, & Mulhern, 2012). Therefore, the first intervention group started with number concept training, followed by basic addition training (the NC-group), while the second intervention group started with number sequence training, followed by basic addition training (the NS-group).

After daily practice for three weeks, significant within-group intervention effects were seen in dot counting fluency and in verbal counting accuracy among the low-performing children in both intervention conditions. The NC-group also improved in basic addition skills, even though those skills were not yet directly supported. After the second intensive 3-week intervention period, both groups showed gains in basic addition.

Among the poorest-performing children, the results revealed condition-specific effects in both intervention groups. The NC-group improved in dot counting fluency, and the NS-group in verbal counting skills. Although the NS-group also improved in dot counting, the gain was larger in the NC-group. After the second intervention period when both groups received basic addition training, the children in the NC-group increased their accuracy in basic addition more than those in the NS-group. This improvement remained stable over the

5-week follow-up period when the difference in gain scores was statistically significant, favoring the NC-group.

The NC-group likely benefited more from massive training of exact non-symbolic and symbolic comparisons and from mapping between different numerical representations as compared to the NS-group which completed and continued number sequences forward and backward. Due to the requirement for exact magnitude and symbolic processing, the children in the NC-group might have paid more attention to exact relations between numbers than the NS-group. Understanding of these relationships might have helped the NC-group benefit from the composing, decomposing, and basic addition tasks practiced during the second intervention period (cf. Geary, 2013; Krajewski & Schneider, 2009).

In sum, the low-performing children seemed to benefit from GGM training regardless of the condition in which they participated. In contrast, among the poorest-performing children, the intervention gain was dependent on the training content, and number concept skills training seemed to offer better transfer to basic addition skills.

5 GENERAL DISCUSSION

Children performing poorly in early number skills seem to have persistent deficits in learning numerical information. These deficits are indicated by atypical number skills development, an increasing performance gap with low- and typically achieving age-peers, and, consequently, less benefit from mathematics instructions in primary school (Morgan et al., 2009, 2011; Murphy et al., 2007; Salaschek et al., 2014; Wong et al., 2015). To avoid cumulative deficits in learning and reduce the need for special education, the necessity for effective early support has also been acknowledged in Finnish Basic Education Act (642/2010, section 30). However, very few theory-based assessment and progress monitoring tools and intervention methods are available for this purpose, especially in the pre-primary education.

This thesis was aimed at examining the intervention effects of a theory-based computer-assisted method (GraphoGame Math, GGM) on early number skills in children most at risk for mathematics difficulties (MD). GGM is based on evidence concerning early number skills development, predictors and features of MD, the principles of effective computerized and non-computerized intervention programs, and user interface features. Given the general need for cost-effective, individualized interventions for children at high risk for learning difficulties (cf. Duncan & Magnuson, 2007; Holmes & Dowker, 2013), the other aim was to evaluate the value of the freely downloadable GGM as a potential evidence-based method for early number skills support. This could offer equal opportunities for cost-effective early intervention in Finnish pre-primary special education settings around the country.

Immediate content-specific intervention effects and their stability were studied using quasi-experimental designs, separate outcome measures, and delayed post-assessments. The aim was to increase knowledge of the benefits of GGM training for the poorest-performing kindergarteners (performance below the 10th percentile as compared to the age-level) because their responsiveness to early number skills intervention (computerized or non-computerized) is relatively understudied. The intervention participants were comparable in terms of their basic number skills, age, pre-primary settings, and socio-economic backgrounds across sub-studies I, II, and III. The majority of the intervention participants were not successful in counting correctly up to 20. Children failed to count forward and backward from a given number, and were incapable of composing non-symbolic and symbolic quantities and solving even the easiest basic addition tasks (such as 2 + 1).

The main findings of this thesis were that 1) the poorest-performing children seemed to benefit from intensified, targeted GGM intervention, and the intervention improvements were stable over the follow-up period, 2) the poorest-performing children displayed specific training effects indicating improvement in the skills practiced, and no transfer to other number skills was found, while gains by low-performing children were more general and transferred to number skills not directly practiced, 3) the requirement to process (i.e., map and compare) between different numerical representations seemed to be beneficial for basic addition skills among the poorest-performing children, and finally 4) despite group-level improvements, individual variations in responsiveness to GGM intervention were observed. In brief, these findings support the theories of early number skills development (Dehaene, 2011; Geary, 2013; Krajewski & Schneider, 2009; von Aster, 2000; von Aster & Shalev, 2007) and are comparable to evidence of the promising effects of CAI (Baroody et al., 2012, 2013; Kucian et al., 2011; Käser et al., 2013) and the meaningful ways to overall manifest improvements among children having (or at risk for) MD or DD (Baker et al., 2002; Fuchs, Fuchs, Powell et al., 2008; Gersten et al., 2009). This thesis also provided support for the suggestion that symbol-quantity processing training might be beneficial for children having (or at risk for) MD (De Smedt & Gilmore, 2011; Rousselle & Noël, 2007). As well, the current findings are in line with previous ones reporting individual variances in responsiveness to interventions among the target group of children (e.g., Fuchs et al., 2012).

Intervention improvements were seen in dot counting, verbal counting, composing, and basic addition skills and tended to depend on the performance in specific training contents (sub-studies I–III). In sub-study I, intervention improvements differed between the two intervention groups which practiced different types of numerical contents. In sub-study II, improvements were seen in the skills practiced and in basic addition but only if the training associated with previous developmental steps (concerning basic skills) was satisfactorily passed and followed with basic addition training (cf. the theoretical model of early number skills development; e.g., Geary, 2013; Krajewski & Schneider, 2009; see also Fig. 1 in section 1.1). Content-specific training effects without transfer to other number skills than those practiced by the children most at risk for MD were also seen in sub-study III. The NC-group improved their dot counting fluency (and not verbal counting skills) as was practiced, and the NS-group improved their verbal counting skills as was practiced. Although the NS-group also showed improvements in dot counting fluency, the gain was larger in the NC-group in this skill. Neither the NC-group nor the NS-group showed improvements in basic addition during the first intervention period. In contrast, low-performing children exhibited more generalized effects on early number skills than only in the skills practiced (sub-study III). For example, although the NC-group received number concept training and the NS-group number sequence training, both groups improved their dot counting fluency and verbal counting skills and achieved similar gains in the basic addition intervention after the second intervention period. The NC-group showed improvements in basic addition after the first intervention period even before starting specific basic addition training.

Two hypotheses of the factors underlying MD have been proposed. First, children with MD show deficits in approximate magnitude processing, which contribute to difficulties learning number symbols and basic arithmetic (the defective number module hypothesis; e.g., Butterworth, 2005; see also Piazza et al., 2010). The NR intervention is based on this first hypothesis. In sub-study I, the NR group improved in basic arithmetic skills after an intensive practice with approximate magnitude comparisons and thus, this hypothesis was somewhat supported. Second, despite intact processing of quantities, children with MD can have deficits in accessing quantities from number symbols, which affects to deficits in learning arithmetic (the access deficit hypothesis; De Smedt & Gilmore, 2011; Rousselle & Noël, 2007). In the present study, the children most at risk for MD showed gains from the GGM intervention which required symbolic number processing skills (NC-component; sub-study III). For example, children needed to match and continuously discriminate the numerical values of the response alternatives in most training trials (> 70 %) because both non-symbolic and symbolic numerical representations were presented as alternatives at the same time. These options could have further affected participants' composing, decomposing, and basic addition skills (following developmental steps) (Geary, 2013; Krajewski & Schneider, 2009; Fig. 1 in section 1.1). Likely for this reason, the NC-group achieved significantly larger gains in basic addition after the entire study process than the NS-group. The trials in NS-component did not as frequently require symbolic number processing (< 20 %). However, this finding requires experimental replications with two separate training contents comprising one that deals purely with quantities without symbolic processing and another with symbols and their corresponding quantities.

It has also been suggested that low-performing children would have deficits in symbolic numerical processing skills and children with DD (here, MD) would have deficits in both non-symbolic and symbolic numerical processing skills (Wong et al., 2015; see also Landerl et al., 2009). However, more knowledge is needed on deficits underlying MD due to variation in task types (cf. Schneider at al., in press; Wong et al., 2015) and sampling and sample characteristics issues (cf. De Smedt et al., 2013). For example, of the skills Wong et al. (2015) assessed, we failed to measure symbolic comparison skills. In sub-study I, the speeded comparison task (similar to that Wong et al., 2015 used) prompted guessing and a clear floor effect among the poorest-performing children. Worth to note, however, according to the game-log data of sub-study II, some of the participants used the 3-week intervention period in practicing only the very basic number skills. These children did proceed slowly through one-to-one correspondence and approximate number comparison, as well as mapping tasks which included the requirement of exact non-symbolic comparison skill. This finding is in line with those by Wong and his colleagues (2015).

Despite promising group-level improvements, individual differences in responsiveness to GGM interventions were found (sub-study II, III; cf. Fuchs et al., 2012). This is why the data analyses of the sub-studies could have focused more on individual-level changes. However, due to the lack of more detailed background information on participants, the single-case analyses were not carried out, with the exception of sub-study II. Nevertheless, the individuality in responses to GGM intervention is difficult to explain by means other than intervention-based factors. The Finnish pre-primary education system is nationwide and each day care center follows the same curriculum instructions (Finnish National Board of Education, 2010; the English version). So, individual improvements cannot likely be explained by variances in pre-primary settings or spontaneous development during such short intervention periods, as was the case in the current sub-studies. Based on previous findings, it seems, for example, that SES explains very little about the variances in learning outcomes in Finland (Aunio & Niemivirta, 2010; Hirvonen et al., 2012; Sahlberg, 2007). This is contrary to some other countries (cf. Jordan & Levine, 2009). Furthermore, the test-retest effect (a potential bias for intervention improvements) was controlled for in each of the three sub-studies at the group level. Since there were no differences between the pre-tests' scores among the poorest-performing children, the averages of those tests were used as the initial performance-levels in sub-studies I and III. If the test-retest effect had produced an improvement in basic addition in the GGM-group in sub-study II, there should have been a comparable effect observed in the children in the wait-list control group because the same assessment tasks at the same time points were repeated in both groups (see Table 2 in sub-study II).

Outside domain-specific factors associated with risk for MD, most of the poorest-performing children had poor proficiency in domain-general skills (visuo-spatial and phonological working memory, vocabulary and rapid naming speed) compared to normative age-level data. Due to small scale of the data and the data analyses methods used, domain-general measures were used only to assess the comparability of the two intervention groups and were not set as covariates for numerical skills or intervention gain scores (sub-studies I, III). However, weaknesses in these skills might have played an explanatory role in the variance in the responsiveness to GGM intervention, even though there was no correlation between the initial level of these skills and their performance during or after the intervention. Nevertheless, based on the findings of sub-study II, performance during GGM practice varied greatly among the poorest-performing children. There were no obvious problems in intervention fidelity, and the groups had comparable initial numerical skills and pre-primary education settings; therefore, GGM seemed to identify responders and non-responders to intervention. Given this individual variance in responsiveness to intervention (cf. Torgesen, 2000), the dual-discrepancy approach for identifying responders and non-responders to intervention could be relevant (poor performance-level and intervention growth rate compared to age-peers; Fuchs & Fuchs, 1998; McMaster et al., 2005). Based on the previous recommendations, the importance of continuous progress monitoring of skills development (Geary, 2011b), and the study results of sub-study II, a triple-discrepancy approach could also be suggested for identifying non-responders (suggested in sub-study II). With this approach, the identification of non-responders would be based on three criteria: 1) poor initial performance-level, 2) poor intervention, and 3) poor follow-up growth rate. Instead of waiting for a child to fail to learn age-relevant skills or to perform poorly in post-test assessments, dynamic testing (cf. Grigorenko, 2009) and ongoing progress monitoring (e.g., Fuchs et al., 2007) might both be beneficial, especially in the target group of the poorest-performing children, as also suggested in the RTI model (Fuchs & Fuchs, 2006; Fuchs et al., 2003). To respond to this question about being (non-)responder, performance during CAI interventions (here, GGM) could be assessed effectively by analyzing the game-log data. Such data have rarely been investigated in treatment studies (however, see Obersteiner et al., 2013).

Of the malleability factors, the roles of instruction and early intervention were taken into account when developing and examining the effects of GGM method, although explicitly investigated in this thesis only in sub-study II. The results of this method could have been positively affected by the use of explicit instructions, pedagogical agents, and tutorials to strengthen the link between sub-skills, and by the step-by-step progression from concrete visualization and manipulation towards semi-concrete and more abstract levels (Dowker, 2001; Fuchs, Fuchs, Powell et al., 2008; Griffin, 2004, 2007; Sarama & Clements, 2009; van Luit & Schopman 2000). In addition, an attractive environment, massive repetition of basic concepts with continuous practice, and the opportunity for early success could have helped task-orientation and thus, played roles in the current findings (cf. Baker et al., 2002; Fuchs, Fuchs, Powell et al., 2008; Gersten et al., 2009). The intensity and individualization of the intervention enhanced by GGM's content-based adaptation and effective feedback system seemed to be beneficial for the children most at risk for MD (here in the CAI context: Cheung & Slavin, 2013; Hasselbring, 1986, Kulik & Kulik, 1991; Kroesbergen & van Luit, 2003; Li & Ma, 2010; Seo & Bryant, 2009; Seo & Woo, 2010; Slavin & Lake, 2008). In sub-study II, some children seemed to benefit from such training more than children in the wait-list control group. This was indicated by raw scores and the fact that after an intensive GGM practice, the difference in basic addition skills was less than 0.5 SD from the age-level mean in 5 of 13 intervention participants.

To conclude, the poorest-performing children need supplemental, intensive support due to their high risk for MD. Specific, adaptive interventions for that purpose should include individualized tutoring and ongoing analysis of performance (e.g., Fuchs, Fuchs, Graddock et al., 2008). Therefore, a theory-based computerized intervention could be a meaningful method to intensify and individualize early number skills support and progress monitoring for the children most at risk for MD. As known, neurobiological deficits underlying

MD can vary among individuals (e.g., Dehaene, 2011; Geary, 2004; Rubinsten & Henik, 2009; von Aster & Shalev, 2007). Environmental, domain-general and domain-specific cognitive skills, malleability, and other general factors can also explain heterogeneity in early number skills performance (Butterworth et al., 2011; Dehaene, 2011; Geary, 2004; Jordan et al., 2003; Kaufmann et al., 2013; Kleemans et al., 2011; Landerl & Moll, 2010; Moll, Snowling, Göbel, & Hulme, 2015; Pennington, 2006; see also Fig. 3 in section 1.2). Since this thesis focused only on domain-general, domain-specific, and malleability factors as indicators of risk for MD (see Fig. 5, in section 2), limitations (see section 5.3), and suggestions for future studies are discussed (see section 5.4).

5.1 Implications for intervention research

The findings of the current thesis were encouraging and in line with previous CAI studies (e.g., Baroody et al., 2012, 2013; Fuchs et al., 2006; Kucian et al., 2011; Käser et al., 2013; Praet & Desoete, 2014; Wilson et al., 2009). As implied in reviews, assessment of the effectiveness of established CAI methods has been difficult due to varying target group characteristics, group sizes, numerical content practiced, instructional components, assessment methods, and the intensity and duration of interventions (Cheung & Slavin, 2013; Kroesbergen & van Luit, 2003; Räsänen et al., 2009; Seo & Bryant, 2009; Slavin & Lake, 2008).

Recently, more detailed descriptions of intervention methods and study procedures have been required for reporting the results. These more detailed reports are thought to deepen knowledge of appropriate support, particularly knowledge of 1) the core intervention components behind the potential positive effects, 2) how the confounding factors are controlled, 3) to whom the potential effects are contrasted, and 4) how the (non-)significant results are interpreted (cf. Dyson, Jordan, Beliakoff, & Hasinger-Das, 2015; Grissom & Kim, 2012; Hill et al., 2008; Kucian et al., 2011; Käser et al., 2013; Praet & Desoete, 2014). Even though the reports related to intervention research are nowadays more detailed, however, to detect the intervention-based factors behind the effects, concise designs are needed. If teacher-directed intervention and CAI are mixed (see Baroody et al., 2012, 2013), the intervention duration is relatively long (Baroody, Purpura, Eiland, & Reid, 2015), or the interventions are carried at homes (Kucian et al., 2011; Käser et al., 2013), the effective intervention components are more difficult to determine due to potential confounding factors (e.g., curriculum-based instruction, maturation, and intervention fidelity).

More precisely, the poorest-performing children do not seem to benefit from general instruction as their age-peers (cf. Morgan et al., 2009, 2011; Murphy et al., 2007), so the effects of instruction or specific interventions are not comparable to groups of low- or typically-performing children (see also Geary, 2011b). It has been recommended to compare these effects to those on a similarly performing control group, instead of better or worse performing ones (Grissom & Kim, 2012; Hill et al., 2008). However, it would be ethically questionable to create an intervention design with an experimental group of the poorest-performing children and to contrast the effects on a control group of children (without acute support) already identified as high risk for MD. In this thesis, the effectiveness of GGM was compared to 1) a performance-level control group which simultaneously underwent NR intervention (sub-study I), and 2) a no-treatment performance-level control group which underwent NR intervention after a short waiting time (wait-list design; sub-study II). As well, 3) the two different GGM components (NC and NS), and their transfer to basic addition skills were compared to each other (sub-study III).

The influence of ethical considerations on the study designs placed some restrictions on examining the effectiveness of GGM. For example, a wait-list design was used in sub-study II because we were interested in the potential intervention effects of GGM and their stability. For ethical purposes, the control group was offered NR training between the third and fourth time points. This opportunity resulted in difficulties comparing the performance of the two intervention groups over the study period due to much longer baseline for the control group and the other type of numerical training applied in the control condition. To compare the effectiveness of GGM against this particular (or any other) method, the third condition with simultaneously administered training should have been designed for contrasting methods (cf. sub-study III). Most importantly, a business-as-usual control design involving the poorest-performing children would not be an ethical intervention design due to the high risk for increasing the cap with low- and typically-achieving children (Geary et al., 2007; Morgan et al., 2009, 2011; Murphy et al., 2007).

The roles of assessment tools in evaluating skill-levels and responsiveness to intervention also deserve discussing. Although using standardized assessment methods is recommended for this purpose (cf. Cook & Cook, 2013; Fuchs, 2003), estimating the initial performance-level, intervention effects, and their effect sizes is challenging when assessing the early number skills of the poorest-performing children. The distributions on standardized tests tend to be positively skewed (the floor effect). For example, in sub-study I, the speeded comparison task and the basic arithmetic test caused guessing and a clear floor effect. Consequently, specific experimental assessment methods were developed, piloted, and used in sub-studies II and III. However, doing so created difficulty evaluating validity and reliability with the small, heterogeneous sub-samples of children in the present studies (cf. Fuchs et al., 2012; Kaufmann et al., 2013; McMaster et al., 2005; Rubinsten & Henik, 2009). These limitations could easily bias the effect size estimations (cf. Cheung & Slavin, 2013; Slavin & Lake, 2008) if overall positive intervention effects can be seen within the poorest-performers (see Torgesen, 2000). However, an evidence-based intervention method requires replication of evidence of its effectiveness through adequate experimental intervention designs with relevant effect sizes (Cook & Cook, 2013, Cook et al., 2009; Gersten, Fuchs et al., 2005). This requirement is challenging to achieve because replication of interventions always requires much effort and resources, especially when conducted with larger samples.

In sum, many methodological challenges persist in intervention research involving small groups of atypical learners. A study design with a stable baseline would help to control for the potential bias in intervention improvements among such participant characteristics (cf. Horner et al., 2005). As mentioned, there were no differences between the two pre-tests' scores among the poorest-performing children in sub-studies I and III. Due to a lack of knowledge on test-retest effects among different performance-level groups, the baseline data of sub-study I was checked for this purpose. The data showed that low-performing children improved their dot counting fluency (p = .061) and basic arithmetic skills (p = .022) more than the poorest-performing children when the pre-test measures were repeated in a week. This is why there is a need to control for test-retest effects and stability of the initial skill-level among atypical learners. Furthermore, if specific assessment tools are used, the initial performance level should be compared to age-relevant reference data in order to define the level of difficulty. As mentioned, both the initial performance level and the responsiveness to intervention in the target group of children seem to be heterogeneous (e.g., Geary, 2011b; Rubinsten & Henik, 2009; Kaufmann et al., 2013; Jordan et al., 2003; Landerl & Moll, 2010). This heterogeneity decreases the likelihood of having two or more comparable intervention groups in a study. Furthermore, to interpret the effects of intervention, special efforts are needed to obtain objective data of participants' 1) performance levels before the study, 2) performance within the conditions, 3) progress during the practice, and 4) fidelity to intervention (cf. Cook et al., 2009; Gersten, Fuchs et al., 2005; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2013).

There are also clear reasons to specify the methods targeting for clinical environment, special, and general educational settings (cf. Fuchs, 2005). Such specification could offer a framework to better evaluate and interpret the effectiveness of the intervention with different sub-samples and varying participant characteristics (see Li & Ma, 2010). This analysis could also help identify the factors vital for supporting early number skills development among children in different skill-level groups (i.e., general support, intensified support, and special support).

5.2 Implications for practitioners and parents

GGM, as an intervention method, could offer an individually adaptive and motivating learning environment, along with high training intensity, continuous feedback, and repetition of basic number skills and domain-specific concepts – all found to be major factors in successful interventions for children having (or at risk for) MD (Fuchs, Fuchs, Powell et al., 2008; Fuchs et al., 2012; Gersten et al., 2009). The increased use of computers, laptops, tablets, and smart phones in Finnish pre-primary education and homes could allow using GGM and other CAI methods to provide targeted support in math during pre-primary education and the first and second grades. GGM can deliver cost-effective assistance for children regardless initial performance level (poor or low performance), teaching resources, and educational setting.

GGM could also be used to identify the individual needs of responders and non-responders to intervention. To be an easy tool for such game-log data analyses for teachers, GGM needs further development. At the moment, teachers can use paper-and-pencil screening tools (administered in groups or individually) and progress monitoring tools (administered individually) developed under the LukiMat-project (learning assessment tools, www.lukimat.fi) (Polet & Koponen, 2011; Salminen & Koponen, 2011).

Due to the large variance in early number skills performance (cf. Aunio & Niemivirta, 2010; Aunola et al., 2004) and skills development (e.g., Geary et al., 2007; Morgan et al., 2009, 2011; Murphy et al., 2007), identification of risk for MD is already highly important in pre-primary education. The deficits influencing early number skills development vary among individuals (e.g., Kaufmann et al., 2013; Rubinsten & Henik, 2009; see also Fig. 3 in section 1.2), so it is reasonable to assess the stability and change in performance level and responsiveness to intensified interventions (Geary, 2011b; Fuchs et al., 2012; Fuchs, Fuchs, Graddock et al., 2008). Assessing individuals' strengths and weaknesses and accordingly planning appropriate personalized support is important. It has been recommended that assessment should be based on the findings from teacher observations, screening and progress monitoring tools, curriculum-based exams, questionnaires, interviews, discussions, and on other potential assessment methods (cf. Dowker, 2009b). With relevant knowledge, adaptive and tailored training should be conducted sufficiently early to prevent and decrease difficulties in learning mathematics. Importantly, there is no reason to wait for failing or post-assessment scores if improvements are not seen during the ongoing intervention (cf. Fuchs et al., 2007, ongoing progress monitoring; Grigorenko, 2009, dynamic testing).

Identification of target skills for early intervention is challenging (Mazzocco, 2005; Mazzocco & Myers, 2003). Therefore, conducting pedagogical discussions about the individuality of needs in early number skills development is highly recommended in groups with different professional contributions and expertise. Such discussions could also relate to reasonable ways and time points at which to assess and support the children most at risk for MD and of how to foster, utilize, and individualize the applications available for daily use in pre-primary and primary education and homes.

5.3 Limitations and strengths

Children with the most severe mathematics learning difficulties seem more likely to have neurobiological deficits (e.g., Dehaene, 2011; Geary, 2004; Rubinsten & Henik, 2009; von Aster & Shalev, 2007) and comorbidity problems (e.g., Jordan et al., 2003; Landerl & Moll, 2010; Moll et al., 2014; Pennington, 2006) than their low- and typically-performing age-peers. However, this thesis did not investigate individual differences in these factors or other behavioral factors behind the risk for MD (see also Fig. 3, in section 1.2). Although the two intervention conditions did not differ in domain-general and domain-specific factors (cf. Kaufmann et al., 2013; von Aster, 2000), age, and pre-primary setting, the small sample sizes in sub-studies prevented examination of the influence of these factors on the intervention effects of GGM. It is important to control for working memory and language skills (or their correlates) when examining the initial performance-level and intervention effects among the children having (or at risk for) MD (cf. Geary, 2013; Kaufmann et al., 2013; Landerl, Göbel, & Moll, 2013; Raghubar, Barnes, & Hecht, 2010), as was done in sub-studies I and III. However, those certain domain-general measures were primarily selected for increasing the comparability of the two intervention conditions and could not be used as covariates for adding knowledge to factors behind deficits in skill-level or responsiveness to GGM intervention. Furthermore, neither socio-economic background nor gender effects were tested. However, there is no clear evidence of these factors' influence on predictions to math achievement (Duncan et al., 2007; Jordan et al., 2007, Jordan et al., 2006; in Finnish context: Aunio & Niemivirta, 2010; Hirvonen et al, 2012). Finally, children having (or at risk for) MD have been found to have problems in their strategy development (e.g., Wylie et al., 2012), but this thesis did not address this topic due to the focus on basic early number skills support and the difficulty embedding strategy training in a computerized format.

Although the intervention and analyses designs used in this thesis were planned and designed taking into account the research questions and target participants (e.g., Cook & Cook, 2013), the small sample sizes complicate interpretation of the intervention results (e.g., Cheung & Slavin, 2013; Li & Ma, 2010). Small sample size reduces the statistical power revealing the less robust intervention effects and tends to produce larger effect size estimates than in large scale studies (e.g., Cheung & Slavin, 2013; Slavin & Lake, 2008). A small sample size also influences the probability of having two (or more) comparable intervention groups if such heterogeneous in factors underlying MD are expected. The intervention conditions of the poorest-performing children could easily differ in individual differences in domain-general and domain-specific skills, as well as responsiveness to intervention and kindergarten instruction. Intervention results could be biased by other effects, such as test–retest effect, data collector bias (cf. sub-study II), and Hawthorne effect (i.e., belonging to a specific condition could produce its own effect on outcomes) (e.g., Adair, 1984). The test-retest bias could be controlled for with a baseline conducted until the substantive trend in skills measured is no longer observable (Horner et al., 2005). However, with similar procedures carried out in each sub-study (teachers nominated candidates who were then individually assessed for MD risk), the proportion of participants included in the main analyses ranged from 5 % to 7 %. This sampling aligned with estimates of MD prevalence (3–7, 5–7, to 5–8 %) (Butterworth et al., 2011; Geary, 2011b, 2004; Landerl & Moll, 2010) adding to knowledge of the effects of GGM intervention for poor early number skills among children most at risk for MD.

More generally, the results could have been affected and the interpretations complicated by the short training period, (lack of) sensitive assessment tools, relatively short follow-up periods, and the attention to the study process and ongoing curriculum activities paid by participating children, kindergarten teachers, and parents. As well, GGM might not have been a sufficiently flexible intervention method to guarantee an equally effective learning environment for all target participants. Large variances in raw and gain scores after intervention periods could be partly explained by the limitations in GGM's contents, layout, adaptation and feedback system etc., even though the game was developed based on prior research.

Despite the limitations, a particular attention to several methodological and ethical decisions has been given during the study process. These choices have been contrasted to the quality indicators for increasing reliability and validity in experimental and quasi-experimental group-level intervention research in special education (Gersten, Fuchs et al., 2005; see also other source e.g., What Works Clearinghouse), as was the case in this thesis. The selection criteria were confirmed by comparing the target group participants' skill-level to that found in various normative age-reference data (see sub-studies I, II, III). The domain-general measures (I, III), children's age (I, II, III), intervention intensity and short duration (I, II, III), and explicit, preset GGM instruction (I, II, III) increased the likelihood that the sub-samples randomized into two intervention conditions were comparable. Kindergarten teachers were instructed to complete training diaries (I, II, III), game-log data were checked and analyzed (II, III), and personal meetings (II), and phone calls (III) were used to assess intervention fidelity. For ethical reasons, the comparison condition was offered another type of intervention during or immediately after studying the effectiveness of GGM (I, II, III).

Paper-and-pencil measures were used to test the transfer effect from the computerized to the typical daily practice context (I, II, III). Separate domain-general measures (instead of using one sum score of numerical competency) were applied to capture and better understand the extent to which the poorest-performing children benefit from GGM practice (I, II, III), and how the potential improvements might remain over a short time period (II: 9-week follow-up; III: 5-week follow-up). Intervention gain scores were used to assess specific intervention effects instead of evaluating the between-group differences in an individual cross-sectional post-assessment point (I, II, III). The effect sizes were calculated for the immediate intervention effect (I, II, III), not for the potential delayed improvements. The latter data analysis could have produced misleading estimations for the specificity of the effect of GGM (i.e., delayed effects might be affected by issues external to the intervention). Furthermore, the test-retest reliability was calculated (I, III) or based on a pilot study conducted before the main study (II). The data collectors did not know to which of the two conditions participants would be randomly assigned after the pre-assessment(s) (I, II, III).

The data collectors were also unfamiliar with the participants (I, III) and expected study outcomes (I, II, III). To quantify the validity and reliability of the specific assessment tools developed for the study purposes, their test-retest effects were piloted (II), or controlled for in a study (I, III).

5.4 Future directions

Knowledge of the theory of early number skills development is becoming more precise. The associations between domain-specific sub-skills (e.g., Bartelet, Vaessen et al., 2014; De Smedt et al., 2013; Dowker, 2009b; Geary, 2013; von Aster & Shalev, 2007), and between domain-general and domain-specific skills (e.g., Bartelet, Ansari et al., 2014; Geary, 2011a; Kleemans et al., 2011; Mazzocco & Grimm, 2013; Zhang et al., 2014) have increased understanding of MD. Awareness of the individuality in responsiveness to intervention and its association with the former skills also adds to knowledge of the phenomenon (e.g., Frijters et al., 2011; Stage, Abbott, Jenkins, & Berninger, 2003; Supekar et al., 2013). However, sampling issues have not yet been widely considered (cf. Geary, 2011b; see also Fig. 2, in section 1.2). Knowledge of MD could be fruitfully increased and deepened by linking identification and intervention research. Doing so would require a longitudinal research design with multidimensional assessment procedures, intervention cycles, and continuous, online progress monitoring. Studies modeling the intervention and follow-up performance between the skill-level groups and within the individuals having (or at risk for) MD (e.g., piecewise methods) would increase information of factors underlying responsiveness to intervention. Single-case experiments, however, would deepen the knowledge on the individuality of MD.

As noted, a business-as-usual control design would not be an ethical intervention design when the poorest-performing children are involved in a study. However, the lack of such a design places restrictions on interpreting the intervention's effectiveness. One way to solve this contradiction could be the switching replications design with a short waiting time. In further studies, it could also be reasonable to examine the effects of an intensified CAI by varying only one element between experiment conditions. For example, the effects of a certain component can be controlled for only if, of two identical experimental conditions, one includes the component, and the other one does not. After acquiring detailed knowledge of the effects of single intervention components, the effects of combined training for domain-general and domain-specific skills could be tested. CAI could also be developed to address deficits in non-numerical, numerical, and other domain-specific cognitive skills (i.e., reading skills) because evidence of the underlying factors of comorbid problems is also rapidly increasing (e.g., Jordan et al., 2003; Pennington, 2006; Moll et al., 2014; Moll et al., 2015; Rubinsten & Henik, 2009; Willcutt et al., 2010; Willcutt et al., 2013). However, there is no clear evidence of interventional outcomes from such combined training (cf. Kaufmann et al., 2013). Carrying out such combined interventions could be interesting, especially with children most at risk for MD who might benefit differently from this kind of support than their age-peers. Combined interventions, however, raise the need for specific assessment and CAI methods which could be develop by an interdisciplinary research group of special education and psychology scholars, game developers, and game-log data statisticians.

Based on this study process, more attention should be paid to individual and content-based adaptation, pedagogical aspects, and feedback systems when developing and studying the effects of any theory-based CAI methods targeted for children (most) at risk (or having) specific learning difficulties.

YHTEENVETO (FINNISH SUMMARY)

Varhaisen tuen tarpeen tunnistamisen ja vaikuttavan tuen tarjoamisen tärkeys on nostettu esiin sekä perusopetuslaissa (PoL § 30; 642/2010) että opetussuunnitelman perusteissa (www.oph.fi), jotka molemmat säätelevät ja ohjaavat esiopetuksen järjestämistä. Tässä väitöskirjassa esitelty Ekapeli-Matikan kehitystyö ja siihen liittyvät vaikuttavuustutkimukset on tehty Niilo Mäki Instituutin koordinoiman LukiMat -hankkeen yhteydessä. Hankkeen pääasiallisena tavoitteena on ollut koota ja tarjota tietoa sekä kehittää tutkimusperustaisia välineitä niin varhaisten taitojen arviointiin kuin tukemiseen, jotta moninaistuvia ja päällekkäistyviä oppimisen vaikeuksia voitaisiin ennaltaehkäistä ja lieventää. Materiaalit ovat ilmaiseksi tarjolla kaikkien esi- ja perusopetuksen ammattilaisten, vanhempien ja muiden lapsen oppimista tukevien ammattiryhmien käyttöön (lisätietoa www.lukimat.fi). Hankkeen rahoittajana on ollut Opetus- ja kulttuuriministeriö vuosina 2007–2015. Ekapeli-Matikan alkuperäinen versio on kehitetty Jyväskylän yliopiston koordinoiman GraphoGame -hankkeen alkuaikoina (Richardson & Lyytinen, 2014).

Tämän väitöskirjan tavoitteena oli tutkia tietokoneavusteisen interventiomenetelmän (Ekapeli-Matikka) vaikuttavuutta varhaisiin numeerisiin taitoihin. Tavoitteena oli kehittää alkuperäisen peliversion pohjalta tutkimustietoon perustuva ja oppimiseen myönteisesti vaikuttava lisäharjoitusväline, jota voisi hyödyntää tehostetun tai erityisen tuen tarpeessa olevien lasten opetuksessa. Kaikissa kolmessa osatutkimuksessa tutkittiin eri peliversion vaikuttavuutta erityisesti niiden esiopetusikäisten lasten taitoihin, joiden lähtötason taidot olivat kaikkein heikoimmat verrattuna ikätasoon (heikoin kymmenes).

Osatutkimukset toteutettiin Itä-Suomessa, Etelä-Suomessa ja Keski-Suomessa vuosina 2007, 2008–2009 ja 2011. Jokaista osatutkimusta varten esiopetusryhmien opettajat nimesivät lapsia, joilla oli erityisen tuen tarvetta varhaisissa matemaattisissa taidoissaan. Kaikki tuen tarpeessa olevat lapset arvioitiin yksilöllisesti ja he saivat osallistua intensiiviseen tietokoneavusteiseen harjoitteluun, mikäli he saivat huoltajiltaan osallistumisluvan. Interventioon osallistuneiden lasten määrä vaihteli osatutkimuksittain riippuen vapaaehtoisten esiopettajien ja heidän opetusryhmiensä lasten kokonaismäärästä. Ensimmäisessä osatutkimuksessa oli mukana 17 lasta (236 lapsen otoksesta, 12 esiopetusryhmästä) ja toisessa osatutkimuksessa 21 lasta (n. 350 lapsen otoksesta, 24 esiopetusryhmästä). Kolmannessa osatutkimuksessa aineisto analysoitiin kahdessa osassa: ensin kaikkien interventiotutkimukseen valittujen 33 lapsen osalta sekä sen jälkeen taidoiltaan kaikkein heikoimpien 14 lapsen osalta (278 lapsen otoksesta, 14 esiopetusryhmästä). Jokaisessa osatutkimuksessa tutkimukseen osallistuneet arvottiin kahteen interventioryhmään. Harjoitusvaikutuksia arvioitiin suhteessa oman interventioryhmän aloitustasoon, perustuen yhteen tai kahteen alkuarviointiin, sekä suhteessa toiseen interventioryhmään. Aineistot analysoitiin parametrittomin tilastollisin menetelmin, koska erityisen tuen tarpeessa olevien lasten määrä oli pieni ja heidän taitotasonsa hyvin heikko, mikä näyttäytyi arviointituloksissa vinoina jakaumina.

Ensimmäisen osatutkimuksen tavoitteena oli selvittää, hyötyvätkö lukujonotaidoiltaan kaikkein heikoimmat esiopetusikäiset lapset ylipäätään lyhyestä ja intensiivisestä tietokoneavusteisesta harjoittelusta. Kaksi interventioryhmää sai harjoitusta eri menetelmillä. Lapset, jotka pelasivat Ekapeli-Matikan ensimmäistä versiota, paransivat lukujonotaitojaan sekä lukumäärän laskemisen taitojaan kolmen viikon päivittäisen harjoittelun jälkeen. Numerorata-peliä (Wilson, Dehaene ym., 2006) pelanneet lapset puolestaan paransivat perusaritmetiikan taitojaan vastaavan, samanaikaisesti toteutetun harjoitusjakson jälkeen. Koska harjoitusmenetelmät erosivat toisistaan, voitiin päätellä, että harjoitellulla sisällöllä on ollut merkitystä sille, missä taidoissa parannus näkyi. Ekapeli-Matikan harjoittelu painottui tarkkaan vertailutaitoon ja lukukäsitetaitoon, kun taas Numeroradan tausta-ajatuksena on tukea likimääräistä vertailutaitoa. Tutkimuksen havainnot nostivat esiin tarpeen kehittää tarkempia menetelmiä kaikkein suurimmassa oppimisvaikeusriskissä olevien lasten varhaisten taitopuutteiden tunnistamiseen. Lisäksi havaittiin, että Ekapeli-Matikan tulisi vahvistaa paremmin varhaisia perusmatemaattisia käsitteitä ennen kuin taidoiltaan kaikkein heikoimmat lapset voivat operoida lukumäärillä tai luvuilla.

Toisen osatutkimuksen tavoitteena oli selvittää sisällöltään ja esitystavoiltaan uudistetun Ekapeli-Matikan vaikuttavuutta yhteenlaskutaitoon niillä lapsilla, joiden kyseiset taidot olivat kaikkein heikoimmat verrattuna ikätasoon. Tutkimusasetelmaan sisällytettiin kaksi interventiojaksoa, jotta Ekapeli-Matikan harjoitusvaikutuksia voitiin seurata muutaman kuukauden ajan. Alkumittauksen jälkeen toinen interventioryhmä harjoitteli Ekapeli-Matikkaa, kun taas toinen ryhmä ei ensimmäisellä jaksolla saanut lisäharjoitusta. Toisen harjoitusjakson aikana jälkimmäinen ryhmä sai Numerorata-harjoitusta, kun taas Ekapeli-Matikkaa ensimmäisellä jaksolla pelanneet eivät saaneet lisäharjoitusta. Tutkimuksen tulokset osoittivat, että ne lapset, jotka olivat edistyneet sujuvasti kolmen viikon päivittäisen harjoittelun aikana Ekapeli-Matikan pelaamisessa, paransivat yhteenlaskutaitojaan eniten ja heidän harjoitusvasteet säilyivät yhdeksän viikon ajan. Toisaalta ne lapset, jotka eivät edistyneet Ekapeli-Matikan pelaamisessa eli juuttuivat harjoittelemaan varhaisia perustaitoja, eivät edistyneet yhteenlaskutaidoissaan välittömästi intervention jälkeen eivätkä seurantajakson aikana. Osa näistä hitaasti edistyneistä lapsista kuitenkin paransi osaamistaan sellaisissa perustavissa numerotaidoissa, joita Ekapeli-Matikan ensimmäiset taitoalueharjoitteet sisälsivät. Havaintojen perusteella voitiin päätellä, että pelin adaptaatio ja palautejärjestelmä tarvitsivat kehitystyötä, jotta ne voisivat paremmin tukea erityisesti lapsia, jotka ovat lähtötason taidoiltaan kaikkein heikoimpia. Tutkimuksen tulokset vahvistivat aikaisempia havaintoja siitä, että kyseisen kohderyhmän lasten harjoitusvasteet ovat usein hyvin yksilöllisiä: vaikka lapset olisivat lähtötason taidoiltaan samankaltaisia, he eivät välttämättä hyödy tietynlaisesta harjoittelusta samalla tavalla. Tähän vaikuttanee oppimisvaikeuksien moninaisuus ja mahdollinen päällekkäisyys sekä yksilölliset eroavaisuudet vaikeuksien syytaustoissa.

Ekapeli-Matikan sisältöjä, esitystapoja, adaptaatiota ja palautejärjestelmää muokattiin ennen kolmatta interventiotutkimusta. Lisäksi peliin sisällytettiin enemmän pedagogisia animaatioita ja ohjeistuksia, jotta kulloinkin harjoiteltavan osataidon käsitteistö tulisi tutuksi jo ennen varsinaista taidon harjoittelua. Perustavat numeeriset taidot jaettiin kolmeen pelisisältöalueeseen: lukukäsitetaitoihin, lukujonotaitoihin ja yhteenlaskutaitoihin. Jokainen näistä alueista sisälsi alkuarviointiosuuden, jonka perusteella pelin aloitustasoa yksilöllistettiin. Tutkimuksessa toinen interventioryhmistä harjoitteli ensimmäisen interventiojakson aikana lukukäsitetaitoja ja toisen jakson aikana yhteenlaskutaitoja. Toinen ryhmä harjoitteli ensimmäisen jakson aikana lukujonotaitoja ja toisen jakson aikana yhteenlaskutaitoja. Tarkasteltaessa ensin kaikkia interventioon osallistuneita lapsia tutkimuksen tulokset osoittivat, että lapset paransivat ensimmäisen kolmen viikon päivittäisen harjoittelujakson jälkeen samankaltaisesti riippumatta harjoitellusta sisällöstä. Tarkasteltaessa taidoiltaan kaikkein heikoimpia lapsia tulokset osoittivat, että harjoitellulla sisällöllä oli merkitystä siihen, missä taidoissa lapset paransivat osaamistaan. Ne lapset, jotka harjoittelivat lukukäsitetaitoja, paransivat nopeassa lukumäärien määrittämisessä. Ne lapset, jotka harjoittelivat lukujonotaitoja, paransivat lukujonotaitojaan. Toisen interventiojakson jälkeen havaittiin, että ne taidoiltaan kaikkein heikoimmat lapset, jotka olivat harjoitelleet ensin lukukäsitetaitoja, hyötyivät enemmän yhteenlaskutaidon harjoittelusta kuin ne lapset, jotka harjoittelivat ensin lukujonotaitoja. Tulos tukee aikaisempia havaintoja siitä, että ne lapset, joilla on vaikeuksia numeeristen taitojen oppimisessa, saattaisivat hyötyä lukukäsitetaitoa vahvistavasta harjoittelusta, koska he näyttäisivät olevan erityisen heikkoja ymmärtämään numerosymboleiden lukumääräisyyttä (esim. De Smedt & Gilmore, 2011).

Väitöskirjan tulosten perusteella Ekapeli-Matikka osoittautui tehokkaaksi keinoksi tukea heikkoja varhaisia numeerisia taitoja esiopetusikäisillä lapsilla. Pelin käyttöä voi siten suositella esiopetukseen ja mahdollisesti myös alkuopetukseen eriyttävänä, tehostetun ja erityisen tuen menetelmänä. Tulokset vahvistavat myös ymmärrystä siitä, millaisia haasteita numeerisilta taidoiltaan kaikkein heikoimpien lasten taitojen arvioinnissa ja tukemisessa voidaan kohdata. Lisätietoa tarvitaan luotettavista tuen tarpeen tunnistamisen ja oppimisen seurannan välineistä, tuen toteutusmahdollisuuksista ja tuen vaikuttavuuden arvioinnista. Tämän väitöskirjan sisältämät osatutkimukset antavat viitteitä siitä, että harjoitellulla sisällöllä on merkitystä, ja että harjoittelun sisällöllinen kohdentaminen on ensisijaisen tärkeää taidoiltaan kaikkein heikoimpien lasten taitojen tukemisessa. Lisäksi näyttäisi siltä, että yksilölliseen adaptaatioon ja palautejärjestelmään tulisi kiinnittää erityistä huomiota, kun tietokoneavusteisia menetelmiä kehitetään tehostetun tuen ja erityisen tuen tarpeisiin.

Yleisenä johtopäätöksenä voidaan todeta, että numeerisilta taidoiltaan kaikkein heikoimpien lasten taitoja tulisi arvioida ja tukea systemaattisesti koko esiopetusvuoden ajan. Monipuolisen arvioinnin tavoitteena on suunnitella kohdennettua ja tarkoituksenmukaista varhaista tukea sitä tarvitseville. Mahdollisen lisäharjoittelun tulisi olla systemaattista, intensiivistä ja kohdennettua ensin niihin taitohierarkian varhaisimpiin osataitoihin, joissa lapsella ilmenee puutteita. Keskeisiä varhaisia taitoja ovat lukumääräisyyden taju, lukumäärän määrittämisen taidot, lukujonotaidot sekä lukusanan, lukumäärän ja numerosymbolien vastaavuuden ymmärtämisen taidot. Tämän jälkeen osa-kokonaisuuksien yhdisteleminen ja kokonaisuuksien osiin jakaminen ovat keskiössä ennen varsinaista ei-symbolista ja symbolista aritmetiikkaa. On tärkeää varmistaa, että perustaidot hallitaan sekä käsitteellisellä että konkreettisella rutiinien tasolla. Vain näin lapselle tulee mahdolliseksi manipuloida lukumääriä, ymmärtää lukujen välisiä suhteita ja suoriutua haastavammistakin matemaattisista operaatioista.

Varhaisen tuen tarpeen tunnistaminen, tuen suunnittelu, toteuttaminen ja sen mahdollinen muokkaaminen näyttäisivät vaativan erityistä pedagogista herkkyyttä, koska lasten taitopuutteissa on yksilöiden välisiä eroja ja oppimisen vaikeudet ovat moninaisia, päällekkäistyviä sekä luonteeltaan hyvin sitkeitä. Tehostetun tai erityisen tuen tarvetta ja vaikuttavuutta tulisikin arvioida mahdollisimman monipuolisin menetelmin, jotta yksilölliset vahvuudet ja heikkoudet sekä oppimiskapasiteetti ja siihen vaikuttavat tekijät voitaisiin tunnistaa. Näin voitaisiin tehdä luotettavammin päätelmiä siitä, onko valittu tuen menetelmä sopiva ja tehokas, ovatko harjoitteluun käytettävä kokonaisaika ja tuen intensiteetti riittäviä, vai onko perusteita vaihtaa ja muokata valittua tuen muotoa ja sen yksilöllisyyden astetta.

REFERENCES

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 69*(2), 334–345.

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2015). Ratio dependence in small number discrimination is affected by the experimental procedure. *Frontiers in Psychology, 6*(1649), 1–13. doi: 10.3389/fpsyg.2015.01649

Ahonen, T., Tuovinen, S., & Leppäsaari, T (2006). *Nopean sarjallisen nimeämisen testi* (2nd ed.). Jyväskylä, Finland: Niilo Mäki Institute.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5). Arlington, VA: American Psychiatric Association.

Ansari, D., Holloway, I. D., Price, G. R., & van Eimeren, L. (2008). Toward a developmental cognitive neuroscience approach to the study of typical and atypical number development. In A. Dowker (ed.), *Mathematical difficulties: psychology and intervention* (pp. 13–43). San Diego, CA: Academic Press.

Aunio, P. (2006). *Number sense in young children -(inter)national group differences and an intervention programme for children with low and average performance* (Doctoral dissertation). Research report 269. Helsinki: University of Helsinki.

Aunio, P., Heiskari, P., van Luit, J. E. H., & Vuorio, J-M. (2015). The development of early numeracy skills in kindergarten in low-, average- and high-performance groups. *Journal of Early Childhood Research, 13*(1), 3–16. doi: 10.1177/1476718X14538722

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20*(5), 427–435. doi: 10.1016/j.lindif.2010.06.003

Aunola, K., Leskinen, E., Lerkkanen, M-K., & Nurmi, J-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*(4), 699–713. doi: 10.1037/0022-0663.96.4.699

Aunola, K., & Räsänen, P. (2007). *The 3-minutes Basic Arithmetic Test.* Niilo Mäki Institute. Unpublished.

Baker, S., Gersten, R., & Lee, D-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103*(1), 51–73.
Retrieved from http://www.jstor.org/stable/1002308

Baroody, A. J., Bajva, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews, 15*, 69–79. doi: 10.1002/ddrr.45

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering at-risk kindergarten children's number sense. *Cognition and Instruction, 30*(4), 437–470. doi: 10.1080/07370008.2012.720152

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal, 1*, 1–41. doi: 10.3102/0002831212473349

Baroody, A. J., Purpura, D. J., Eiland, M. D., & Reid, E. E. (2015). The impact of highly and minimally guided discovery instruction on promoting the learning of reasoning strategies for basic add-1 and doubles combinations. *Early Childhood Research Quarterly, 30*, 93–105. doi: 10.1016/j.ecresq.2014.09.003

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities, 35*(3), 657–670. doi: 10.1016/j.ridd.2013.12.010

Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *Journal of Experimental Child Psychology, 117*, 12–28. doi: 10.1016/j.jecp.2013.08.010

Basic Education Act (628/1998). Amended by Act (642/2010). Suomen laki ja asetuskokoelma.

Batchelor, S., Inglis, M., & Gilmore, C. (2015). Spontaneous focusing on numerosity and the arithmetic advantage. *Learning and Instruction, 40*, 79–88. doi: 10.1016/j.learninstruc.2015.09.005

Becker, H. J. (1992). Computer-based integrated learning systems in the elementary and middle grades: A critical review and synthesis of evaluation reports. *Journal of Educational Computing Research, 8*(1), 1–41. doi: 10.2190/23BC-ME1W-V37U-5TMJ

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*(4), 333–339. doi: 10.1177/00222194050380040901

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46*(2), 545–551. doi: 10.1037/a0017887

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327–336. doi: 10.1016/j.lindif.2011.01.007.

Björn, P. M., Aro, M. T., Koponen, T. K., Fuchs, L. S., & Fuchs, D. H. (2015). The many faces of special education within RTI frameworks in the United States and Finland. *Learning Disability Quarterly*. doi: 10.1177/0731948715594787

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 289–328. doi: 10.1080/19345740802400072

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*(3), 273–293. doi: 10.1207/S15326942DN1903_3

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46*(1), 3–18. doi: 10.1111/j.1469-7610.2005.00374.x

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science, 332*, 1049–1053. doi: 10.1126/science.1201536

Carnine, D. (1997). Instructional design in mathematics for students with learning disabilities. *Journal of Learning Disabilities, 30*(2), 130–141. doi: 10.1177/002221949703000201

Chambers, J. A., & Sprecher, J. W. (1980). Computer assisted instruction: Current trends and critical issues. *Communications of the Association for Computing Machinery, 23*(6), 332–342. doi: 10.1145/358876.358881

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

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology, 33*(3), 300–317. doi: 10.1080/87565640801982387

Christensen, C. A., & Gerber, M. M. (1990). Effectiveness of computerized drill and practice games in teaching basic math facts. *Exceptionality, 1*, 149–165. doi: 10.1080/09362839009524751

Clements, D. H. (2002). Computers in early childhood mathematics. *Contemporary Issues in Early Childhood, 3*(2), 160–180. doi: org/10.2304/ciec.2002.3.2.2

Clements, D. H. (2004). Major themes and recommendations. In D. H. Clements, & J. Sarama (Eds.), *Engaging young children in mathematics, standards for early childhood mathematics education* (pp. 7–72). Mahwah, NJ: Erlbaum.

Cook, B. G., & Cook, S. C. (2013). Unraveling evidence-based practices in special education. *The Journal of Special Education, 47*(2), 71–82. doi: 10.1177/0022466911420877

Cook, B. G., Tankersley, M., & Landrum, T. J. (2009). Determining evidence-based practices in special education. *Exceptional Children, 75*(3), 365–383. Retrieved from http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&sid=75a11115-bab9-4bfd-b9f4-18b70eb30267%40sessionmgr115&hid=118

Cordes, S., & Brannon, E. M: (2008). Quantitative competencies in infancy. *Developmental Science, 11*(6), 803–808. doi: 10.1111/j.1467-7687.2008.00770.x

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. Unpublished doctoral dissertation. Montreal, Canada: McGill University.

Cotton, K. (1991). Computer-assisted instruction. *School Improvement Research Series*, 1–17.

Dehaene, S. (1992). Varietes of numerical abilities. *Cognition, 44*, 1–42.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* Oxford, UK: University Press.

Dencla, M. B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colours, letters and numbers by normal children. *Cortex, 10*(2), 186–202.

De Smedt, B., & Gilmore C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*, 278–292. doi:10.1016/j.jecp.2010.09.003

De Smedt, B., Noël, M-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*, 48–55. doi: 10.1016/j.tine.2013.06.001

Desoete, A., & Grégoire, J. (2006). Numerical competence in young children and in children with mathematics learning disabilities. *Learning and Individual Differences, 16*, 351–367. doi: 10.1016/j.lindif.2006.12.006

Devine, A., Soltész, F., Nobes, A., Goswami, U., & Szücs, D. (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learning and Instruction 27*, 31–39. doi:10.1016/j.learninstruc.2013.02.004

DiPerna, J. C., Lei, P-W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the early childhood longitudinal study - kindergarten cohort. *Journal of Educational Psychology, 99*(2), 369–379. doi: 10.1037/0022-0663.99.2.369

Dowker, A. (2001). Numeracy recovery: A pilot scheme for early intervention with young children with numeracy difficulties. *Support for Learning, 16*, 6–10. doi: 10.1111/1467-9604.00178

Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 324–332. doi: 10.1177/00222194050380040801

Dowker, A. (2008a). Individual differences in numerical abilities in preschoolers. *Developmental Science, 11*(5), 650–654. doi: 10.1111/j.1467-7687.2008.00713.x

Dowker, A. (2008b). Numeracy Recovery with children with arithmetical difficulties: Intervention and research. In A. Dowker (ed.), *Mathematical Difficulties: Psychology and Intervention* (pp. 181-202). Oxford: Elsevier.

Dowker, A. (2009a). Use of derived fact strategies by children with mathematical difficulties. *Cognitive Development, 24*(4), 401–410. doi: 10.1016/j.cogdev.2009.09.005

Dowker, A. (2009b). *What works for children with mathematical difficulties? The effectiveness of intervention schemes*. London: DCSF Publications.

Dowker, A., & Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *The British Psychological Society, Monograph Series 2*(7), 65–81. doi:10.1348/97818543370009X12583699332492

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

Duncan, G. J., & Magnuson, K. (2007). Penny wise and effects size foolish. *Child Development Perspectives, 1*(1), 46–51.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test – Revised.* Circle Pines, MN: American Guidance Service.

Dyson, N., Jordan, N.C., Beliakoff, A., & Hasinger-Das, B. (2015). A kindergarten number sense intervention for low-achieving children with contrasting practice conditions. *Journal for Research in Mathematics Education, 46*(3), 331–370. doi: 10.5951/jresematheduc.46.3.0331

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

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, England: SAGE Publications Ltd.

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory & Cognition, 38*(7), 951–961. doi:10.3758/MC.38.7.951

Finnish National Board of Education. (2010). Core curriculum for preschool education in Finland. Downloadable (in English) from http://www.oph.fi/download/153504_national_core_curriculum_for_pre-primary_education_2010.pdf

Fletcher-Flinn, C. M., & Gravatt, B. (1995). The Efficacy of Computer Assisted Instruction (CAI): A meta-analysis. *Journal of Educational Computing Research*, 12(3), 219–241.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: from identification to intervention.* New York: The Guilford Press.

Frijters, J. C., Lovett, M. W., Steinbach, K. A., Wolf, M., Sevcik, R. A., & Morris, R. D. (2011). Neurocognitive predictors of reading outcomes for children with reading disabilities. *Journal of Learning Disabilities, 44*(2), 150–166. doi: 10.1177/0022219410391185

Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10,* 29–44. doi: 10.1016/j.edurev.2013.05.003

Fuchs, L. S. (2005). Prevention research in mathematics: Improving outcomes, building identification models, and understanding disability. *Journal of Learning Disabilities, 38*(4), 350–352. doi: 10.1177/00222194050380041201

Fuchs, D., & Fuchs, L. S. (2001). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children, 38*(1), 57–61.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly 41*(1), 93–99. doi: 10.1598/RRQ.41.1.4

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13*(4), 204–219.

Fuchs, L. S., & Fuchs, D. (2009). On the importance of a unified model of responsiveness to intervention. *Child Development Perspectives, 3*(1), 41–43.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*(3), 257–269. doi: 10.1177/0022219412442167

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73*(3), 311–330.

Fuchs, L. S., Fuchs, D., Compton, D. L., Wehby, J., Schumacher, R. F., Gersten, R., & Jordan, N. C. (2015). Inclusion versus specialized intervention for very-low-performing students: What does access mean in an era of academic challenge? *Exceptional Children, 81*(2), 134–157. doi: 10.1177/0014402914551743

Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., & Hamlett, C. L. (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*(3), 491–509. doi: 10.1037/0022-0663.100.3.491

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities, 39*(5), 467–475. doi: 10.1177/00222194060390050701

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31*(2), 79–92. doi: 10.2307/20528819

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18*(3), 157–171. doi: 10.1111/1540-5826.00072

Fuson, K. C. (1988). *Children's counting and concepts on number.* New York: Springer-Verlag.

Fuson, K. C. (2009). Avoiding misinterpretations of Piaget and Vygotsky: Mathematical teaching without learning, learning without teaching, or helpful learning-path teaching? *Cognitive Development, 24*, 343–361. doi: 10.1016/j.cogdev.2009.09.009

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*(1), 4–15. doi: 10.1177/00222194040370010201

Geary, D. C. (2011a). 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. (2011b). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics, 32*(3), 250–263. doi:10.1097/DBP.0b013e318209edef

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23–27. doi: 10.1177/0963721412469398

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment, 27*(3), 265–279. doi: 10.1177/0734282908330592 [*Finnish version: Numerosetit* (Koponen, T., Salminen, J., Räsänen, P., & Manninen, J. 2008.) Niilo Mäki Institute. Unpublished.]

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77*, 236–263. doi: 10.1006/jecp.2000.2561

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities, 45*(4), 291–307. doi: 10.1177/0022219410392046

Geary, D., C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*(3), 277–299. doi: 10.1080/87565640801982361

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. R. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202–1242. doi: 10.3102/0034654309334431

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149–164. doi: 10.1177/001440290507100202

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38,* 293–304. doi: 10.1177/00222194050380040301

Griffin, S. (2004). Building number sense with Number Worlds: A mathematics program for young children. *Early Childhood Research Quarterly, 19*, 173–180. doi: 10.1016/j.ecresq.2004.01.012

Griffin, S. (2007). Early intervention for children at risk of developing mathematical learning difficulties. In D. B. Berch, & M. M. M. Mazzocco (Eds.), *Why Is Math So Hard for Some Children?* (pp. 373–395). Baltimore, ML: Paul H. Brookes Publishing Co.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disability, 42*(2), 111–132. doi: 10.1177/0022219408326207

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

Hannula, M. M. (2005). *Spontaneous focusing on numerosity in the development of early mathematical skills.* (Doctoral dissertation). Turun yliopiston julkaisuja B 282. Turku: University of Turku.

Hannula, M. M., & Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learning and Instruction, 15*(3), 237–256. doi: 10.1016/j.learninstruc.2005.04.005

Hannula, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology, 107*(4), 394–406. doi: 10.1016/j.jecp.2010.06.004

Hannula, M. M., Räsänen, P., & Lehtinen, E. (2007). Development of counting skills: Role of spontaneous focusing on numerosity and subitizing-based enumeration. *Mathematical Thinking and Learning, 9*(1), 51–57. doi: 10.1080/10986060709336605

Hannula-Sormunen, M. M., Lehtinen, E., & Räsänen, P. (2015). Preschool children's spontaneous focusing on numerosity, subitizing, and counting skills as predictors of their mathematical performance seven years later at school. *Mathematical Thinking and Learning, 17*(2-3), 155–177. doi: 10.1080/10986065.2015.1016814

Hasselbring, T. S. (1986). Research on the effectiveness of computer-based instruction: a review. *International Review of Education, 32*(3), 313–324. doi: 10.1007/BF02426065

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177. doi: 10.1111/j.1750-8606.2008.00061.x

Hirvonen, R., Tolvanen, A., Aunola, K., & Nurmi, J-E. (2012). The developmental dynamics of task-avoidant behavior and math performance in kindergarten and elementary school. *Learning and Individual Differences, 22*, 715–723. doi: 10.1016/j.lindif.2012.05.014

Holmes, W., & Dowker, A. (2013). Cutch Up Numeracy: a targeted intervention for children who are low-attaining in mathematics. *Research in Mathematics Education, 15*(3), 249–265. doi: 10.1080/14794802.2013.803779

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. doi: 10.1177/001440290507100203

Hunt, J. H., Valentine, C., Bryant, D. P., Pfannenstiel, K. H., & Bryant, B. R. (2015). Supplemental mathematics intervention: How and why special educators intensify intervention for students with learning disabilities. *Remedial and Special Education.* doi: 10.1177/0741932515597293

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*(3), 834–850. doi: 10.1111/1467-8624.00571

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*(1), 36–46. doi: 10.1111/j.1540-5826.2007.00229.x

Jordan, N. C., Kaplan, D., Nabors Oláh, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153–175. doi: 10.1111/j.1467-8624.2006.00862.x

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

Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews, 15*, 60–68. doi: 10.1002/ddrr.46

Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Göbel, S. M., & Grabner, R. H., … Nuerk, H-C. (2013). Dyscalculia from a developmental and differential perspective. *Frontiers in Psychology, 4*(516), 1–5. doi: 10.3389/fpsyg.2013.00516

Kleemans, T., Peeters, M., Segers, E., & Verhoeven, L. (2012.) Child and home predictors of early numeracy skills in kindergarten. *Early Childhood Research Quarterly, 27*, 471–477. doi: 10.1016/j.ecresq.2011.12.004

Kleemans, T., Segers, E., & Verhoeven, L. (2011.) Precursors to numeracy in kindergarners with specific language impairment. *Research in Developmental Disabilities, 32*, 2901–2908. doi: 10.1016/j.ridd.2011.05.013

Kohli, N., Sullivan, A. L., Sadeh, S., & Zopluoglu, C. (2015). Longitudinal mathematics development of students with learning disabilities and students without disabilities: A comparison of linear, quadratic, and piecewise linear mixed effects model. *Journal of School Psychology, 52*, 105–120. doi: 10.1016/j.jsp.2014.12.002

Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J-E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skills. *Journal of Experimental Child Psychology, 97*, 220–241. doi: 10.1016/j.jecp.2007.03.001

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*(1), 162–175. doi: 10.1037/a0029285

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY. A developmental neuropsychological assessment manual*, Finnish Edition: NEPSY II - Lasten neuropsykologinen tutkimus (Korkman, Kirk, & Kemp, 2008). Helsinki, Finland: Psykologien Kustannus Oy

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*, 516–531. doi:10.1016/j.jecp.2009.03.009

Kroesbergen, E. H., & van Luit, J. (2003). Mathematics interventions for children with special educational needs. A meta-analysis. *Remedial and Special Education, 24*(2), 97–114. doi: 10.1177/07419325030240020501

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönman, C., Plangger, F., … von Aster, M. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage, 57*(3), 782–795. doi: 10.1016/j.neuroimage.2011.01.070

Kulik, C-L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*, 75–94. doi: 10.1016/0747-5632(91)90030-5

Kulik, J. A., Kulik, C-L. C., & Bangert-Drowns, R. L. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior, 1*, 59–74.

Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93*(3), 451–455. doi: 10.1037/0022-0663.93.3.451

Käser, T., Baschera, G-M., Kohn, J., Kucian, K., Richtmann, V., Grond, U., … von Aster, M. (2013). Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Frontiers in Psychology, 4*, Advance online publication. doi: 10.3389/fpsyg.2013.00489

Käser, T., Kucian, K., Ringwald, M., Baschera, G. M., von Aster, M., & Gross, M. (2011). Therapy software for enhancing numerical cognition. In J. Özyurt, A. Anschütz, S. Bernholt, & J. Lenk (Eds.), *Interdisciplinary perspectives on cognition, education and the brain* - Hanse-Studies (Vol. 7, pp. 207–216). Oldenburg, Germany: BIS-Verlag.

Lachance, J. A., & Mazzocco, M. M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences, 16*, 195–216. doi: 10-1016/j.lindif.2005.12.001

Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cogitive profiles. *Journal of Experimental Child Psychology, 103*, 309–324. doi: 10.1016/j.jecp.2009.03.006

Landerl, K., Köbel, S. M., & Moll, K. (2013). Core deficit and individual manifestations of developmental dyscalculia (DD): The role of comorbidity. *Trends in Neuroscience and Education, 2*, 38–42. doi: 10.1016/j.tine.2013.06.002

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *Journal of Child Psychology and Psychiatry, 51*(3), 287–294. doi: 10.1111/j.1469-7610.2009.02164.x

LeFevre, J-A., Fast, L., Skwarchuk, S-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*(6), 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x

Lepola, J., Niemi, P., Kuikka, M., & Hannula, M. M. (2005). Cognitive-linguistic skills and motivation as longitudinal predictors of reading and arithmetic achievement: A follow-up study from kindergarten to grade 2. *International Journal of Educational Research, 43*, 250–271. doi: 10.1016/j.ijer.2006.06.005

Leskinen, M., & Salminen, J. (2015). Ennalta ehkäisevä tuki: Suuntaviivoja tutkimusperustaisuudelle. *NMI Bulletin 25*(3), 22–35.

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. doi: 10.1007/s10648-010-9125-8

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*(5), 451–459. doi: 10.1177/0022219408321126

Mattinen, A. (2006). Huomio lukumääriin. Tutkimus 3-vuotiaiden lasten matemaattisten taitojen tukemisesta päiväkodissa (Doctoral dissertation). Turun yliopiston julkaisuja C 247. Turku: University of Turku.

Mazzocco, M. M. M. (2005). Challenges in identifying target skills for math disability screening and intervention. *Journal of Learning Disabilities, 38*(4), 318–323. doi: 10.1177/00222194050380040701

Mazzocco, M. M. M., & Grimm, K. J. (2013). Growth in rapid automatized naming from grades K to 8 in children with math or reading disabilities. *Journal of Learning Disabilities, 46*(6), 517–533. doi: 10.1177/0022219413477475

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia, 53*(1), 218–253. doi: 10.1007/s11881-003-0011-7

Mazzocco, M. M. M., & Räsänen, P. (2013). Contributions of longitudinal studies to evolving definitions and knowledge of developmental dyscalculia. *Trends in Neuroscience and Education, 2*, 65–73. doi: 10.1016/j.tine.2013.05.001

Mazzocco, M. M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice, 20*(3), 142–155. doi: 10.1111/j.1540-5826.2005.00129.x

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471–490. doi: 10.1016/j.ecresq.2006.09.003

McCollister, T. S., Burts, D. C., Wright, V. L., & Hildreth, G. J. (1986). Effects of computer-assisted instruction and teacher-assisted instruction on arithmetic task achievement scores of kindergarten children. *Journal of Educational Research, 80*(2), 121–125.

McMaster, K. N., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71(4),* 445–463.

Mevarech, Z. R., & Rich, Y. (1985). Effects of computer-assisted mathematics instruction on disadvantaged pupils' cognitive and affective development. *Journal of Educational Research, 79*(1), 5–11.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*(3), 272–277.

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and gender differences. *Plos One, 9*(7), 1–9. doi: 10.1371/journal.pone.0103537

Moll, K., Snowling, M. J., Göbel, S. M., & Hulme, C. (2015). Early language and executive skills predict variations in number and arithmetic skills in children at family-risk of dyslexia and typically developing controls. *Learning and Instruction, 38*, 53–62. doi: 10.1016/j-learninstruc.2015.03.004

Mononen, R. (2014). *Early mathematics interventions: Supporting young children with low performance in mathematics* (Doctoral Dissertation). Research report 362. Helsinki: University of Helsinki.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*(1–2), 99–113. doi: 10.1023/B:TRUC.0000021811.66966.1d

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*(4), 306–321. doi: 10.1177/0022219408331037

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: who falls increasingly behind? *Journal of Learning Disabilities, 44*(5), 472–488. doi: 10.1177/0022219411414010

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*(5), 458–478. doi: 10.1177/00222194070400050901

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction, 23*, 125–135. doi: 10.1016/j.learninstruc.2012.08.004

Okolo, C. M. (1991). Learning and behaviorally handicapped students' perceptions of instructional and motivational features of computer assisted instruction. *Journal of Research on Computing in Education, 24*(2), 171–188. doi: 10.1080/08886504.1991.10782000

Okolo, C. M. (1992). The effect of computer-assisted instruction format and initial attitude on the arithmetic facts proficiency and continuing motivation of students with learning disabilities. *Exceptionality, 3*(4), 195–211. doi: 10.1080/09362839209524815

Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2006). Computer-assisted teaching and mathematical learning in Down syndrome children. *Journal of Computer Assisted Learning, 22*, 298–307. doi: 10.1111/j.1365-2729.2006.00179.x

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*, 385–413. doi: 10.1016/j.cognition.2006.04.008

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*, 33–41. doi: 10.1016/j.cognition.2010.03.012

Piazza, M., Mechelli,A., Butterworth, B., & Price. C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *NeuroImage, 15*, 435–446. doi:10.1006/nimg.2001.0980

Polet, J., & Koponen, T. (2011). LukiMat - Oppimisen arviointi: Matematiikan tuen tarpeen tunnistamisen välineet esikouluun. Tekninen opas. Downloadable from http://www.lukimat.fi

Praet, M., & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerized kindergarten interventions: A randomized controlled study. *Teaching and Teacher Education 39*, 56–65. doi: 10.1016/j.tate.2013.12.003

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110–122. doi: 10.1016/j.lindif.2009.10.005

Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology, 4*(141), 649–666. doi: 10.1037/a0027520

Richardson, U., & Lyytinen, H. (2014). The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments 10*(1), 39–60. Downloadable from http://www.humantechnology.jyu.fi/articles/volume10/2014/Richardson_Lyytinen.pdf

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburg (Ed.), *The Development of Mathematical Thinking* (pp. 153–196). New York, NY: Academic Press.

Rousselle, L. & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*, 361–395. doi:10.1016/j.cognition.2006.01.005

Rubinsten, O., & Henik, A. (2009). Developmental dyscalculia: Heterogeneity may not mean different mechanisms. *Trends in Cognitive Science, 13*(2), 92–99. doi: 10.1016/j.tics.2008.11.002

Räsänen, P., Salminen, J. Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development, 24*, 450–472. doi: 10.1016/j.cogdev.2009.09.003

Sahlberg, P. (2007). Education policies for raising student learning: the Finnish approach. *Journal of Education policy, 22*(2), 147–171. doi: 10.1080/02680930601158919

Salaschek, M., Zeuch, N., & Souvignier, E. (2014). Mathematics growth trajectories in first grade: Cumulative vs. compensatory patterns and the role of number sense. *Learning and Individual Differences, 35*, 103–112. doi: 10-1016/j-lindif.2014.06.009

Salminen, J. & Koponen, T. (2010). The Rapid Naming of Small Quantities Test. Niilo Mäki Institute. Unpublished.

Salminen, J., & Koponen, T. (2011). LukiMat - Oppimisen arviointi: Matematiikan oppimisen seurannan välineet. Käsikirja. Niilo Mäki Instituutti. Downloadable from http://www.lukimat.fi

Salminen, J., Räsänen, P., Koponen, T., & Aunio, P. (2008a). The Basic Addition Fluency Test. Niilo Mäki Institute. Unpublished.

Salminen, J., Räsänen, P., Koponen, T., & Aunio, P. (2008b). The Non-Symbolic Story Problems Test. niilo Mäki Institute. Unpublished.

Salminen, J., Räsänen, P., Koponen, T., & Aunio, P. (2008c). The Number Word – Quantity Correspondence Test. Niilo Mäki Institute. Unpublished.

Salonen, P., Lepola, J., Vauras, M., Rauhanummi, T., Lehtinen, E., & Kinnunen, R. (1994). *Diagnostiset testit 3. Motivaatio, Metakognitio ja Matematiikka* [Diagnostic Tests 3: Motivation, metacognition and mathematics]. Turku, Finland: Centre for Learning Research, University of Turku.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children.* New York, NY: Routledge.

Schmidt, M., Weinstein, T., Niemic, R., & Walberg, H. J. (1985–86). Computer-assisted instruction with exceptional children. *The Journal of Special Education, 19*(4), 493–501.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (in press). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science.*

Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2015). Spontaneous non-verbal counting in toddlers. *Developmental Science*, 1–9. doi: 10.1111/desc.12299

Seo, Y-J., & Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education, 53*, 913–928. doi: 10.1016/j.compedu.2009.05.002

Seo, Y-J. & Woo, H. (2010). The identification, implementation, and evaluation of critical user interface design features of computer-assisted instruction programs in mathematics for students with learning disabilities. *Computers & Education, 55*(1), 363–377. doi: 10.1016/j.compedu.2010.02.002

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. doi: 10.3102/0034654307313795

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., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237–250. doi: 10.1111/1467-9280.02438

Skagerlund, K., & Träff, U. (2014). Number processing and heterogeneity of developmental dyscalculia: Subtypes with different cognitive profiles and deficits. *Journal of Learning Disabilities.* doi: 10.1177/0022219414522707

Skwarchuk, S-L., Sowinski, C., & LeFevre, J-A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model. *Journal of Experimental Child Psychology, 121*, 63–84. doi: 10.1016/j.jecp.2013.11.006

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best evidence synthesis. *Review of Educational Research, 78*(3), 427–455. doi: 10.3102/0034654308317473

Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities, 36*(1), 24–33. doi: 10.1177/00222194030360010401

Starkey, P., & Cooper, R. G. Jr, (1980). Perception of numbers by human infants. *Science, 210*(4473), 1033–1035. doi: 10.1126/science.7434014

Supekar, K., Swigart, A. G., Tenison, C., Jolles, D. D., Rosenberg-Lee, M., Fuchs, L., & Menon, V. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. *PNAS, 110*(20), 8230–8235. doi: 10.1073/pnas.1222154110

Suppes, P. (1968). Computer-assisted instruction: An overwiev of operations and problems. In A. J. H. Morrel (Ed.), *Information processing 68.2., Proceedings of IFIP Congress, Edinburgh* (pp. 1103–1113). Retrieved from http://suppescorpus.stanford.edu/articles/comped/102.pdf

Suppes, P., & Jerman, M. (1969). Computer assisted instruction at Stanford. *Educational Technology, 9*, 22–24.

Suppes, P., & Morningstar, M. (1969). Computer-assisted instruction. *Science, 166*, 343–350.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes.

Review of Educational Research, 68(3), 277–321. doi: 10.3102/00346543068003277

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2013). Intervention fidelity in special and general education research journals. *The Journal of Special Education, 47*(1), 3–13. doi: 10.1177/0022466911419516

Toll, S., van Viersen, S., Kroesbergen, E. H., & van Luit, J. E. H. (2015). The development of (non-)symbolic comparison skills throughout kindergarten and their relations with basic mathematical skills. *Learning and Individual Differences, 38*, 10–17. doi: 10.1016/j.lindif.2014.12.006

Torgesen, J. K. (2000). Individual differences in response to early intervention in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*(1), 55–64.

Van Luit, J. E. H., & Schopman, E. A. M. (2000). Improving early numeracy of young children with special educational needs. *Remedial and Special Education, 21*(1), 27–40. doi: 10.1177/074193250002100105

Van Luit, J. E. H., van de Rijt, B. A. M., & Aunio, P. (2006). *Early Numeracy test, Finnish Edition: [Lukukäsitetesti]*. Helsinki, Finland: Psykologien Kustannus Oy.

Von Aster, M. G. (2000). Developmental cognitive neuropsychology of number processing and calculation: varietes of developmental dyscalculia. *European Child & Adolecent Psychiatry, 9*(II), 41–57. doi: 10.1007/s007870070008

Von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology, 49*(11), 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

What Works Clearinghouse. The website on scientific evidence related to education practice and policy. The website is administered by the Institute of Education Sciences, U.S. http://ies.ed.gov/ncee/wwc/ (sited 13.9.2015)

WHO (World Health Organization). (2015). *International Statistical Classification of Diseases and Related Health Problems* (10th Revision; ICD-10). Chapter V: Mental and behavioural disorders, Disorders of psychological development, F81.2 Specific disorder of arithmetical skills.

Willcutt, E. G., Pennington, B. F., Duncan, L., Smith, S. D., Keenan, J. M., Wadsworth, S., DeFries, J. C., & Olson, R. K. (2010). Understanding the complex etiologies of developmental disorders: Behavioral and molecular genetic approaches. *Journal of Developmental and Behavioral Pediatrics, 31*(7), 533–544. doi: 10.1097/DBP.0b013e3181ef42a1

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *Journal of Learning Disability, 46*(6), 500–516. doi: 10.1177/0022219413477476

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education, 3*(4), 224–234. doi: 10.1111/j.1751-228X.2009.01075.x

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006). Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions, 2*(19), 1–14. doi: 10.1186/1744-9081-2-19

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of The Number Race, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions, 2*(20), 1–36. doi: 10.1186/1744-9081-2-20

Wong, T. T.-Y., Ho, C. S.-H., & Tang, J. (2015). Defective number sense of impaired access? Differential impairments in different subgroups of children with mathematics difficulties. *Journal of Learning Disabilities*. doi: 10.1177/0022219415588851

Wright, R. J. (2003). Mathematics recovery: A program of intervention in early number learning. *Australian Journal of Learning Disabilities, 8*(4), 6–11. doi: 10.1080/19404150309546744

Wright, R. J., Martland, J., & Stafford, A. K. (2008). *Early numeracy: Assessment for teaching and intervention* (2nd ed.). London: Paul Chapman Publishing.

Wylie, J., Jordan, J. A., & Mulhern, G. (2012). Strategic development in exact calculation: Group and individual differences in four achievement subtypes. *Journal of Experimental Child Psychology, 113*(1), 112–130. doi: 10.1016/j.jecp.2012.05.005

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*, 155–193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24*, 220–251.

Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M-K., Nurmi, J-E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development, 85*(3), 1091–1107. doi: 10.1111/cdev.12173

APPENDIXES

Appendix 1

Training content of the original version of GraphoGame Math

Range	Task type	Concept
1–3	approximate comparison	smallest, biggest, the most, the least
	number word - quantity mapping	number words 1–3
	number word - quantity - number symbol mapping	
2–5	number word - quantity mapping	number words 2–5
	number word - quantity - number symbol mapping	
	number word - number symbol mapping	
	number after	
	number before	
	symbolic single-digit addition	
	symbolic single-digit subtraction	
3–7	number word - quantity mapping	number words 3–7
	number word - quantity - number symbol mapping	
7–10	number word - quantity - number symbol mapping	number words 7–10
1–10	number word - quantity - number symbol mapping	number words 1–10
5–10	number word - number symbol mapping	number words 5–10
2–10	number word - number symbol mapping	number words 2–10
	symbolic single-digit addition	
	symbolic single-digit addition	
	symbolic single-digit addition	
	symbolic single-digit addition	
	symbolic single-digit subtraction	
10–15	number word - number symbol mapping	number words 10–15
…	…	…
until 30	training continued with similar contents	number words 15–30

Note. Range = Corresponding number area used in specified training level

Appendix 2

Training content of the second version of GraphoGame Math

Range	Task type	Concept
0–5	one-to-one correspondence	equal amount
	approximate non-symbolic comparison	more, less the most, the least
	non-symbolic ordering	order, first, smallest and biggest amount
	number word - quantity mapping (objects)	number words 1–5
	number word - quantity mapping (dots)	
	quantity - quantity mapping (sounds and objects)	
	object counting (sub-set from a set)	equal amount
	exact non-symbolic comparison	one more, one less two more, two less
	object counting (composing two sets of objects/dots)	equal amount, altogether
	object counting (decomposing set of objects/dots)	
	story problems	one; two; three more, first, after, altogether, number words 1–5
	mental addition	adding, altogether, number words 1–5
	number word - quantity - number symbol mapping	number word, quantity, number symbol, written symbol, number words 1–5
	number word - quantity - number symbol mapping	number word zero, no objects, number symbol zero
…	…	…
	symbolic comparison	smaller number, larger number
	symbolic ordering	order of the numbers, number sequence
	number before - after	number words 10–15
	exact symbolic comparison	one more, one less two more, two less
	symbolic composing and decomposing	equal amount, altogether
	non-symbolic addition with symbolic addition	adding, equal, altogether, operational symbols (plus, equal), number words 1–5
	symbolic addition	
…	…	…
until 10	similar contents repeated by weighting addition	

Note. Range = Corresponding number area used in specified training level. … = Non-symbolic contents were repeated before symbolic representations along basic addition were in the focus.

Appendix 3 (A–C)

Training content of the third version of GraphoGame Math

Appendix 3 A

Range	Task type in Number Concept Component	Concept
0–5	Tutorials (presented before actual training)	more, less, the most, the least, equal, number symbol, number word, one more, one less, two more, two less, number words 1–2, number symbol 2
0–20	Assessment point (presented before actual training)	equal amount, correspondence, one more, one less, two more, two less
0–5	one-to-one correspondence	equal amount
	approximate non-symbolic and symbolic comparison	more, less
	number word - quantity - number symbol mapping	number words 0–5
	approximate non-symbolic and symbolic comparison	most, least
	exact non-symbolic and symbolic comparison	one more, one less two more, two less
3–7	contents repeated based on assessment point	(see above)
6–10	contents repeated based on assessment point	
0–10	contents repeated based on assessment point	
10–20	contents repeated based on assessment point	

Note. Range = Corresponding number area used in specified training level.

Appendix 3 B

Range	Task type in Number Sequence Component	Concept
0–5	Tutorials (presented before actual training)	number sequence, order, forward, backward, before, after, the largest amount first, the smallest amount first, from largest to smallest amount, from smallest to largest amount, number words 1–3, (number symbols 1–5)
0–20	Assessment point (presented before actual training)	number sequence forward, and backward
0–5	continuing non-symbolic and symbolic sequences	forward, from smallest to largest amount and vice versa
	fulfilling non-symbolic and symbolic sequences	
	ordering objects and number symbols	order, the smallest amount first, the largest amount first
	number before - after	number words 0–5
	“number line task” with starting point only	number sequence, missing number symbol
3–7	contents repeated based on assessment point	(see above)
6–10	contents repeated based on assessment point	
0–10	contents repeated based on assessment point	
10–20	contents repeated based on assessment point	

Note. Range = Corresponding number area used in specified training level.

Appendix 3 C

Range	Task type in Basic Addition Component	Concept
0–5	Tutorials (presented before actual training)	adding, altogether, equal amount, add, plus, sum, number words 1–3, (operational symbols plus and equal, number symbols 1–5)
0–20	Assessment point (presented before actual training)	fulfilling symbolic addition tasks
0–5	non-symbolic and symbolic composing	quantity, number symbol, equal amount, altogether
	non-symbolic and symbolic decomposing	(see above), set, subset
	fulfilling non-symbolic and symbolic addition task	addition fact, altogether, operational symbols (plus, equal)
	constructing symbolic addition	number symbols, addition fact, operational symbols plus and equal
	mental addition	number words 0–5, adding, makes altogether, sum
	comparison of presented addition facts	addition fact, sum, response, the largest
	ordering of presented addition facts by sum	
	correspondence between sum and addition fact	number words 0–5, addition fact, sum
3–7	contents repeated based on assessment point	(see above)
6–10	contents repeated based on assessment point	
0–10	contents repeated based on assessment point	
10–20	contents repeated based on assessment point	

Note. Range = Corresponding number area used in specified training level.

ORIGINAL PAPERS

I

PREVENTIVE SUPPORT FOR KINDERGARTENERS MOST AT-RISK FOR MATHEMATICS DIFFICULTIES: COMPUTER-ASSISTED INTERVENTION

by

Jonna Salminen, Tuire Koponen, Pekka Räsänen, & Mikko Aro, (2015)

Mathematical Thinking and Learning

Reproduced with kind permission from the Taylor & Francis Group.

Preventive Support for Kindergarteners Most At-Risk for Mathematics Difficulties: Computer-Assisted Intervention

Jonna Salminen and Tuire Koponen

Pekka Räsänen and Mikko Aro

Jonna Salminen, Department of Education, Special Education, University of Jyväskylä; Tuire Koponen, Niilo Mäki Institute, Jyväskylä; Pekka Räsänen, Niilo Mäki Institute, Jyväskylä; Mikko Aro, Department of Education, Special Education, University of Jyväskylä.

The data for this study was collected as part of the LukiMat project. The project is supported by the Ministry of Education and Culture in Finland since 2007. I wish to thank the participating day care centers, the kindergarteners and their teachers as well as the parents. The proposals of the Finnish Advisory Board on Research Integrity concerning the ethical questions relating to research were followed during the study. The study will be part of a doctoral dissertation.

Correspondence concerning this article should be addressed to Jonna Salminen, Department of Education, Special Education, University of Jyväskylä, PO box 35 40014 Jyväskylä, Finland, jonna.salminen@jyu.fi

Preventive Support for Kindergarteners Most At-Risk for Mathematics Difficulties: Computer-Assisted Intervention

Weaknesses in early number skills have been found to be a risk factor for later difficulties in mathematical performance. Nevertheless, only a few intervention studies with young children have been published. In this study, the responsiveness to early support in kindergarteners with most severe difficulties was examined with two different computer programs. Two intervention groups were matched by age, visuo-spatial and phonological working memory, as well as early number skills. After a short and intensive computerized intervention, the results indicated significant intervention effects for verbal counting Wilcoxon *ES (r)* = .46, and dot counting fluency, *r* = .52, when practiced with GraphoGame Math, as well as for basic arithmetic, *r* = .63, when practiced with Number Race. The findings suggest that a targeted computerized practice can produce specific training effects in kindergarteners most at-risk for mathematics difficulties. The results are discussed with regard to practical implications for educational game development.

Mathematics difficulties (MD), typically seen as deficits in basic arithmetic (Geary, 2011), can be predicted rather reliably from kindergarten (e.g., McClelland, Acock, & Morrison, 2006; Morgan, Farkas, & Wu, 2009; Murphy, Mazzocco, Hanich, & Early, 2007). It is known that the difference in early number skills is notable between low-performing children and typically performing ones, and the gap seems to grow during the following years (e.g., Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Desoete & Grégoire, 2006; Jordan, Kaplan, Locuniak, & Ramineni, 2007). According to Geary (2011) the children with the most severe difficulties (scoring below the 10th percentile on a standardized mathematic achievement test) differ from their low achieving age-peers (scoring between the 11th and 25th percentile) as having even more persistent deficits (see also Chong & Siegel, 2008; Geary, 2013; Geary, Hoard, & Bailey, 2012). Such deficits or delays can be seen in processing numbers, learning arithmetic procedures, retrieving arithmetic facts, and in working memory (Geary, 2011). The growth rates of these sub-groups (performance below 10th, or between 11th and 25th percentiles) differ from each other when proceeding from kindergarten to later, primary grade levels (e.g., Geary, Hoard, Nugent, & Byrd-Craven, 2008; Morgan et al., 2009; Murphy et al., 2007). The children with the most severe difficulties fall even more behind. These findings underline the need to develop and investigate tools to effectively support such kindergarteners, especially those most at-risk for severe disability in learning arithmetic which per se is a critical feature of mathematics difficulties (Butterworth, Varma, & Laurillard, 2011; Geary, 2011). In order to improve such tools, it is important to understand what the key early number skills are, and how do they develop.

EARLY NUMBER SKILLS

It has been observed that young children can recognize small quantities by subitizing (Dehaene, 2011; Starkey & Cooper, 1980). This term describes the process by which children quickly see how many items are in a given space. Researchers have noted that this ability is inborn and is restricted to small quantities (e.g., Piazza, Mechelli, Butterworth, & Price, 2002; Reeve, Reynolds, Humberstone, & Butterworth, 2012). The ability to discriminate small quantities is a prerequisite for learning the meaning of small number words such as *one*, *two*, and *three*, which are typically learned by children by the time they are 3.5 years old (e.g., Wynn, 1990). Enumerating quantities of more than three requires counting (see Dehaene, 2011, p. 56–59), or detecting the number of subsets for forming the whole quantity that is conceptual subitizing (Sarama & Clements, 2009).

The development of verbal counting involves several steps. Young children start verbal counting by repeating number words first as a rhyme, starting from number one, without necessarily knowing their numerical values (Fuson, 1988). After that, number words can be perceived as separate items and children learn the correct order of the small number words (Fuson, 1988). The development of verbal counting skill is important for later arithmetic performance at school age (e.g., Desoete & Grégoire, 2006; Koponen, Salmi, Eklund, & Aro, 2013; Lepola, Niemi, Kuikka, & Hannula, 2005). More specifically, it has been found that verbal counting predicts both procedural calculation and fact retrieval fluency (Koponen, Aunola, Ahonen, & Nurmi, 2007).

During early number skill development, the knowledge of number word sequence becomes integrated with the cardinality meaning of number words in order to be used in counting as a tool for determining the exact magnitude of objects, that is object counting. Hence, exact object counting demands understanding that the last counted number word represents the total amount of items, as well as other counting principles (e.g., one-to-one correspondence between count words and objects being counted, stable order of the count words and cardinality) (e.g., Clements, 2004; Sarama & Clements, 2009; Gelman & Gallistel, 1978; Wynn, 1992). It seems that object counting fluency remains rather stable between different skill-level groups (Reeve et al., 2012; 6 year old kindergarteners were followed for 5 years). Also, early number competence including both verbal counting and object counting as well as other numerical skills is found to highly predict later mathematics achievement at school (Jordan et al, 2007; Mazzocco & Thompson, 2005), and more specifically calculation fluency (Locuniak & Jordan, 2008) as well as applied problem solving

(Jordan, Kaplan, Ramineni, & Locuniak, 2009). Furthermore, when exact object counting skills develop, they can be used for composing and decomposing quantities (Baroody, 1987; Sarama & Clements, 2009) as well as for adding and taking away. These skills are prerequisites for fluency in later calculation skills. Respectively, non-fluency with arithmetic combinations is a critical characteristic of MD (Gersten, Jordan, & Flojo, 2005).

Children with atypical number skill development seem to have deficits in symbolic comparison, while the evidence for similar deficits in non-symbolic comparison remains rather contradictory (see Bartelet, Vaessen, Blomert, & Ansari, 2014; De Smedt, Noël, Gilmore, & Ansari, 2013). Kindergarteners with the weakest performance in counting and other early number skills (performance below 10%) appear to have a slower rate of growth than low performing children (performance from 11% to 25%) through the third grade, a trend that might continue in later grades (Murphy et al., 2007). Overall, the definitions of at-risk status (e.g., Mazzocco, 2005) and MD (e.g., Butterworth et al., 2011; Fuchs, 2005; Geary, 2004, 2011, 2013) are currently becoming more specific. Understanding the heterogeneity among the lowest performers has implications for early identification and suggests the need for intensified and individualized support (Geary, 2011).

COMPUTER-ASSISTED INTERVENTION

Recently, the use of computers in daily kindergarten activities has increased greatly. A variety of computers, laptops, and tablets are available with downward trend of costs. Researchers have considered the benefits of computer use for mathematics learning for decades. Computers can provide developmentally appropriate experiences for children (Clements, 2002), as well as motivate (Becker, 1992; De Smedt et al., 2013) and activate (Chambers & Sprecher, 1980) children. Furthermore, they provide immediate and continuous feedback as well as repetitive practice, all of which are found to be important for children with weak skills (Hasselbring, 1986). Li and Ma (2010) recently concluded that computer technology is more effective with regard to mathematics achievement in special needs students than in general education students. As Slavin and Lake

(2008) summarized, a reasonable use of computers can provide mathematics exercises tailored to individual needs, and adaptive software can identify child's strengths and weaknesses to fill possible gaps. These findings are consistent with other meta-analyses and reviews in which low achievers and at-risk learners progressed more than other students when computer-assisted intervention (CAI) was used (e.g., Kroesbergen & Van Luit, 2003; Kulik & Kulik, 1991; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009). However, Räsänen (in press) reviewed that in recent decades the main trend of the effectiveness of CAI on numerical skills has been declining, not increasing. Similar observations on the trend have been made in three recent meta-analyses (Cheung & Slavin, 2013; Christmann & Badgett, 2003; Li & Ma, 2010).

The effectiveness of CAI has been difficult to establish in research literature due to varying study designs, target group definitions, and reports of the content being practiced (e.g., Räsänen et al., 2009; Seo & Bryant, 2009; Slavin & Lake, 2008). For this reason, concerning children in primary grades, Räsänen et al. (2009) were able to calculate effect sizes for only five CAI studies in which pre- and post-test scores with standard deviations for both intervention and control groups were reported. Seo and Bryant (2009) also faced several methodological problems in analyzing the effects of CAI studies in children with learning disabilities (see also Cheung & Slavin, 2013; Slavin & Lake, 2008).

There are only a few previous CAI studies of early number skills for young children (5−7-year olds) at-risk for difficulties in learning mathematics. Baroody, Eiland, Purpura, and Reid (2013) recently reported highly robust CAI effects for children at-risk of not learning the add-1 rule of basic addition. These children received the intervention in two stages: first, children played manual games, and then they had guided computer practice sessions. Researchers have conducted a similar study with significant effects for kindergarteners regarding the add-0 and add-1 rules (Baroody, Eiland, Purpura, & Reid, 2012). With the first graders at-risk for reading and math difficulties, the effect of CAI was significant for addition fact fluency (sum ≤ 18) (Fuchs et al., 2006). Children's basic arithmetic (Praet & Desoete, 2014), numeral recognition (McCollister, Burts, Wright, & Hildreth, 1986), enumeration (Ortega-Tudela & Gómez-Ariza, 2006), and symbolic comparison (Wilson, Dehaene, Dubois, & Fayol, 2009) skills have also been significantly enhanced by CAI. Even in younger children (4-year olds) studies have shown positive effects on pre-mathematical knowledge (Howard, Watson, Brinkley, & Ingels-Young, 1994; Elliot & Hall, 1997). On the other hand, Din and Calao (2001) found no statistically significant results in mathematics when low SES children played several educational video games. Detailed information of these CAI intervention studies are summarized in Table 1.

Insert Table 1 here.

In the literature regarding intervention effectiveness, a central issue is the duration and intensity of the practice. In CAIs lasting for several months, a semester, or a whole year, the effects seem to be less clear than in shorter interventions of four weeks or less, irrespective of target group age (e.g., Kroesbergen & Van Luit, 2003; Kulik & Kulik, 1991). More focused interventions to specific skills, higher intensity, and more homogeneous target groups may explain the larger effect sizes in short interventions (Räsänen, in press). However, the finding also suggests that even very short but intensive interventions can be used to produce significant gains if the content of the practice is aligned with the needs of the learner.

THE CURRENT STUDY

This study examines the effects of two freely downloadable, adaptive mathematical computer programs on kindergarteners most at-risk for mathematics difficulties. Here, most at-risk status signifies poor performance in verbal counting (≤ the 10th percentile), along with significantly slower dot counting, weaker basic arithmetic, as well as visuo-spatial and phonological working memory skills as compared to a reference group (not most at-risk for MD). In the current study, verbal counting level was used as inclusion criterion because it seems to be a strong predictor of later arithmetic achievement at school (Aunio & Niemivirta, 2010; Aunola et al., 2004; Koponen et al., 2013; Lepola et al., 2005). Verbal counting was assessed with similar types of tasks as the aforementioned studies.

There were also other reasons for selecting verbal counting instead of all early number skills as inclusion criterion. We did not use object counting as inclusion criterion because it seems to be more effective at differentiating between the lowest and typically achieving children at somewhat older age levels (approximately 8−14 years old; e.g., De Smedt, et al., 2013). From a more technical point of view, the performance in our number comparison task was not used as inclusion criterion. The children were asked to determine, as quickly as possible, which of the two presented symbols was larger on the computer screen (on the left or right side on the screen) by clicking the respective mouse button. The time pressure increased the number of errors and even caused guessing in the most at-risk children. Thus, neither the fluency in number comparison nor the accuracy (not properly measured in the current study) were analyzed further, even though number comparison seems to differentiate children with and without (risk for) MD, and is related to arithmetic skills (e.g., Bartelet, Vaessen, et al., 2014; Skagerlund &Träff, 2014). Further, fluency in basic arithmetic at kindergarten age is related to familiarity with arithmetic symbols, as reflected in low general performance level at the pre-assessments in the current study. Therefore, basic arithmetic was not suitable as inclusion criterion.

The main purpose of this study was to examine if short and intensive practice with computer programs can support the early number skills (verbal counting, object counting or basic arithmetic) in kindergarteners most at-risk for MD. Here, short and intensive practice period means training for three weeks, 10−15 minutes per day (c.f., durations and intensiveness in earlier studies described in Table 1). We also examined the between-condition differences in potential intervention gain scores. Finally, we examined the association of the gain scores with the total intervention exposure. In this study, two intervention conditions were used. One focused on exact numerical processes, and the other on approximate numerical processes. Therefore, the total intervention exposure was contrasted with the potential positive gains within conditions. Due to the fact that the evidence of CAI effectiveness in kindergarteners most at-risk for MD is still largely missing, exact hypotheses were not set for the current study.

METHOD

Participants

To conduct our study, we first obtained written permission from a municipal official in charge of day care in a city in Eastern Finland. Next, the official recruited voluntary teachers from day care centers to operate as coordinators for the intervention. We requested written permission from parents whose child took part in the kindergarten curriculum at any of these day care centers (*12* day care centers, altogether *236* kindergarteners). We informed parents of the purpose of the study, and of their right to discontinue the participation at any point. Of the resulting group of children, candidates (n = 30) were nominated into intervention group based on the teachers' observations of who needed extra support for early number skills. If the original number of children was 10 or less per kindergarten group, the teacher was asked to nominate one candidate. If the number of children was 11−21; 22−35; 36 or more per group, the teacher was asked to nominate 2; 3; and 4 candidates, respectively. To form the reference group (n = 30), teachers also nominated one peer-control for each candidate from the same kindergarten group. The peer-control was selected on the basis of having the nearest birthday to the candidate, and of not being in need of extra support for early number skills. Nomination was followed by two individual pre-tests of cognitive abilities, early number skills, and control measures for the candidates and peer-controls that were the reference children. All assessments were administered by the first author and a research assistant. Both have experience assessing young children, but were unfamiliar with the participants in this study. The sample was homogenous in cultural background, and all participants were native speakers of Finnish.

The findings concerning the total sample of children (*n* = 60) have earlier been published in a separate study (Räsänen et al., 2009). For the current study, only those target children who performed below the 10th percentile of the reference group (not most at-risk for MD) in verbal counting were included. Verbal counting (count on, count backward, and skip count by 2) was used as criterion task because of its importance in early number skill development, and in learning more complex mathematics at school age. Based on the inclusion criteria, the study sample consisted of 17 intervention children (7 % of the original sample; *n* = 236), and the reference group (*n* = 30; 13 boys, 17 girls; mean age = 78.8 months, *SD* = 3.3) was used only to determine the risk level in verbal counting, and for testing the test−retest reliabilities for early number skill measures. The children without the risk status for MD typically have mastered prerequisite numerical skills at kindergarten age. For example, in our study 23 of 30 children in the reference group reached the maximum score in counting skills at post-assessment (without extra support). For this reason, the reference group data are not analyzed further, and the group comparisons were not carried out between intervention and reference conditions.

In the current study, poor performance in verbal counting means that the majority of our intervention children (*n* = 17) could not even count correctly up to 20 (64.7 % of participants), and they also failed at more complex tasks, such as counting backward from 12 to 8 (76.5 %) or skip counting by 2 up to 10 (70.6 %) in February, during their kindergarten year. This skill level was comparable to the level of children scoring below the 10th percentile in a normative sample of Finnish kindergarteners (*n* = 502) that was collected for a nationally normed assessment test that included number knowledge, number concept, verbal counting, and non-verbal calculation tasks (see technical manual; Polet & Koponen, 2011). In this normative sample, collected in January-February during kindergarten year, 61.2 % of the poorest performers (lowest 10 %, *n* = 49) could not count on up to 20, 85.7 % could not count backward from 12 to 8, and 87.8 % could not skip count by 2 up to 10. Among the rest of the children (i.e. performance above the 10th percentile; *n* =

453) the corresponding failure percentages were 3.8 %; 15.9 %; and 21.2 %. In the reference group of the current study (not most at-risk for MD, *n* = 30) the respective percentages were 6.7 %; 16.7 %; and 16.7 %.

For ethical reasons, all parents provided written permission for their child to participate in both the assessments and the intervention. All parents were also informed of the research project as follows: The Ministry of Education and Culture in Finland (2007−2013) has funded a research project during which a research-based web service for learning challenges in early reading and mathematics will be created, and the effectiveness of certain educational computer games for early support will be studied. In addition, all children in the 12 participating day care centers were allowed to use the intervention programs after the actual study if their parents gave consent that they could do so.

Design and Materials

The study took place at day care centers for six weeks from February to April. During this period, all 17 participants followed the normal kindergarten curriculum. According to the Finnish National Board of Education (2010; downloadable in English), the purpose of Finnish pre-primary education is that “the child develops learning-to-learn skills and positive self-image; as well as acquires basic skills, knowledge and capabilities from different areas of learning in accordance with their age and abilities”. Understanding of concepts, classification, comparison, and sorting are specified as objectives for early mathematics (p. 11−12). Pre-primary education also aims to develop children’s concentration, listening, communication and thinking skills. The children participate in pre-primary educational activities for five days a week, three hours per day. Usually, formal activities include some training for learning letter names and sounds, as well as number symbols. The activities aim also to support social skills: how to follow instructions, how to work in a group, how to co-operate with peers, and how to take care of oneself and one’s own responsibilities. In Finland, 97−98 % of the cohort takes part in the free pre-primary education (The Finnish Ministry of Education and Culture, 2013).

The first pre-test consisted of two tasks for assessing more general and non-domain specific skill levels of the intervention groups (visuo-spatial and phonological working memory); four tasks for assessing early number skills (verbal counting, dot counting fluency, number comparison, basic arithmetic); and one control task unrelated to the intervention (rapid naming). The second pre-test consisted of the aforementioned four number skill tasks and rapid naming. After these tests, the kindergarteners most at-risk (17) were randomly divided into two intervention conditions. Therefore, 9 children (7 boys, 2 girls; mean age = 80.1 months, *SD* = 4.5) were instructed to practice with GraphoGame Math (GGM group), and 8 children (4 boys, 4 girls; mean age = 78.4 months, *SD* = 4.1) to practice with Number Race (NR group). At the beginning of the intervention there were no significant differences in the visuo-spatial skills or phonological working memory, early number skills, or the control task between these two groups (see Table 2). Both groups received intensive intervention for 3 weeks, for 10−15 minutes per day. Finally, all children were post-tested using the aforementioned four early number skill tasks and rapid naming. As mentioned earlier, due to the very low accuracy of identifying number symbols, the number comparison task was excluded from our analyses.

Intervention Conditions

Both intervention tools − GraphoGame Math, (in Finnish and Swedish) and Number Race (open source for multiple languages) − are freely available for children, teachers, and parents. An updated version of GGM can be downloaded from an online educational service (www.lukimat.fi), and NR has its own website (http://thenumberrace.com/nr/home.php) from which a detailed user guide can be downloaded.

GraphoGame Math. Originally GraphoGame Math (GGM) has been designed as part of the GraphoGame project at the University of Jyväskylä in Finland (see Richardson & Lyytinen, 2014).

GGM is targeted primarily at children between 6 and 8 years old. The main purpose of the game version we used in the current study was to support acquisition of basic mathematical concepts and skills, such as, dot counting; the correspondence of number word, quantity and number symbol; basic addition; and basic subtraction skills. GGM consisted of several different tasks that were presented in 50 fields of game content, with approximately 1000 items in total. In all trials the child was instructed to respond by choosing the corresponding visual stimulus according to an auditory cue by clicking on the correct item presented on the screen among incorrect alternatives using the mouse's left button.

GGM included tasks in which the exact relations between numbers were practiced. For example, one type of task required the child to identify a correct number neighbour for a verbally presented number word (number before / number after). This activity was intended to strengthen especially the child's verbal number list (see Fuson, 2009; c.f., Wright, 2003), and thus, verbal counting. GGM also aims at practicing object counting and cardinality through tasks in which the child heard a number word (e.g., "four"), and the ball with the corresponding amount of dots (among other balls with different amounts of dots) had to be clicked (here, four dots). Finally, in GGM, basic arithmetic was practiced through tasks in which the child hears a sum (e.g., "five"), and the ball with the corresponding calculation (e.g., 4 + 1) must be clicked. Analogously, basic subtraction was practiced through tasks in which the child hears a difference (e.g., "two"), and the ball with the corresponding calculation (e.g., 4 - 2) must be clicked. Each task in GGM included a time pressure element created by the slow descent of visual objects on the screen. The child needed to choose the corresponding visual stimulus according to an auditory cue before the stimuli (the correct one among the distractors) had fallen down to be eaten by a "pac-man"-like game figure.

The adaptation in GGM was based on gradually increasing complexity of the content (starting from non-symbolic comparisons and continuing to object counting; number concept training; number neighbors activation; symbolic comparisons; and basic arithmetic). Also, the number range widened gradually (1−3; 1−6; 1−10; 10−20; 20−30), and the better the child performed the more alternatives (as distractors) appeared on the screen. The adaptation algorithm aimed at keeping the individual accuracy rate at around 85%, which meant that GGM kept the child practicing at certain sub-task until the child managed to reach the pre-determined performance level, before letting the child to proceed for the more demanding training of the next sub-skill.

GGM gave immediate, continuous, and delayed feedback. After a successful trial, the child heard a sound signalling a correct response. The selected stimulus stopped, and a yellow star outline appeared, while the incorrect stimuli continued to fall down. After an unsuccessful trial, the child heard a sound signalling an incorrect response, and the incorrect stimulus stopped, while the correct stimulus got a green outlining. After a predefined set of trials, the child received feedback according to the success during the set; this feedback came in the form of butterflies whose colors indicated the child’s accuracy level. The child also saw the total playing time as a progressive bar on the screen.

Number Race. Number Race (NR) is aimed primarily at children between 5 and 8 years old. The original purpose of NR was to remediate dyscalculia by enhancing quantity representation (Wilson, Dehaene, et al., 2006; Wilson, Revkin, et al., 2006). More specifically, Number Race aims to enhance and automatize number processing, the mental number line, as well as skills in counting, basic addition, and subtraction (c.f., designers’ definitions; www.thenumberrace.com; see “How it works”). Within the game, the child is instructed to choose the larger of the two quantities presented visually by concrete objects (coins or coconuts), symbols, or basic addition and/or subtraction calculations (see also Wilson et al., 2009, p. 227).

The NR has been developed specifically to support the learning of children with MD. In this study, we sought to examine the specific effects of NR practice on early number skills in kindergarteners most at-risk for MD. In NR, verbal counting and verbal number lists were implicitly practiced. After each selection the child made between two presented quantities (“selection screen"), a race track appeared ("game board screen"). The child needed to click a square on the track whose order on a path corresponded to the number of quantities selected. After that, the game moved the child's character on the track (a non-numerical path) while simultaneously repeating the number words aloud. The child also clicked a square on the track which corresponded to the number of quantities the enemy character received. This was followed by the aforementioned action. Although dot counting was not explicitly practiced, children could use counting or conceptual subitizing in tasks where they were supposed to select the larger of the two presented set of objects (with a range of 1−9). Finally, basic arithmetic was practiced with a similar type of selection task during which the child saw two arithmetic calculations instead of objects/number symbols. The child had to select the one that produced the larger solution. The calculations were presented both as addition (e.g., 2 + 1 vs. 3 + 2); and subtraction (4 - 1 vs. 3 - 2) tasks; or the two task types were mixed (e.g., 2 + 2 vs. 3 - 1).

The adaptation in NR was based on numerical distance, notation, and time pressure being related to the child's performance. For example, the differentiation was supposed to be easier between two distant quantities than between two closer ones (see also Dehaene, 2011, pp. 60−61). As such, in NR, numerical notation changed sensitively between concrete and more complex notations. In terms of the time pressure, after a certain number of successful trials, the enemy character (located on the top of the screen) moved actively for being quicker than the child in reaching the larger amount of the two quantities. In the version of NR we used in our study, the number range in the comparison varied from 1 to 9, and each race track consisted of 40 steps. The item selection algorithm of the game tried to keep the probability of success above 75%.

NR also gave immediate, continuous and delayed feedback. Every time the child managed to choose the bigger quantity, the child heard the sound of applause; conversely, if the child chose the smaller quantity, the child heard a short sound signalling an incorrect response. Every time the child won a single track, the child could unlock a fish (underwater) or a butterfly (jungle). If the child won many tracks, the child was allowed to unlock new characters to use for playing.

Cognitive Skills Measures Administered in the First Pre-Test

Visuo-spatial Working Memory. The Corsi blocks task is a widely used test designed to assess visuo-spatial working memory (Corsi, 1972; Milner, 1971). A board (8x10 inch) with wooden cubes (1.25 inch) comparable to the original test was used. The child was asked to touch the cubes in the same serial order according to a given model. The span increased by one after every two sets. If the child gave two consecutive incorrect responses, the testing was discontinued. For each set the child correctly repeated, one point was awarded (for a maximum of 16 points). The sum was used in the analyses. Cronbach's alpha for the Corsi blocks tapping task has been found to be .61 (e.g., Busch, Farrel, Lisdahl-Medina, & Krikorian, 2005).

Phonological Working Memory. The Nonword repetition task from the Neuropsychological tests for Children (NEPSY; Korkman, Kirk, & Kemp, 2008) was used to assess phonological working memory. In this task, the child was asked to repeat non-words, which were orally given by the tester, one at a time. There were 16 items that increased in length and complexity. If the child gave four consecutive incorrect responses, the testing was discontinued. The score was the number of correctly repeated items. The sum score was included in the analyses. Cronbach's alpha in this test has been found to be .71 (Korkman, 2000).

Early Number Skills Measures Administered in Two Pre-Tests and Post-Test

Verbal Counting. Verbal counting skills were measured by three separate verbal counting tasks adapted from the Early Numeracy Test (Van Luit, Van de Rijt, & Aunio, 2006). In the counting forward subtest the child was asked to count forward starting from number 1. For correctly reaching the number words 2−9 one point was awarded. For reaching the number words 10−19 two points were awarded. For reaching 20−23 three points, and reaching 24 four points were awarded. In the counting backward subtest, the child was asked to count backward from 15. The number words for 15, 14 and 13 were given as a model by the tester. If the child was able to count backward correctly until the number words 12−10, the child received one point; reaching 9−6 two points; 5−2, three points; and reaching 1 four points were awarded. In the skip counting subtest, the child was asked to count every second number word starting from 2. The tester provided number words 2, 4, and 6 as a model to begin. If the child was able to continue to the number words 8 or 10 one point was awarded; to 12, 14, 16, or 18 two points; to 20 or 22, three points; and to 24 four points were awarded. A sum score of these three subtests (for a maximum of 12 points) was used in the analyses. Cronbach's alpha was .79 in the first, and .78 in the second pre-test. The Spearman correlation coefficient for test−retest in the sample ($n = 47$) was .73, $p < .001$ (two-tailed).

Object counting. Object counting was assessed by a task in which one to six black randomly arranged dots were presented on a computer screen. The child was asked to say the number of dots aloud as quickly as possible. If the child responded correctly, the tester clicked the mouse's left button. If the child responded incorrectly, the tester clicked the mouse's right button. This test consisted of 4 practice items and 18 test items, with three presentations of one- to six-dot items each. The number of correct responses and reaction times were scored. Because the accuracy of recognizing the dots was over 85% for every task in every assessment point among all participants, the accuracy score was excluded from the analyses. The median of the reaction times for each dot group (1−6) was used for computing two variables. Subitizing fluency (the mean of median reaction times for correctly recognizing dot groups 1−3) was used as a variable according to earlier studies (see Bartelet, Ansari, Vaessen, & Blomert, 2014; subitizing range). Cronbach's alpha was .76 in the first, and .85 in the second pre-test. The other variable used was dot counting fluency (the mean of median reaction times for correctly recognizing dot groups 4−6) based on Bartelet, Ansari, and colleagues (2014; counting range). Cronbach's alpha was .63 in the first, and .60 in the second pre-test. The Spearman correlation coefficient for test–retest in the sample ($n = 47$) was .76, $p < .001$ (two-tailed) in subitizing fluency, and .66, $p < .001$ (two-tailed) in dot counting fluency.

Basic Arithmetic. Basic arithmetic was measured by a paper and pencil test consisting of two parts: 1) concrete object counting (3 tasks) and 2) symbolic calculation parts (28 tasks). The symbolic calculations included tasks like the following: 2 + 1 = __; 4 – 1 = __; 7 + __ = 14; 15 – __ = 9; 3 + 4 + 6 = __; __ – 3 = 10; 16 = 9 + __ (Aunola & Räsänen, 2007). The test began with the symbolic calculations. The child was instructed to resolve as many of the problems as possible in 3 minutes. A stopwatch was used to measure time. If the child could not solve any calculation items, the child was asked to count objects (circles and squares) from three separate pictures and to add the corresponding number symbol next to each picture. The test was originally developed for the longitudinal data collection and thus, the test included multiple arithmetic combinations for avoiding ceiling effect in later primary school grades. The score for basic arithmetic skills was the sum of correct responses. Those who managed to calculate at least one symbolic problem were automatically given three points for object counting. The maximum score was 31 (3 + 28). The Spearman correlation coefficient for test–retest in the sample ($n = 47$) was .82, $p < .001$ (two-tailed).

Control Measure Administered in Two Pre-Tests and Post-Test

Rapid Naming. The test of Rapid serial naming (RAN) of colors (Denckla & Rudel, 1974; standardized Finnish version by Ahonen, Tuovinen, & Leppäsaari, 2006) was included in all three assessments to control for the specificity of the intervention effects. RAN consisted of five colored squares (black, red, yellow, green, and blue) each repeated several times in pseudorandom order, with no consecutive presentations of the same color. Altogether 50 stimuli were arranged in five rows. Before the test, practice items were presented to ensure that the child knew the names of colors. The child was instructed to name all stimuli as quickly and accurately as possible. A stopwatch was used to measure the time for completion, which was used in the analyses. The

Spearman correlation coefficient for test–retest in the sample ($n = 47$) was .83, $p < .001$ (two-tailed).

Procedure

The children were assessed individually at each of the following three time points: February, February–March, and April. Each assessment session was held in a quiet, separate room in the day care center and lasted approximately 20 to 30 minutes.

After two pre-tests, the children were randomly allocated into two intervention conditions, practicing with either GGM or NR. The children were instructed to play individually with their headphones on (without tutoring) for 12 to 15 times in a 3-week period during their kindergarten hours. Each session was instructed to last 10 to 15 minutes. The study aimed for a minimum exposure to practice time at 120 minutes, which was realized in both intervention conditions. The kindergarten teachers organized the intervention sessions, and helped the children to log in, and log out of the intervention games. To assess intervention fidelity, the teachers also reported the number and length of each session in a practice diary.

Data Analyses

The average scores of two pre-tests of each early number skill (verbal counting, subitizing fluency, dot counting fluency and basic arithmetic) and the control (RAN) measure were used as the initial level score. The Corsi blocks task and the Non-word repetition task were measured once in the first pre-test.

The analysis was made using SPSS version 20. Non-parametric methods were used for the analyses because the variables were not normally distributed and the sample sizes were small (*GGM* = 9, *NR* = 8). Therefore, the Wilcoxon signed-rank test was used to analyze within-group intervention effects. The results were interpreted with exact, one-tailed p values. To calculate the within-group effect sizes of the Wilcoxon signed-rank test, the following formula was used: ES (r)

= Wilcoxon $Z / \sqrt{N}$, where N is the number of observations (Field, 2013). The between-group differences in the initial level, the intervention gain scores, and the total exposure to intervention were analyzed by the Mann-Whitney U-test. Here the results were interpreted with exact, two-tailed *p* values. Table 2 presents the intervention group averages at the initial and post-test levels, as well as the significant within-group gain scores.

RESULTS

The effect of intervention for the GGM group on verbal counting was statistically significant, Wilcoxon $Z = -1.95$, $p = .031$, $r = .46$ (Table 2). There was also a significant intervention effect in dot counting fluency, Wilcoxon $Z = -2.19$, $p = .014$, $r = .52$ (Table 2). Altogether 6 children of 9 achieved higher raw scores in verbal counting, and 8 children of 9 were more fluent in dot counting after the intervention. The child with a slower speed in the post-test improved the most in accuracy. Overall, the significant change in dot counting fluency did not result from a lower accuracy level in the post-test. In contrast, the children retained their accuracy level, or were more accurate in the post-test. There was no significant improvement in basic arithmetic, or control task, rapid naming.

For the NR group, a significant intervention effect was seen in basic arithmetic, Wilcoxon $Z = -2.53$, $p = .008$, $r = .63$ (Table 2). Altogether 6 children of 8 achieved higher raw scores in basic arithmetic after NR practice. There was no significant improvement in verbal counting, dot counting fluency, or control task, rapid naming.

Finally, there was a significant between-group difference in intervention gain scores of basic arithmetic ($U = 13.0$, $Z = -2.30$, $p = .014$, $r = .56$), favoring the NR group. There were no other between-group differences in gain scores of early number skills or rapid naming.

The fidelity of intervention was satisfactory in both groups, with all participants reaching the target of 120 practice minutes. The total exposure times ranged from 142 minutes (approx. 9.5 minutes per day) to 237 minutes (approx. 15.8 minutes per day) in the GGM group and from 169 minutes (approx. 11.3 minutes per day) to 350 minutes (approx. 23.3 minutes per day) in the NR group. One child who played 350 NR minutes was an outlier in the sample. After excluding the outlier, the exposure times in the NR group ranged from 169 to 260 minutes (approx. 17.3 minutes per day). However, the outlier was included in non-parametrical analyses based on rank-orders instead of mean scores.

When comparing the exposure to intervention between the GGM and NR groups, there was no significant difference in the number of sessions practiced (GGM: $M = 10.78$, $SD = 2.05$, $MD = 11.00$; NR: $M = 11.25$, $SD = 1.83$, $MD = 11.00$; Mann-Whitney $U = 30.5$, $Z = -0.55$, $p = .619$). However, the difference in total playing time in minutes reached significance ($U = 15.5$, $Z = -1.97$, $p = .049$), indicating longer playing times for the NR group (GGM: $M = 188.00$, $SD = 26.33$, $MD = 190.00$; NR: $M = 232.14$, $SD = 56.31$, $MD = 220.50$). This difference might be due to instructions for playing NR: the children were instructed to end their session only after finalizing an uncompleted race track (from the start point to finish). This instruction was given to ensure that their progress would be recorded per each session. Because of differences in exposure times, and because of the different numerical processes built-in to the two games, the association between playing times and intervention gain scores was analyzed within sub-groups. The results indicated a non-significant correlation (Spearman’s rho) between GGM minutes played and gain scores of verbal counting (.29) and between GGM minutes played and gain scores in dot counting fluency (.08). There also was no significant correlation between sessions played and the aforementioned gain scores (.49; .03, respectively). In the NR group, there was no significant correlation between NR minutes played and gain scores in basic arithmetic (.55), or between sessions played and gain scores in basic arithmetic (.10).

Insert Table 2 here.

DISCUSSION

The main purpose of this study was to examine if short and intensive practice with mathematical computer programs can support early number skills in kindergarteners most at-risk for MD. The results indicated a significant intervention effect for verbal counting and dot counting fluency when the children practiced with GraphoGame Math (GGM), and in basic arithmetic when they practiced with Number Race (NR). The effect sizes were relatively large for all improvements at group level (r = .46; .52; .63, respectively; c.f., Cohen, 1992). Between-group difference was found in gain scores of basic arithmetic, favoring the NR group.

It is unlikely that the practice effects found in both intervention groups were due to maturation, kindergarten teaching, or test–retest effect because the intervention period was short, the found intervention effects were group specific, and the test–retest effect was controlled for with two pre-tests. Moreover, if the observed improvements were due to domain general factors, parallel gains could be expected in the control measure since all assessment tasks shared the fluency requirement. However, there were no within-group effects in the control measure (rapid naming).

The improvements in verbal counting and dot counting fluency in the GGM group can be explained with the nature of the GGM practice itself: it focused on exact discriminations of numerical representations. Also the time limit in each trial encouraged fluency. In addition to counting, GGM practice might have strengthened the concept of cardinality, the relations between numbers (number neighbors, number comparison), and the ability to detect quickly the sub-groups of objects. For example, in each dot counting trial, the child needed to count different quantities in order to pick out the correct stimulus among incorrect ones within a limited time. Therefore, GGM could have directed the children towards using faster and more efficient strategies in determining the number of objects. This would mean seeing a set of five dots as a combination of three-and-two instead of counting the dots one by one. This would reflect conceptual subitizing (Sarama & Clements, 2009).

The results are encouraging since verbal counting is shown to have a strong connection to later arithmetic at school age (e.g., Desoete & Grégoire, 2006; Koponen et al., 2013; Lepola et al., 2005). There is also evidence that fluency in object counting is rather stable between different fluency-level groups (e.g., Reeve et al., 2012). In previous CAI studies, positive effects on object counting accuracy in low-performing children have been reported (Ortega-Tudela & Gómez-Ariza, 2006; Table 1). At-risk children also have benefited from computerized practice (Elliot & Hall, 1997; Table 1). The latter gain was seen on a larger achievement test (TEMA-2) containing object counting as one sub-skill. Hence, it is difficult to conclude whether or not the gain in Elliot's and Hall's study (1997) resulted specifically from an improvement in object counting, or if it simply reflected general improvements in all sub-skills.

Also, a significant effect of intervention was found in basic arithmetic in the NR group. In earlier NR studies, positive effects on arithmetic skills have been found in school-aged children with specific MD status (Wilson, Revkin, et al., 2006, the original version of NR) and without it (see Obersteiner, Reiss, & Ufer, 2013; two different experimental versions of NR were used). The effect is logical considering the content of the game. In NR, the quantities are first presented as concrete objects and number symbols, but quite soon also as basic addition and/or subtraction calculations (ongoing adaptation in notation, numerical distance, and time pressure). As an example, after the child has chosen a calculation like 3 + 2, NR repeats aloud "you chose - three plus two - equals five" while simultaneously presenting all symbols "3 + 2 = 5". This might help children to learn the association between verbal (spoken) and written (numbers and arithmetical symbols i.e. plus, minus, equal) representations of basic addition and/or subtraction calculations. This finding is encouraging since early number combination and story problem solving skills seem to predict later calculation procedures and applied problem solving (e.g., Jordan et al., 2009), and the effect size was relatively large. In earlier CAI studies (other than NR), positive intervention effects in basic addition have been found in groups of at-risk children with performance below the 25th percentile (e.g., Baroody et al., 2012, 2013; Fuchs et al., 2006).

Furthermore, the between-group comparison revealed a significant difference in gain scores of basic arithmetic, favoring the NR group. This result could be due to structure of the games. In NR, basic addition and/or subtraction calculations are presented quite soon after concrete objects and number symbols since the numerical notation is adaptive for accuracy. Therefore, the content varies more continuously in NR than in GGM, which is divided into different levels. This type of variation might mean that the children are exposed to arithmetic practice regardless of the number of NR minutes or sessions played. Such an ongoing sensitivity in adaptation of NR could explain the between-group difference in gain scores on basic arithmetic. Indeed, NR focuses on the approximate numerical processes for determining which of the two arithmetic calculations should be selected for receiving a larger amount of objects; however, the children might have needed to estimate in more detail (or even calculate) the sums and differences of the presented calculations. In GGM, by contrast, the numerical content was organized so that basic arithmetic was hierarchically the highest sub-skill practiced; thus, it is possible that the children did not reach the highest training level during the intervention period. GGM had an adaptation that kept the child practicing specific sub-skills until the satisfied performance level was achieved. Additionally, arithmetic practices were perhaps too complex in the current GGM version, even if it did expose the certain basic concepts. In GGM, the child heard a sum, and the correct calculation must be selected among a number of alternatives. In GGM’s arithmetic tasks, the operation symbols (plus, minus) were visually presented, but the symbols were not verbally presented at all, unlike in NR. Therefore, both treatments focused merely on procedurally oriented addition and subtraction training. It is probable that such practice should come only after the conceptually oriented training in this target group. As suggested, a good conceptual knowledge allows an efficient application of calculation procedures (Dowker, 2009).

We also examined if the intervention benefit was related to the intervention exposure. The results revealed that the gain scores of verbal counting and dot counting fluency did not correlate significantly with the amount of GGM sessions or minutes played. Despite the significant improvement in basic arithmetic in the NR group, the gain score was not related to total NR sessions or minutes played either. This finding could most likely be explained by the adaptation, which individualizes the practice in both games. The success rate is approximately 85 % in GGM (after achieving a certain amount of correct answers, the child is allowed to move to the next game level); and in NR, the content varies frequently depending on the child's performance. For this reason, there may be variation in children's exposure times for different sub-skill training. Perhaps this variation explains why significant correlations were not found between minutes played, sessions practiced, and certain intervention gains.

In sum, the results of this study are in line with some earlier studies in which CAI has been shown to be effective especially for children with weak skills (e.g., Li & Ma, 2010) over short, intensive practice periods (e.g., Kroesbergen & Van Luit, 2003; Kulik & Kulik, 1991). There are also suggestions that a well-planned adaptive practice is able to identify children's strengths and weaknesses as well as fill their individual gaps (Fuchs, 2005; Fuchs, Fuchs, & Compton, 2012), also when offered in computerized format (Hasselbring, 1986; Slavin & Lake, 2008). Hence, it seems reasonable to offer specific number skill training for kindergarteners most at-risk for MD, especially because it is known that early difficulties tend to be very persistent within this group (e.g., Geary et al., 2008; Morgan et al., 2009; Murphy et al., 2007). Furthermore, it might be worth noting that individually targeted practice with computers allows teachers to concentrate on other methods to enhance children's learning (Clements, 2002), and to enrich their experience.

The specific features of the used games might explain different types of effects observed. GGM focused on exact numerical processes and cardinality, and had an effect on verbal counting and dot counting fluency. On the other hand, NR focused on approximate numerical processes, which means that neither the cardinality nor exact dot counting was practiced explicitly. After each quantity selection, the game moved characters along a track while the number words were simultaneously repeated which supported directly verbal counting (counting forward) (see the images in Wilson et al., 2009, p. 227). However, the number range was perhaps too concise (1−9) in NR to produce effects for the used verbal counting measures (range 1−24). As already discussed, both assessed and trained arithmetic tasks having a procedural orientation (instead of a conceptual one) might have had an effect on the effect sizes observed in a group of most at-risk for MD.

As mentioned, the adaptation in both games increased the variation in practiced content. Children might have been exposed somewhat differently to specific sub-skill training within a short intervention period of three weeks. It is also possible that our assessment tools were not sensitive enough to pick up development for all skills assessed. As noted earlier, the number comparison task used emphasized speeded responding and was unfit for the most at-risk participants of the current study. There was also a floor effect in basic arithmetic task due to its original purpose in a longitudinal data collection for avoiding a ceiling effect in later primary school grades. It is obviously not straightforward to create assessment materials responsive enough for the whole range of early number skill levels. On the other hand, standardized assessment tools usually are not specific enough for the targeted training contents. As such, this study should be considered a preliminary approach for assessing the intervention effects in children most at-risk for MD.

There are also other limitations in the study. The small sample sizes, like we had in this study, might create a lack of power for revealing less robust effects of the interventions. Our inclusion criterion was stricter than typically used: the children with a performance below the 10th percentiles in verbal counting were included. Obviously, this limited scope means that the results should be interpreted with caution and await replications with larger samples. In further studies, an experimental design with a business-as-usual control group would also be useful in determining the specific effects of the intervention.

In a number of studies it has been pointed out that within small samples the effect sizes tend to be relatively large (e.g., Cheung & Slavin, 2013; Slavin & Lake, 2008). For this reason, instead of effect size values, different benchmarks for interpreting the effectiveness (relevant to intervention, target population, and outcome measures) should be used (Grissom & Kim, 2012; Hill, Bloom, Black, & Lipsey, 2008). In our case, the effect size comparisons were difficult to carry out since the inclusion criterion was stricter, and intervention duration and intensity differed in the current study as compared to earlier CAI studies (see Table 1). Although the effect sizes exceeded the long-time averages presented in meta-analyses (see Räsänen, in press), they should be evaluated with caution. In addition, standardized tests have been suggested for identifying target children (Mazzocco, 2005) and for measuring outcomes (Slavin & Lake, 2008) when studying the real transfer benefit of interventions, or for proposing any method as an evidence-based practice (e.g., Cook, Tankersley, & Landrum, 2009). Nonetheless, in this study, the practice targeted the specific skills of a particular group of children in need of early support. In other words, the purpose of this study was not to evaluate the transfer effects or to compare the effects to the normally-developing reference group that already performs at the ceiling in many prerequisite early number tasks. Finally, it would be useful to have an access to game log data, and to conduct long-term post-assessments.

Some practical implications in terms of developing educational games are worth discussing. Even though it seems that a short and intensive computerized practice can produce condition specific effects with regard to the specific group of children most at-risk for MD, developers should carefully focus on coherent intervention principles. The practice should include explicit instructions; step-by-step procedures; simultaneous training for both concepts and concrete operations; immediate, continuous, and delayed feedback; a motivating environment; and ongoing assessment (cf. Baker, Gersten & Lee, 2002; Fuchs et al., 2008; Gersten et al., 2009). Both intervention programs used in the current study (GraphoGame Math and Number Race) cover the majority of the aforementioned principles, but there is room for improvement. This means not only a sufficient and rich numerical content with multidimensional task types, but also a carefully planned MD-appropriate user interface; content-based adaptations; and a more pedagogical feedback system. Due to a persistent nature of deficits in arithmetic the early intervention should strengthen both incipient number skills and basic addition/subtraction in a meaningful way. The conceptual basis should be practiced before more procedurally oriented training starts. All of these aforementioned components could have specific influences on the desired immediate and long-term effects. To develop such an appropriate and sensitive tool, multidisciplinary efforts are needed, including mathematics education and psychology researchers, game developers, and big log-data statisticians.

As a recommendation for further studies, the effectiveness of CAI in children most at-risk for MD should be examined with larger samples. As studies have shown individual variation in intervention responsiveness to be large in a group of children with MD (e.g., Dowker & Sigley, 2010; Fuchs et al., 2012; Geary, 2011), we would recommend examining the effects of a tailored, targeted training based on qualified screening assessments. The potential intervention benefits also should be followed by delayed assessments (e.g., Fuchs et al., 2006; Wilson et al. 2009). In addition, game log data could provide more detailed information on an individual level, as well as generate insight into how children actually act while using the programs (see Käser et al., 2011 for more on using log file data in analyses of CAI effectiveness). This type of data could help determine the individual patterns of development within the CAI and give a deeper understanding of the vital factors in producing better learning.

References

Ahonen, T., Tuovinen, S., & Leppäsaari, T (2006). *Nopean sarjallisen nimeämisen testi* (2nd ed.). Jyväskylä, Finland: Niilo Mäki Institute.

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20*(5), 427−435. doi: 10.1016/j.lindif.2010.06.003

Aunola, K., Leskinen, E., Lerkkanen, M-K., & Nurmi, J-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*(4), 699–713. doi: 10.1037/0022-0663.96.4.699

Aunola, K., & Räsänen, P. (2007). *The 3-minutes Basic Arithmetic Test.* Niilo Mäki Institute. Unpublished.

Baker, S., Gersten, R., & Lee, D-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103*(1), 51−73. Retrieved from http://www.jstor.org/stable/1002308

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education, 18*(2), 141−157. http://www.jstor.org/stable/749248

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering at-risk kindergarten children's number sense. *Cognition and Instruction, 30*(4), 437–470. doi: 10.1080/07370008.2012.720152

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal, 1*, 1−41. doi: 10.3102/0002831212473349

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities 35*(3), 657−670. doi: 10.1016/j.ridd.2013.12.010

Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D., (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *Journal of Experimental Child Psychology 117*, 12–28. doi: 10.1016/j.jecp.2013.08.010

Becker, H. J. (1992). Computer-based integrated learning systems in the elementary and middle grades: A critical review and synthesis of evaluation reports. *Journal of Educational Computing Research, 8*(1), 1−41. doi: 10.2190/23BC-ME1W-V37U-5TMJ

Busch, R. M., Farrell, K., Lisdahl-Medina, K., & Krikorian, R. (2005). Corsi block-tapping task performance as a function of path configuration. *Journal of Clinical and Experimental Neuropsychology, 27*(1), 127−134. doi: 10.1080/138033990513681

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science, 332*, 1049−1053. doi: 10.1126/science.1201536

Chambers, J. A., & Sprecher, J. W. (1980). Computer Assisted Instruction: Current Trends and Critical Issues. *Communications of the Association for Computing Machinery, 23*(6), 332−342. doi: 10.1145/358876.358881

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

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology, 33*(3), 300–317. doi: 10.1080/87565640801982387

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, 1*, 91–104. Retrieved from http://www.editlib.org/p/18866

Clements, D. H. (2002). Computers in early childhood mathematics. *Contemporary Issues in Early Childhood, 3*(2), 160−180. doi: org/10.2304/ciec.2002.3.2.2

Clements, D. H. (2004). Major themes and recommendations. In D. H. Clements, & J. Sarama (Eds.), *Engaging young children in mathematics, standards for early childhood mathematics education* (pp. 7–72). Mahwah, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 122*(1), 155−159. doi: 10.1037/0033-2909.112.1.155

Cook, B. G., Tankersley, M., & Landrum, T. J. (2009). Determining evidence-based practices in special education. *Exceptional Children, 75*(3), 365–383. Retrieved from http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&sid=75a11115-bab9-4bfd-b9f4-18b70eb30267%40sessionmgr115&hid=118

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*(2).

Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* Oxford, UK: University Press.

Dencla, M. B., & Rudel, R. (1974). Rapid “automatized” naming of pictured objects, colours, letters and numbers by normal children. *Cortex, 10*(2), 186−202.

De Smedt, B., Noël, M-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children’s mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*, 48–55. doi: 10.1016/j.tine.2013.06.001

Desoete, A., & Grégoire, J. (2006). Numerical competence in young children and in children with mathematics learning disabilities. *Learning and Individual Differences, 16*, 351–367. doi: 10.1016/j.lindif.2006.12.006

Din, F. S., & Calao, J. (2001). The effects of playing educational video games on kindergarten achievement. *Child Study Journal, 31*(2), 95–102. http://clem.mscd.edu/~sandersc/3310%20ed%20video%20article.pdf

Dowker, A. (2009). Use of derived fact strategies by children with mathematical difficulties. *Cognitive Development, 24*(4), 401−410. doi: 10.1016/j.cogdev.2009.09.005

Dowker, A., & Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *The British Psychological Society, Monograph Series 2*(7), 65–81. doi:10.1348/97818543370009X12583699332492

Elliot, A., & Hall, N. (1997). The impact of self-regulatory teaching strategies on ”at-risk” preschoolers’ mathematical learning in a computer-mediated environment. *Journal of Computing in Childhood Education, 8*(2/3), 187–198.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE Publications Ltd.

Finnish Ministry of Education and Culture. (2013). Selvitys esiopetuksen velvoittavuudesta [Report of obligation to participate in Finnish kindergarten curriculum], Finland: Ministry of Education and Culture.

Finnish National Board of Education. (2010). Core curriculum for preschool education in Finland. Downloadable (in English) from http://www.oph.fi/download/153504_national_core_curriculum_for_pre-primary_education_2010.pdf

Fuchs, L. S. (2005). Prevention research in mathematics: Improving outcomes, building identification models, and understanding disability. *Journal of Learning Disabilities, 38*(4), 350–352. doi: 10.1177/00222194050380041201

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*(3), 257–269. doi: 10.1177/0022219412442167

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., & Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities, 39*(5), 467–475. doi: 10.1177/00222194060390050701

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31*(2), 79–92. doi: 10.2307/20528819

Fuson, K. C. (1988). *Children's counting and concepts on number.* New York: Springer-Verlag.

Fuson, K. C. (2009). Avoiding misinterpretations of Piaget and Vygotsky: Mathematical teaching without learning, learning without teaching, or helpful learning-path teaching? *Cognitive Development, 24*, 343–361. doi: 10.1016/j.cogdev.2009.09.009

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*(1), 4–15. doi: 10.1177/00222194040370010201

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics, 32*(3), 250–263. doi:10.1097/DBP.0b013e318209edef

Geary, D. C. (2013). Learning disabilities in mathematics: Recent advances. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (second edition) (pp. 239–255). New York: Guilford Press.

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities, 45*(4), 291–307. doi: 10.1177/0022219410392046

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*(3), 277–299. doi: 10.1080/87565640801982361

Gelman, R., & Gallistel, C. R. (1978). *The child’s understanding of number.* Cambridge, MA: Harvard University Press.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. R. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202–1242. doi: 10.3102/0034654309334431

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*, 293–304. doi: 10.1177/00222194050380040301

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

Hasselbring, T. S. (1986). Research on the effectiveness of computer-based instruction: a review. *International Review of Education, 32*(3), 313–324. doi: 10.1007/BF02426065

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177. doi: 10.1111/j.1750-8606.2008.00061.x

Howard, J. R., Watson, J. A., Brinkley, V. M., & Ingels-Young, G. (1994). Comprehension monitoring, stylistic differences, pre-math knowledge, and transfer: A comprehensive pre-math/spatial development computer-assisted instruction (CAI) and LOGO curriculum designed to test their effects. *Journal of Educational Computing Research, 11*(2), 91–105. doi: 10.2190/8WK8-Q6T1-2V0R-K2RW

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*(1), 36–46. doi: 10.1111/j.1540-5826.2007.00229.x

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

Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J-E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skills. *Journal of Experimental Child Psychology, 97*, 220–241. doi: 10.1016/j.jecp.2007.03.001

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*(1), 162–175. doi: 10.1037/a0029285

Korkman, M. (2000). *Nepsy. Käsikirja II.* Helsinki, Finland: Psykologien Kustannus Oy.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY. A developmental neuropsychological assessment manual*, Finnish Edition: NEPSY II - Lasten neuropsykologinen tutkimus (Korkman, Kirk, & Kemp, 2008). Helsinki, Finland: Psykologien Kustannus Oy

Kroesbergen, E. H., & Van Luit, J. (2003). Mathematics interventions for children with special educational needs. A meta-analysis. *Remedial and Special Education, 24*(2), 97–114. doi: 10.1177/07419325030240020501

Kulik, C-L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*, 75–94. doi: 10.1016/0747-5632(91)90030-5

Käser, T., Kucian, K., Ringwald, M., Baschera, G. M., von Aster, M., & Gross, M. (2011). Therapy software for enhancing numerical cognition. In J. Özyurt, A. Anschütz, S. Bernholt, & J. Lenk (Eds.), *Interdisciplinary perspectives on cognition, education and the brain* (pp. 207−216). Oldenburg, BIS-Verlag. (Hanse-Studies Vol. 7)

Lepola, J., Niemi, P., Kuikka, M., & Hannula, M. M. (2005). Cognitive-linguistic skills and motivation as longitudinal predictors of reading and arithmetic achievement: A follow-up study from kindergarten to grade 2. *International Journal of Educational Research, 43*, 250–271. doi: 10.1016/j.ijer.2006.06.005

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. doi: 10.1007/s10648-010-9125-8

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*(5), 451–459. doi: 10.1177/0022219408321126

Mazzocco, M. M. M. (2005). Challenges in identifying target skills for math disability screening and intervention. *Journal of Learning Disabilities, 38*(4), 318−323. doi: 10.1177/00222194050380040701

Mazzocco, M. M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice, 20*(3), 142–155. doi: 10.1111/j.1540-5826.2005.00129.x

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471–490. doi: 10.1016/j.ecresq.2006.09.003

McCollister, T. S., Burts, D. C., Wright, V. L., & Hildreth, G. J. (1986). Effects of computer-assisted instruction and teacher-assisted instruction on arithmetic task achievement scores of kindergarten children. *Journal of Educational Research, 80*(2), 121–125.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*(3), 272−277.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*(4), 306−321. doi: 10.1177/0022219408331037

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*(5), 458−478. doi: 10.1177/00222194070400050901

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction, 23*, 125−135. doi: 10.1016/j.learninstruc.2012.08.004

Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2006). Computer-assisted teaching and mathematical learning in Down syndrome children. *Journal of Computer Assisted Learning, 22*, 298−307. doi: 10.1111/j.1365-2729.2006.00179.x

Piazza, M., Mechelli,A., Butterworth, B., & Price. C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *NeuroImage, 15*, 435−446. doi:10.1006/nimg.2001.0980

Polet, J., & Koponen, T. (2011). LukiMat - Oppimisen arviointi: Matematiikan tuen tarpeen tunnistamisen välineet esikouluun. Tekninen opas. Downloadable from http://www.lukimat.fi

Praet, M., & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerized kindergarten interventions: A randomized controlled study. *Teaching and Teacher Education 39*, 56−65. doi: 10.1016/j.tate.2013.12.003

Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology, 4*(141), 649–666. doi: 10.1037/a0027520

Richardson, U., & Lyytinen, H. (2014). The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*

10(1), 39–60. Downloadable from http://www.humantechnology.jyu.fi/articles/volume10/2014/Richardson_Lyytinen.pdf

Räsänen, P. (in press). Computer-assisted interventions on basic number skills. In R. Cohen Kadosh, & A. Dowker (Eds.), *Oxford handbook of numerical cognition*. London: Oxford University Press.

Räsänen, P., Salminen, J. Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development, 24*, 450−472. doi: 10.1016/j.cogdev.2009.09.003

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children.* New York, NY: Routledge.

Seo, Y-J., & Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education, 53*, 913–928. doi:10.1016/j.compedu.2009.05.002

Skagerlund, K., & Träff, U. (2014). Number processing and heterogeneity of developmental dyscalculia: Subtypes with different cognitive profiles and deficits. *Journal of Learning Disabilities.* doi: 10.1177/0022219414522707

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best evidence synthesis. *Review of Educational Research, 78*(3), 427−455. doi: 10.3102/0034654308317473

Starkey, P., & Cooper, R. G. Jr, (1980). Perception of numbers by human infants. *Science, 210*(4473), 1033–1035. doi: 10.1126/science.7434014

Van Luit, J. E. H., Van de Rijt, B. A. M., & Aunio, P. (2003). Early Numeracy test, Finnish Edition: Lukukäsitetesti (Van Luit, Van de Rijt, & Aunio, 2006). Helsinki, Finland: Psykologien Kustannus Oy.

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education, 3*(4), 224–234. doi: 10.1111/j.1751-228X.2009.01075.x

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006). Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions, 2*(19), 1–14. doi: 10.1186/1744-9081-2-19

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of The Number Race, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions, 2*(20), 1–36. doi: 10.1186/1744-9081-2-20

Wright, R. J. (2003). Mathematics recovery: A program of intervention in early number learning. *Australian Journal of Learning Disabilities, 8*(4), 6–11. doi: 10.1080/19404150309546743

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*, 155−193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology, 24*, 220−251.

Table 1

Descriptions of Computer-Assisted Intervention Studies in At-Risk Children

Study	Sample	Age	Status	Time	Sessions	Main result (ES / ES_m)[k]
Baroody et al. (2012)	15+13	5.58	at-risk MD[b]	19	20[h] + 20 x 30 min	Addition ($ES_m = .30$)
Baroody et al. (2013)	43+21	6.5	at-risk MD[b]	20	20[h] + 20 x 30 min	Addition (n+1/1+n) ($ES = .39$)
Din et al. (2001)[a]	24+23	5.5	low SES	11	55 x 40 + 30 min[i]	No significant training effect
Elliot et al. (1997)[a]	18+18	4	at-risk LD[c]	6	15 x 20 min	TEMA-2 ($ES = .54$)
Fuchs et al. (2006)	16+17	6−7	at-risk MD[d]+RD[e]	18	50 x 10 min	Addition ($ES = .44$)
Howard et al. (1994)	32	4.42	middle SES	9	27 x 15 min	TEMA (ES[l])
McCollister et al. (1986)[a]	28+25	5−6	low / middle SES	< 2	10 x 10−15 min	Numeral recognition ($ES = .43$)
Ortega-Tudela et al. (2006)	10+8	6.55	low performing[f]	21	15 x 35 min	Enumeration ($ES_m = .61$)
Praet et al. (2014)	44+39+49	5.67	middle SES	5	8 x 25 min	Arithmetic ($ES_m = .39$)[m]
Räsänen et al. (2009)	29+15+15	6.55	low performing[g]	3	15 x10−15 min	Symbolic comparison ($ES = .22$)[n]
Wilson et al. (2009)	27+26	5.6	low SES	14	6 + 4 x 20 min[j]	Symbolic comparison ($ES_m = .39$)[o]

Note. Sample = Sample sizes (treatment + control) used for analyses. Age = Mean age in years (*SD*s were not always reported). MD = mathematics difficulties. SES = socioeconomic status. LD = learning difficulties. RD = reading difficulties. Time = Time used for practice in weeks. Main result = Main significant training effects. $ES_{(m)}$ = Effect size$_{(\text{mean})}$. TEMA = Test of Early Mathematics Ability (Ginsburg & Baroody, 1983, 1990, 2003).

[a] Training in pairs. [b] In the bottom 25th percentile in TEMA-3 (and/or personal and/or familial risk factors). [c] Identified as being at-risk of early learning difficulties. [d] In the bottom 25th percentile in arithmetic and [e] in the bottom 17th percentile in reading. [f] Not acquired counting and cardinality concepts according to a curriculum-based evaluation. [g] Identified by kindergarten teachers as having a need for numerical support. [h] 20 sessions with manual games, 20 with CAI. [i] 40 min at kindergarten, 30 min at home. [j] Six sessions with math, four sessions with reading software. [k] If not reported, effect sizes (Cohen's d) were calculated from reported statistics by recommended formulas (Cohen, 1992), and thereafter all d values were converted to r values for being comparable to the current study results. [l] Could not be calculated. [m] Sample sizes of low performers (<25th percentile) in three conditions were not reported. Between-groups *ES*s were calculated for the main sample. [n] Counted as an *ES* for both intervention conditions. [o] Counted from quadratic analyses for the lowest performers.

Table 2

Group Means, Standard Deviations, Medians, and Significant Within-Group Gain Scores, as well as Between-Group Comparisons

	Intervention condition					
		GGM ($n = 9$)		NR ($n = 8$)		Group comparisons
Variable	Time point	*M (SD)*	*MD*	*M (SD)*	*MD*	*Mann-Whitney U*
Cognitive skills						
Corsi blocks (max. 16)	T1	4.67 (1.00)	5.00	5.25 (1.04)	5.00	
Non-words (max. 16)	T1	7.56 (2.70)	9.00	8.25 (1.49)	8.00	
Early number skills						
Verbal counting (max. 12)	T1	**4.17 (1.94)**	4.00	5.81 (1.62)	6.25	
	T2	**6.11[b] (3.48)***	7.00	7.50 (3.51)	7.50	
Subitizing fluency (seconds)	T1	1.42 (0.18)	1.38	1.56 (0.27)	1.63	
	T2	1.31 (0.39)	1.23	1.43 (0.36)	1.45	
Dot counting fluency (seconds)	T1	**3.82 (0.44)**	**3.73**	3.88 (0.96)	3.59	
	T2	**3.11[c] (0.61)***	**3.24**	3.62 (0.82)	3.58	
	gain	-0.71 (0.71)	-0.96	-0.26 (0.75)	-0.37	
Basic arithmetic (max. 31)[a]	T1	3.67 (1.32)	4.00	**4.94 (1.32)**	**4.25**	
	T2	3.89 (1.36)	4.00	**5.75[d] (1.17)***	**5.00**	NR > GGM**
	gain	0.22 (0.97)	0.00	0.81 (0.37)	1.00	NR > GGM*
Control task						
Rapid naming (seconds)	T1	72.39 (12.11)	69.50	67.63 (17.44)	65.50	
	T2	64.89 (14.13)	58.00	65.63 (12.98)	65.50	

Note. GGM = Graphogame math group; NR = Number race group; T1 = initial level; T2 = post-test.
Within-group effects of intervention are shown in boldface.
[a]Time limited task. [b]Wilcoxon *ES* (*r*) = .46; [c].52; [d].63.
* $p < .05$. ** $p < .01$.

II

INDIVIDUAL VARIANCE IN RESPONSIVENESS TO EARLY COMPUTERIZED MATHEMATICS INTERVENTION

by

Jonna Salminen, Tuire Koponen, Markku Leskinen,
Anna-Maija Poikkeus, & Mikko Aro, (2015)

Learning and Individual Differences

Reproduced with kind permission by Elsevier.

Learning and Individual Differences 43 (2015) 124–131

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Individual variance in responsiveness to early computerized mathematics intervention

Jonna B. Salminen [a,*], Tuire K. Koponen [b], Markku Leskinen [a], Anna-Maija Poikkeus [c], Mikko T. Aro [a]

[a] *Department of Education, Special Education, P.O. Box 35, FI-40014, University of Jyvaskyla, Finland*
[b] *Niilo Mäki Institute, Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland*
[c] *University of Jyvaskyla, Department of Teacher Education, P.O. Box 35, FI-40014, University of Jyvaskyla, Finland*

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form 16 February 2015
Accepted 11 September 2015

Keywords:
Computer-assisted intervention (CAI)
Kindergarten
Early number skills
Game-log data
Response to intervention (RTI)

ABSTRACT

We examined the effects of short, intensive computerized intervention in early number skills for kindergarteners with poor addition skills (below 1.5 *SD*). The mathematical content of the software was hierarchically organized, starting from one-to-one correspondence, comparing and ordering, and proceeding via number concept and counting to basic addition. The results showed positive within-group effects for basic addition (Wilcoxon *ES* (r) = .59), verbal counting (.56), and the Number Sets Test (.45; see Geary, Bailey, & Hoard, 2009). The effects remained stable over a 9-week follow-up period. However, there was no significant between-group difference in terms of gain scores as compared to a wait-list control group. Based on game-log data, individual variance in responsiveness to the intervention was analyzed. Even though the findings suggest that adaptive, hierarchically organized content could provide effective support for some children with poor early number skills, more specific instruction and feedback system are needed in individualizing interventions.

1. Introduction

Basic arithmetic (i.e., addition and subtraction skills) is an important predictor for later school mathematics achievement above and beyond the influence of intelligence (Duncan et al., 2007; Geary, 2011a; Geary, Hoard, & Bailey, 2012). Difficulties in basic arithmetic and fact retrieval are very persistent (Geary, 2011b), and they constitute a core feature of mathematics difficulties (MDs) (Gersten, Jordan, & Flojo, 2005). Accumulated knowledge about the development of early number skills as a basis for arithmetic has helped to advance means of early identification and support (Butterworth, 2005; Morgan, Farkas, & Wu, 2009). However, there is still a need to examine effective, evidence-based intervention methods (Butterworth, Varma, & Laurillard, 2011; Jordan & Levine, 2009), especially among kindergarteners with poor early number skills (performance below 10th percentile; Morgan et al., 2009; Murphy, Mazzocco, Hanich, & Early, 2007).

Early arithmetic skill seems to develop via different hierarchical modules of sub-skills. In this regard, approximate magnitude discrimination is the innate skill of subitizing small sets of objects and quickly differentiating which of the two sets of objects is the larger when the difference between the quantities is significant enough (Geary, 2013; Dehaene, 2011). The ability to recite number words evolves later with the development of expressive language skills, which thus enables meaningful counting (Krajewski & Schneider, 2009). After that, the link between the number words, quantities and symbols becomes precise (Geary, 2013; Krajewski & Schneider, 2009). This developmental step is required for describing the exact amount of quantities exceeding the subitizing range. Furthermore, an explicit number system, including an understanding of the relationships between numbers, is the next vital step for composing and decomposing, and thus, basic arithmetic skills (cf. Geary, 2013; Krajewski & Schneider, 2009). Von Aster (2000) has noted that individual development is also dependent on the maturation of semantic, verbal, and visual/symbolic modules, which are all necessary for calculation skills.

Due to the diversity of the deficits associated with MDs (e.g., Rubinsten & Henik, 2009), as well as the heterogeneity among individuals with MDs (e.g., Geary, 2004; Jordan, Hanich, & Kaplan, 2003; Von Aster & Shalev, 2007), there is a call for tailored interventions (Dowker, 2001; Geary, 2011b; Slavin & Lake, 2008), and continuous evaluation and identification of children who are not responding to support (Fuchs, Fuchs, & Compton, 2012). Therefore, adaptive intervention programs with dynamic, simultaneous assessment tools could be beneficial in identifying children with (or at-risk for) MDs, and for assisting teachers in planning individualized support.

A well-planned computer-assisted intervention (CAI) offers several possibilities for tailored practice (Seo & Bryant, 2009; Slavin & Lake, 2008), even though the main trend in terms of the effectiveness of CAI on number skills has been suggested to be in decline in recent decades

---

* Corresponding author.
E-mail addresses: jonna.salminen@jyu.fi (J.B. Salminen), tuire.koponen@nmi.fi (T.K. Koponen), markku.leskinen@jyu.fi (M. Leskinen), anna-maija.poikkeus@jyu.fi (A.-M. Poikkeus), mikko.aro@jyu.fi (M.T. Aro).

http://dx.doi.org/10.1016/j.lindif.2015.09.002

(Cheung & Slavin, 2013). The factors behind CAI's positive effects on arithmetic development are challenging to identify due to variations among the target group's characteristics, group sizes, practiced numerical content and the instructional components, and the interventions' intensiveness in previous studies (see Table 1). In order to further develop the use of computers in learning, it is important to establish the specific components needed for effective, tailored intervention.

The core components of effective numerical intervention for children with (or at-risk for) MDs include explicit instructions; repetitive training in basic concepts; step-by-step proceeding; early success; immediate corrective, continuous, and cumulative feedback; and a motivating environment with which to maintain task-orientation (Baker, Gersten, & Lee, 2002; Fuchs et al., 2008; Gersten et al., 2009). Basic skills should be addressed before more complex ones (Dowker, 2001; Sarama & Clements, 2009), and sub-skills should be integrated rather than addressed separately (Fuchs et al., 2012). Further, the relationship between non-symbolic and symbolic notations should be emphasized (Griffin, 2004; Van Luit & Schopman, 2000). However, it seems that previous CAIs for arithmetic have mainly consisted of drill-based practice used for automatizing fact retrieval (see Table 1). Despite generally positive results, the stability of improvements stemming from the intervention is rarely reported. Nonetheless, Kucian et al. (2011) and Schoppek and Tulis (2010) have demonstrated the delayed effects of a short-term intervention that utilized a variety of mathematics content (see descriptions in Table 1).

There is no pre-existing evidence concerning the effects of gradually enhancing children's development of early number skills on learning arithmetic. In addition to group-level intervention effects, we wanted to evaluate individual performance during practice. Therefore, the study goals were: 1) to investigate the group-level effects of short and intensive intervention with a GraphoGame Math program (GGM; see description in Section 2.4.) and to assess the stability of any effects; 2) to contrast the intervention participants' gain scores with the performance level of a wait-list control group; and 3) to evaluate the responsiveness to GGM intervention at an individual level by analyzing the game-log data.

2. Method

Kindergarten teachers from 24 different day care centers in southern Finland were each asked to nominate two children from their group as candidates most in need of extra mathematics intervention. With their parents' permission, the children ($n = 48$) participated in assessments and computer-assisted intervention. All *48* children were individually tested to determine their early number skills. For ethical reasons, all the children were included in the intervention, even though the assessments indicated that the group of candidates included false positives. Participants were randomized into either 1) a GGM group ($n = 24$) that took part in GraphoGame Math practice during the first intervention period and had no practice during the second intervention period; or 2) a control group ($n = 24$) that had no extra practice during the first intervention period and participated in another numerical practice during the second intervention period. All participants were native speakers of Finnish.

2.1. Study participants

Of the *48* participants, *21* children fulfilled the criteria for poor addition skills (1.5 *SD* below the normative age level; below 7th percentiles) and were thus included in the analyses. The inclusion criterion was being unable to solve more than three simple addition problems (e.g., $2 + 1$, $1 + 3$, $3 + 2$) in a time-limited task, as compared to the age-level mean score of 18 out of 45. The reference data had been collected by research assistants for another study one month prior to when the data collection for the current study was carried out (*reference sample* $n = 77$; *mean age* = 74.2 months, $SD = 3.6$). A sub-sample of *13* children (*4* boys and *9* girls; each from different day care centers) formed the GGM group (*mean age* = 78.6 months, $SD = 5.4$). One participant was not present for the third assessment due to illness. The missing data point was replaced by adding the mean gain score of the GGM group (the difference between the third and second assessments) to the child's score in the second assessment. This did not have an effect on the statistical significance of the findings. A sub-sample of *8*

Table 1
Descriptions of the trained contents in computer-assisted intervention studies of basic arithmetic.

Study	*n*	Age	Status	Duration	Training sessions	Description of the training	Effect
Baroody et al. (2012)[a]	28	5.58	at-risk[d]	19 weeks	20 + 20 × 30 min	Subitizing, enumeration, numeral recognition, transcoding and addition + addition (add-0/1)	+
Baroody et al. (2013)[a]	64	6.5	at-risk[d]	20 weeks	20 + 20 × 30 min	Transcoding, verbal counting, object counting, numerical relations, written numbers, arithmetic + addition (add-1/near doubles)	+
Christensen and Gerber (1990)	60	8.80	LD[e]	2 weeks	13 × 6 min	Drilling of single-digit addition facts	+
Fuchs et al. (2006)	33	6–7	at-risk MD[d] + RD	18 weeks	50 × 10 min	Retrieving addition and subtraction facts	+
Hativa and Shorer (1989)	211	8–11	low SES	Semester	3 times a week	Mixed types of arithmetical contents	−
Kraus (1981)[b]	19	7–8	TA	2 weeks	5 + 5 × 1 min	Filling the missing addends to addition combinations	+
Kucian et al. (2011)[b]	32	9.5	DD	13 weeks	25 × 15 min	Locating numbers of dots, digits, sums and differences to number line	++
Käser et al. (2013)[b]	32	7–11	MD	12 weeks	30 × 20 + 30 × 20 min	Number representation, varied types of addition and subtraction tasks	+
Mevarech and Rich (1985)	376	8–11	LA	Semester	Once a week	Mixed types of arithmetical contents	+
Obersteiner et al. (2013)	147	6.91	TA	4 weeks	10 × 30 min	Two versions of Number Race (c.f., Wilson, Dehaene et al. 2006)	+
Okolo (1992)	41	9–12	LD	9 weeks	4 × 20 min + 15 min	Mapping presented responds for addition or multiplication facts	+
Schoppek and Tulis (2010)	110	8.7	TA	10 weeks	7 × 60 min	Solving arithmetical equations and word problems (addition, subtraction, multiplication, division), number comparison, number line	+
Schoppek and Tulis (2010)	94	9.1	TA	10 weeks	7 × 45 min	Described above	++
Shin et al. (2006)	46	7–8	Middle SES	18 weeks	3–4 × 15 min a week	Drilling of addition, subtraction or their mixed combinations	+
Trifiletti, Frith, and Armstrong (1984)	21	9–15	LD[d] + MD	Semester	40 min a day	Mathematics readiness, addition, subtraction, multiplication, division and fraction	+
Wilson, Dehaene, Dubois, and Fayol (2009)[c]	53	5.6	Low SES	14 weeks	6 + 4 × 20 min	Number Race: Approximate comparison between quantities, number symbols and/or addition and subtraction facts	−
Wilson, Revkin, et al. (2006)	9	7–9	LA[f]	10 weeks	20 × 30 min	Number Race: described above	+

Note. Age = Mean age in years (as originally reported). LD = learning difficulties. MD = mathematics difficulties. RD = reading difficulties. SES = socioeconomic status. TA = typically achieving. DD = developmental dyscalculia. LA = low achieving. Effect = immediate (+) and long-term effects (++) on arithmetic.
[a]Training started with manual games. [b]Training operated at homes. [c]Training mixed with reading software. [d]Cut-off point (not always reported) below 25; [e]16; [f]37 percentiles.

children (5 boys and 3 girls; each from different day care centers) formed the control (CTR) group (*mean age* = 76.6 months, *SD* = 4.3).

2.2. Design

The study was carried out in the day care centers for 14 weeks, from November to February, including four non-computerized assessment points and two cycles of intervention. After the first assessment (Time 1; week 1), the participants were randomly divided into two groups (week 2). During weeks 3–5, the GGM group played GraphoGame Math, while the CTR group received no extra intervention. After this 3-week cycle, the second assessment was conducted (Time 2; week 6). Next followed a break (weeks 7–9) due to the holiday season. The third assessment was administered after the break (Time 3; week 10). During weeks 11–13, the GGM group received no extra intervention, while the CTR group was offered another type of numerical intervention. Finally, the fourth assessment (Time 4; week 14) was conducted. Due to the different characters of the numerical intervention for the CTR group, comparisons are only made for the gain scores until the third assessment.

During the study, all participants followed their normal kindergarten curriculum. According to the Finnish National Board of Education (2010; downloadable in English), an understanding of concepts, classification, comparison, and sorting are specified as objectives for early mathematics in pre-primary education (pp. 11–12).

2.3. Early number skill measures

To assess the reliability the measures of early number skills were piloted before the study. The pilot study sample consisted of kindergarten children from southern Finland ($n = 34$). Pearson correlation coefficients for test–retest results (over a week) are reported here separately for each measure.

2.3.1. Enumeration

The enumeration task consisted of a number concept task involving 18 large wooden beads (Salminen, Räsänen, Koponen, & Aunio, 2008a). The child was asked to pick up the number of beads that the tester had requested and then put them on the table. Six different quantities were requested: 3, 6, 8, 10, 13, and 17. The score was the number of correct answers (maximum of 6 points). The Pearson correlation coefficient for test–retest in the pilot study was .72, $p < .001$ (2-tailed).

2.3.2. Verbal counting

Three different verbal counting tasks were adapted from Diagnostic Tests 3 (Salonen et al., 1994). In the first subtest, the child was asked to count forward starting from 1. A correct response containing number words between 2 and 9 was scored as 1 point, 10–19 as 2 points, 20–29 as 3 points, and 30 as 4 points. In the second subtest, the child counted 4 steps forward from a given number word: 3, 8, 12, and 19. In the third subtest, the child was asked to count 4 steps backwards from a given number word: 4, 8, 12, and 23. In the two latter subtests one point was given for each correctly performed number word sequence. A sum score of the three subtests was used in the analyses (maximum of 12 points). The Pearson correlation coefficient for test–retest in the pilot study was .80, $p < .001$ (2-tailed).

2.3.3. Number sets test

The original Number Sets Test (Geary et al., 2009) consists of four different parts. In this paper and pencil test, the child was asked to determine as quickly and accurately as possible whether pairs or trios of object sets and/or Arabic numbers match a standard number (5 and 9). Part A involves objects, Part B objects and numbers. Before starting the test, the target number 4 was used when explaining the task. The target number 3 was administered as practice. In this study, only the target number 5 was used. The child was asked to circle the correct pairs or trios that matched 5. A stopwatch was used to measure the time limit of 1 min per part. The sensitivity for identifying matching pairs or trios (correct responses minus incorrect ones) was scored for both parts, although only Part A was used for the analyses because of a floor effect in Part B. The maximum score was 18 points. The Pearson correlation coefficient for test–retest in the pilot study was .82, $p < .001$ (2-tailed).

2.3.4. Basic addition

The basic addition task included single-digit addition problems ($a + b = \_\_$) in which the sum was 10 or less (Salminen, Räsänen, Koponen, & Aunio, 2008b). The items were pseudo-randomly ordered. Problems in which the sum was five or less were presented first, while highly similar problems were not permitted to follow each other (e.g., $2 + 2 = \_\_$ and $2 + 3 = \_\_$). Add-zero problems were excluded. The task $1 + 1 = \_\_$ was administered as practice, while the tasks $3 + 3 = \_\_$ and $1 + 5 = \_\_$ (presented vertically) were used for demonstrating a vertical list. The child was instructed to work through the item list as quickly as possible and to respond orally. The time limit of 3 min was measured with a stopwatch. The score was the sum of the correct responses (maximum of 45 points). The Pearson correlation coefficient for test–retest in the pilot study was .90, $p < .001$ (2-tailed).

2.4. Intervention

The original version of GraphoGame Math (GGM) was designed as part of the GraphoGame project at the University of Jyväskylä in Finland (see Richardson & Lyytinen, 2014; updated version in Finnish and Swedish at www.lukimat.fi). GGM has been modified and further developed as a tool for individual intervention in early number skills for children aged 6 and 7 who are at-risk for MD.

GGM is theory-based and follows the hierarchy of developmental steps in early arithmetic development (e.g., Geary, 2013). The task types are created on the basis of findings from longitudinal studies. To decrease the risk of MD, one-to-one correspondence, comparing, ordering, and object counting should already be emphasized and strengthened at kindergarten age (Desoete & Grégoire, 2006; Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006; Reeve, Reynolds, Humberstone, & Butterworth, 2012). Both number concept (Geary, 2011a; McClelland, Acock, & Morrison, 2006) and counting skills (Aunio & Niemivirta, 2010; Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Koponen, Salmi, Eklund, & Aro, 2013) are also included in GGM. The evidence-based intervention principles and components recommended for children with (or at-risk for) MDs (e.g., Dowker, 2001; Fuchs et al., 2008; Gersten et al., 2009; Kucian et al., 2011) as well as a personalized user interface with explicit auditory instructions, non-distracting graphics and sound, non-animated stimuli, and corrective and supportive feedback (Seo & Woo, 2010) are also embedded into GGM.

The adaptation in GGM is based on the item difficulty being contingent on the child's performance. The numerical notation proceeds from concrete (objects) to semi-concrete (dots), and finally, to abstract (number words, symbols). Correspondingly, the number range changes from the initial 0–5 to 5–10. An accuracy rate of 85% is required in order to pass a level.

GGM includes six warm-up levels to familiarize the child with the game. After that, one-to-one correspondence (levels 1–3), approximate comparison (levels 4–5), ordering (levels 6–9), and number word — quantity mapping (levels 10–11) are trained. Next, object counting (12–13), exact comparison (14–17), composing (18–19), and decomposing (20–21) are practiced. Finally, conceptual (22–23) and mental addition (24) then symbolic addition (25 —) are practiced. While basic addition is the focus after level 21, GGM provides periodic repetition of all the aforementioned basic number content through levels 25–106. Each level consists of approximately 15 trials.

GGM includes two types of trials: either the child selects (with a mouse) the correct target stimulus (e.g., the number 4) among the alternatives to match an auditory cue (e.g., “four”), or the child organizes (with a mouse) randomly ordered stimuli (e.g., four cards depicting varying numbers of beavers) to match an auditory cue (e.g., “Order, the smallest amount first.”). The child clicks the stimuli one-by-one from the bottom of the screen to fill four empty places at the top of the screen. In each trial, the child confirms their response by clicking the “check” button. In the basic addition tasks, the child selects a correct addend/sum to match an auditory cue (selecting trial) or completes/formulates different types of calculations (cf. Riley, Greeno, & Heller, 1983) by organizing addends/sum and/or the arithmetical symbols plus/equal (organizing trial). The two trial types are present throughout the game, but are never featured together within a single level. In GGM, all stimuli remain on the screen until the child clicks on them. There is no time pressure to make the selections. If the child does not click on any stimulus within 30 s, the game pauses. These pauses are excluded from the active training time.

GGM provides immediate corrective, continuous, and cumulative feedback. After a correct response, the child hears an expression of approval (e.g., “Great,” or “Well done!”), and the correct stimulus is highlighted in green, while the incorrect stimuli disappear. As corrective feedback after an incorrect response, the child hears a supportive expression (e.g., “Not exactly,” or “Try again”), while the incorrect stimulus is highlighted in red and becomes unselectable. The child selects again until the correct stimulus is found and highlighted in green. After three incorrect selections, GGM provides the correct solution. During each level, continuous feedback is offered in the form of a constantly updated bar at the bottom of the screen showing the proportion of correct responses. After each level, cumulative feedback is offered in the form of stickers available for the child to choose from. The more correct responses the child gives, the more stickers are available. These stickers are collected in a personal, virtual sticker album. This activity is excluded from the active training time.

2.5. Procedure

Kindergarten teachers, trained in the assessment procedure, assessed the children individually at each of the four time points. The assessment tasks were presented in a fixed order: enumeration, verbal counting, Number Sets Test, and basic addition. The assessment sessions took place in a quiet, private room at each child's day care center and lasted approximately 20 min. During the 3-week intervention periods, the teachers were asked to arrange a total of between 12 to 15 individual practice sessions during normal kindergarten hours, with 4 to 5 sessions per week. Each session lasted 10 to 15 min. The teachers assisted the children with the headphones (equalizing the audio levels) and helped them to start and end their sessions. Individual game-log data on intervention fidelity and performance during practice was collected throughout the study.

2.6. Data-analysis

Non-parametric tests were used in the analysis (with SPSS version 20) because the distributions were positively skewed and the sample sizes were small. The Wilcoxon signed-rank test was used for analyzing the within-group (GGM) effects. To calculate the within-group effect sizes for the Wilcoxon signed-rank test results, the following formula was used: ES $(r) = \text{Wilcoxon } Z/\sqrt{N}$, where N is the number of observations (Field, 2013). The Mann–Whitney U-test was used to analyze the between-group differences (GGM/CTR) at the initial level, as well as to compare the intervention gain scores. Further, to calculate the between-group effect sizes for the Mann–Whitney U-test results, the aforementioned formula was used. The within-group results and the between-group results in gain scores are interpreted with exact, one-tailed p values. The results for the initial level group comparisons are interpreted with exact, two-tailed p values.

3. Results

The results are presented in three parts. First, the group-level effects are reported ($GGM = 13$). Second, the between-group differences for the GGM and the CTR groups in terms of gain scores are presented. Finally, the descriptive statistics for individual variance in responsiveness to GGM intervention, as well as individual performance based on game-log data are presented. The means of the raw scores, standard deviations, and medians of the two groups at different time points, as well as the statistically significant intervention effects with the gain scores are shown in Table 2.

Table 2
Group performance scores in different time points and between-group comparisons.*

Intervention condition						
		GGM ($n = 13$)		CTR ($n = 8$)		Group comparisons[a]
Variable (max.)	Time	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*	Mann–Whitney U-test
Enumeration (6)	1	4.77 (1.09)	5.00	5.25 (0.71)	5.00	GGM = CTR
	2	4.92 (1.32)[b]	6.00	5.25 (1.04)	5.50	
	Gain 1–2	0.15 (1.21)	0.00	0.00 (0.53)	0.00	GGM = CTR (.13)
	3	5.13 (1.20)	5.71	5.00 (1.51)	6.00	
	4	5.31 (1.03)	6.00	NA	NA	
Verbal counting (12)	1	**3.92 (1.89)**	**4.00**	6.88 (2.80)	5.50	GGM < CTR*
	2	**5.54 (2.22)**[c]	**6.00**	8.00 (2.33)	8.00	
	Gain 1–2	1.62 (1.19)	2.00	1.13 (1.55)	1.00	GGM = CTR (.14)
	3	**5.63 (1.80)**	**5.14**	7.63 (3.25)	8.50	
	4	**7.31 (2.81)**	**8.00**	NA	NA	
Number Sets Test (Part A, 18)	1	**3.08 (3.33)**	**3.00**	1.75 (1.83)	1.50	GGM = CTR
	2	**5.46 (2.44)**[d]	**6.00**	5.88 (5.14)	4.50	
	Gain 1–2	2.38 (3.23)	2.00	4.13 (4.79)	3.50	GGM = CTR (.18)
	3	**5.99 (3.36)**	**7.86**	5.88 (2.30)	6.50	
	4	**5.69 (3.23)**	**6.00**	NA	NA	
Basic addition (45)	1	**1.08 (1.32)**	**0.00**	0.38 (1.06)	0.00	GGM = CTR
	2	**9.00 (6.27)**[e]	**10.00**	5.25 (4.95)	5.00	
	Gain 1–2	7.92 (6.05)	7.00	4.88 (5.08)	3.50	GGM = CTR (.25)
	3	**10.18 (7.37)**	**10.00**	6.25 (4.83)	6.00	
	4	**12.08 (6.79)**	**13.00**	NA	NA	

Note. GGM = GraphoGame Math; CTR = performance level control group; NA = non-applicable.
Significant within-group intervention effects are shown in boldface.
[a] Between-group comparisons are made at time point 1, and for gain scores between time points 1–2 (Wilcoxon *ES* (*r*) in parenthesis).
[b] Within-group Wilcoxon *ES* (*r*) = .06 (close to ceiling); [c].56; [d].45; [e].59.
* $p < .05$.

3.1. The within-group effects of GraphoGame Math intervention

The Wilcoxon test showed a significant group-level improvement in basic addition, verbal counting skills, and the Number Sets Test (Part A) between the first and second assessments ($Z = -2.99$, $p = .001$, $r = .59$; $Z = -2.87$, $p = .002$, $r = .56$; $Z = -2.29$, $p = .021$, $r = .45$, respectively). There was no significant immediate effect on enumeration (the scores were close to ceiling; see Table 2).

The Wilcoxon test also revealed that the improvement in basic addition, verbal counting skills, and the Number Sets Test (Part A) remained stable between the first and third assessments ($Z = -2.98$, $p = .001$; $Z = -2.84$, $p = .002$; $Z = -2.36$, $p = .016$, respectively), as well as between the first and fourth assessments ($Z = -2.94$, $p = .001$; $Z = -2.92$, $p = .002$; $Z = -2.29$, $p = .019$, respectively).

3.2. Intervention effects as compared to the control group

At the first assessment point, there were no significant differences between the two groups (GGM/CTR) in basic addition, enumeration, or the Number Sets Test (Geary et al., 2009). However, the CTR group had better verbal counting skills than the GGM group (Mann–Whitney $U = 18.5$, $Z = -2.46$, $p = .012$; Table 2).

After the GGM intervention, the differences in gain scores between the GGM and CTR groups did not reach significance (Table 2). Due to the non-significant differences, the between-group comparisons were not analyzed further.

3.3. Individual variance in responsiveness to intervention

The children in the GGM group reached the instructed number of intervention sessions ($M = 15.31$, $SD = 1.75$, $Mdn = 15.00$) and minutes (active playing: $M = 103.08$, $SD = 25.87$, $Mdn = 100.00$). However, the game-log data of the *13* children showed a large variance in their performance. The individual progress through the hierarchically organized game content varied greatly, despite initially similar levels of early addition skills. Some of the participants needed a lot of time to complete the easiest non-symbolic contents.

Two figures depict the individual performance during the intervention. Fig. 1 depicts the association between the highest played game level (ranging from 12 to 68) and the gain scores for the intervention period in basic addition (ranged from −1 to 15 points). The cases are rank-ordered by the highest played game level they reached in the 3-week intervention period. Level 21 is marked in Fig. 1 as a starting point for basic addition contents. In general, it seems that the higher the child progressed in GGM levels, the better the gain in basic addition. This notion receives further support from the significant correlation between the highest level played and the intervention gain scores in basic arithmetic (Spearman's *rho* .80, two-tailed $p = .001$).

As seen in Fig. 1, all the participants passed the first 11 non-symbolic levels (containing one-to-one correspondence, approximate comparison, ordering, and number word — quantity mapping). There was a significant correlation between the rank-orders based on the highest level played and the minutes used for passing the easiest levels (Spearman's *rho* .85, two-tailed $p < .001$). This reflects the fact that the more time was used for the easiest non-symbolic contents, the fewer levels were passed. The time used for passing the easiest 11 levels is contrasted with the gain scores of the Number Sets Test (non-symbolic task) in Fig. 2.

Analysis of the game-log data reveals at least three kinds of profiles. First, there were the children who proceeded rather quickly in both basic number and addition levels, and also improved their skills. Second, there were the children who did not reach the addition levels and who did not improve their addition skills, but who still improved their basic number skills. Third, there were the children who did not seem to respond to the intervention at all, and who improved neither their addition nor basic number skills.

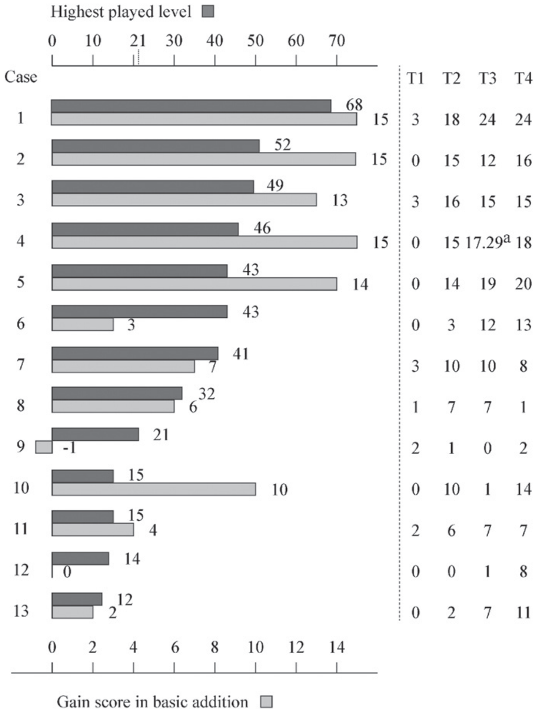

Fig. 1. The association between the highest played GraphoGame Math level and the gain scores of the intervention period in basic addition. Case = intervention participants as a rank-ordered list based on the highest played game level; T1−T4 = individual raw scores in different assessment points in basic addition; Time point 1−Time point 4. ^{a}Missing value replaced by adding the mean gain score of the intervention group in basic addition to the case's score in the second assessment.

4. Discussion

The purpose of this study was to investigate whether intensive computer-assisted intervention (CAI) targeting early number skills and utilizing the recommended principles of effective interventions (Baker et al., 2002; Fuchs et al., 2008; Gersten et al., 2009) and a personalized user interface (Seo & Woo, 2010) would produce positive intervention effects in kindergarteners with poor addition skills. Over the course of a 3-week period, daily practices with GraphoGame Math resulted in within-group improvements in basic addition, verbal counting, and the Number Sets Test. The effect sizes were relatively large (cf. Cohen, 1992), and the intervention improvements remained stable over the 9-week follow-up period. However, there was no between-group difference in gain scores as compared to a wait-list control group.

Previously, stable effects have been reported mainly for teacher-directed training with specific intervention programs (e.g., Dowker, 2001; Griffin, 2004; Van Luit & Schopman, 2000; Wright, 2003), but not after relatively short CAI (e.g., Fuchs et al., 2006). However, utilizing diverse content Schoppek and Tulis (2010); children without mathematics difficulties) and Kucian et al. (2011; children with developmental dyscalculia) reported stable effects in primary school-aged children after intensive CAI. Baroody, Eiland, Purpura, and Reid (2012, 2013) also used mixed training content when introducing CAI to at-risk young children (performance below 25%), but the authors followed the effect only for 2 weeks.

Overall, our results are partially in line with earlier studies in which short, intensive (Cheung & Slavin, 2013) and repetitive training with

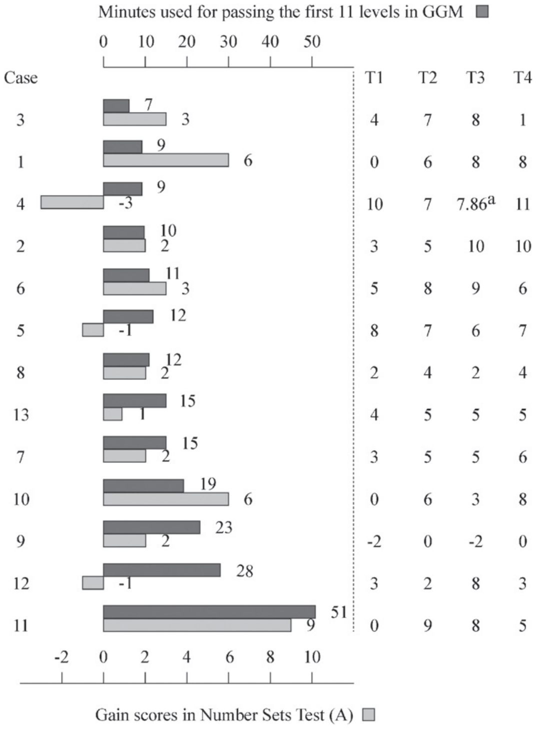

Fig. 2. The association between the time used for passing the basic levels in GraphoGame Math and the gain scores of the intervention period in the Number Sets Test (Part A). Case = intervention participants' case number is based on the highest played game level (presented in Fig. 1); T1 − T4 = individual raw scores in different assessment points in the Number Sets Test (Geary et al., 2009); Time point 1 − Time point 4. ^{a}Missing value replaced by adding the mean gain score of the intervention group in the Number Sets Test to the case's score in the second assessment.

immediate corrective, continuous, and cumulative feedback has been found to be beneficial for children with (or at-risk for) MD (Baker et al., 2002; Fuchs et al., 2008; Gersten et al., 2009). However, there were no significant between-group differences in terms of intervention gain scores. This could be partly due to the large variance in raw and gain scores revealing individual differences in responsiveness to the intervention. In general, a 3-week intervention period might be too short for children severely at-risk for problems in math development. Also, the sensitivity of assessment tools is often an issue in intervention research. The individuality in responsiveness to intervention, and the attention the children and/or kindergarten teachers paid to the assessments and daily kindergarten activities during the ongoing study, could have affected the variances in outcome measures in such a small scale study. The small sample size both reduces the power for revealing less robust effects, and produces over estimated effect sizes (e.g., Cheung & Slavin, 2013; Slavin & Lake, 2008).

The heterogeneity in responsiveness to the intervention and lack of group-level differences suggest that closer attention should be paid to individual development during the intervention. A dual-discrepancy approach has been suggested for discriminating between responders and non-responders to intervention (poor performance level and intervention growth rate as compared to age-peers; Fuchs & Fuchs, 1998; McMaster, Fuchs, Fuchs, & Compton, 2005). Nonetheless, children with (or at-risk for) MD are not homogeneous in terms of the deficits associated with MD and responsiveness to intervention (e.g., Fuchs et al., 2012; Geary, 2004; Jordan et al., 2003; Rubinsten & Henik, 2009). In order to avoid an inaccurate interpretation of (non-)responding based only on the aforementioned criteria, intervention fidelity and/or outcome measures, performance during intervention should also be assessed in more detail. The utilization of game-log data in CAI provides an opportunity to quantify and evaluate individual development during the intervention. However, it has rarely been utilized in analyses or described (except cf. Obersteiner, Reiss, & Ufer, 2013). In this study, the analyses revealed that performance during the intervention varied greatly between participants, despite no problems in intervention fidelity according to both the game-log data and the kindergarten teachers' reports. There were also no marked differences in initial levels of addition skill observed and assessed by the teachers.

Fig. 1 shows that the intervention participants who reached the higher game levels generally improved more in basic addition. Due to the hierarchical structure of the game, addition was practiced after the basic levels were satisfactorily passed (after level 21; see Fig. 1). Participants who passed the basic levels (except in case number 6) showed larger gain scores for basic addition. It seems logical that the more the child was exposed to basic addition training, the larger the benefit. However, using GGM did not result in an improvement in basic addition skills for all participants.

In sum, it is apparent that children with (or at-risk for) MD would benefit from their progress being monitored (Geary, 2011b). From a pedagogical point of view, in order to detect the non-responders, individual skill levels should be followed both during and after the intervention. In our study, the clearest non-responder was case 9, who used a lot of time to pass the basic levels and who did not improve in any early number skill measures after the intervention or during the follow-up period. Alternatively, some children demonstrated delayed effects in basic addition (cases 6, 12, and 13; Fig. 1) and in the Number Sets Test (cases 2, and 12; Fig. 2). Case numbers 8 and 10 provide examples of this delayed change in number skill development. To summarize, the performance level of the children who did not respond to the intervention as expected needs monitoring. Non-responders may need more time, or a more varied and individualized intervention, in order to improve their early number skills. Even though the intervention approach used in the current study was based on the theoretical model of early number skill development (cf. Geary, 2013) and the recommended intervention components (cf. Fuchs et al., 2008), it seems that for ensuring progress through the hierarchical content more specific instructional support is needed for overcoming the individual bottlenecks. This could mean a combination of explicit instructions and massed practice with scaffolding cues towards correct responses and sustainable learning.

4.1. Limitations and implications for future research

Even though the assessments included measures with the test–retest reliability information from a pilot study, a baseline would have allowed for a better control for the test–retest effects. When proposing any program as an evidence-based practice for special education, a baseline should be included in order to control for possible confounding factors (Cook & Cook, 2013; Cook, Tankersley, & Landrum, 2009). Obviously, more well-controlled studies of the effectiveness of intensive CAI with larger MD samples are needed, but also, a *triple-discrepancy approach* could be discussed as a way of identifying non-responders (discrepant from age-peers in performance level and intervention growth rate, but also in delayed effect).

4.2. Conclusions

The findings of this study suggest that a carefully planned computerized intervention can support the development of early number skills in some kindergarteners who exhibit poor addition skills. It can provide a tool for close monitoring of the children who progress slowly or who are not responding to intervention, and who are thus in need of more tailored support. Repetition of basic mathematical concepts, proceeding from the concrete to the abstract, and exposure to hierarchically organized content can strengthen a child's early number skills. However, the individual responsiveness to intervention (Fuchs et al., 2012) needs to be carefully evaluated. It could also be worth to pay more attention to the adaptation and pedagogically meaningful instruction during the response process in order to ensure development during the practice and/or identifying the individual needs of (non-)responders to intervention.

Acknowledgments and ethical conducts

The developmental work on GraphoGame Math is carried out as part of the LukiMat-project, which was funded by the Ministry of Education and Culture in Finland (DNRO 20/240/2006) (2007–2015). The data used in this study was collected during the same project. The authors declare no potential conflicts of interest with respect to the authorship or publication of this data. The guidelines of the Finnish Advisory Board on Research Integrity were carefully followed in the study. All parents provided written permission for their child to participate in both the assessments and the intervention. All parents were also informed of their right to discontinue participation at any time. Children on the waiting list who were identified as displaying risks in their development of early number skills were also given an opportunity to receive extra support after a short period. In addition, all children in the participating day care centers were allowed to use the intervention programs after the study period if their parents gave consent for them to do so.

References

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences*, *20*(5), 427–435. http://dx.doi.org/10.1016/j.lindif.2010.06.003.

Aunola, K., Leskinen, E., Lerkkanen, M. -K., & Nurmi, J. -E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, *96*(4), 699–713. http://dx.doi.org/10.1037/0022-0663.96.4.699.

Baker, S., Gersten, R., & Lee, D. -S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal*, *103*(1), 51–73 retrieved from http://www.Jstor.org/stable/1002308.

Baroody, A.A.J., Eiland, M.D., Purpura, D.J., & Reid, E.E. (2012). Fostering at-risk kindergarten children's number sense. *Cognition and Instruction*, *30*(4), 437–470. http://dx.doi.org/10.1080/07370008.2012.720152.

Baroody, A.A.J., Eiland, M.D., Purpura, D.J., & Reid, E.E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal*, *1*, 1–41. http://dx.doi.org/10.3102/0002831212473349.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*(1), 3–18. http://dx.doi.org/10.1111/j.1469-7610.2005.00374.x.

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science*, *332*, 1049–1053. http://dx.doi.org/10.1126/science.1201536.

Cheung, aA.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. http://dx.doi.org/10.1016/j.edurev.2013.01.001.

Christensen, C.A., & Gerber, M.M. (1990). Effectiveness of computerized drill and practice games in teaching basic math facts. *Exceptionality*, *1*, 149–165. http://dx.doi.org/10.1080/09362839009524751.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *122*(1), 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155.

Cook, B.G., & Cook, S.C. (2013). Unraveling evidence-based practices in special education. *Journal of Special Education*, *47*(2), 71–82. http://dx.doi.org/10.1177/0022466911420877.

Cook, B.G., Tankersley, M., & Landrum, T.J. (2009). Determining evidence-based practices in special education. *Exceptional Children*, *75*(3), 365–383 [retrieved from http://web.Ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&sid=75a11115-bab9-4bfd-b9f4-18b70eb30267%40sessionmgr115&hid=118].

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford, UK: University Press.

Desoete, A.A., & Grégoire, J. (2006). Numerical competence in young children and in children with mathematics learning disabilities. *Learning and Individual Differences*, *16*, 351–367. http://dx.doi.org/10.1016/j.lindif.2006.12.006.

Dowker, A.A. (2001). Numeracy recovery: A pilot scheme for early intervention with young children with numeracy difficulties. *Support for Learning*, *16*, 6–10. http://dx.doi.org/10.1111/1467-9604.00178.

Duncan, G.J., Dowsett, C.J., Claessens, A.A., Magnuson, K., Huston, aA.C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. http://dx.doi.org/10.1037/0012-1649.43.6.1428.

Field, A.A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE Publications Ltd.

Finnish National Board Of Education (2010). Core curriculum for preschool education in Finland. Downloadable (in English) from http://www.oph.fi/download/153504_national_core_curriculum_for_pre-primary_education_2010.pdf

Fuchs, L.S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice*, *13*(4), 204–219.

Fuchs, L.S., Fuchs, D., & Compton, D.L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities*, *45*(3), 257–269. http://dx.doi.org/10.1177/0022219412442167.

Fuchs, L.S., Fuchs, D., Hamlet, C.L., Powell, S.R., Capizzi, A.M., & Seethaler, P.M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities*, *39*(5), 467–475. http://dx.doi.org/10.1177/00222194060390050701.

Fuchs, L.S., Fuchs, D., Powell, S.R., Seethaler, P.M., Cirino, P.T., & Fletcher, J.M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly*, *31*(2), 79–92. http://dx.doi.org/10.2307/20528819.

Geary, D.C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*(1), 4–15. http://dx.doi.org/10.1177/00222194040370010201.

Geary, D.C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science*, *22*(1), 23–27. http://dx.doi.org/10.1177/0963721412469398.

Geary, D.C., Bailey, D.H., & Hoard, M.K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment*, *27*(3), 265–279. http://dx.doi.org/10.1177/0734282908330592 [Finnish version: Numerosetit (Koponen, T., Salminen, J., Räsänen, P., & Manninen, J. 2008.) Niilo Mäki Institute. Unpublished.].

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

Geary, D.C. (2011b). *Journal of Developmental and Behavioral Pediatrics*, *32*(3), 250–263. http://dx.doi.org/10.1097/DBP.0b013e318209edef.

Geary, D.C., Hoard, M.K., & Bailey, D.H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities*, *45*(4), 291–307. http://dx.doi.org/10.1177/0022219410392046.

Gersten, R., Jordan, N.C., & Flojo, J.R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*, 293–304. http://dx.doi.org/10.1177/00222194050380040301.

Gersten, R., Chard, D.J., Jayanthi, M., Baker, S.K., Morphy, P., & Flojo, J.R. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, *79*(3), 1202–1242. http://dx.doi.org/10.3102/0034654309334431.

Griffin, S. (2004). Building number sense with number worlds: A mathematics program for young children. *Early Childhood Research Quarterly*, *19*, 173–180. http://dx.doi.org/10.1016/j.ecresq.2004.01.012.

Hativa, N., & Shorer, D. (1989). Socioeconomic status, aptitude and gender differences in CAI gains of arithmetic. *The Journal of Educational Research*, *83*(1), 11–21 Retrieved from http://www.jstor.org/stable/27540363.

Jordan, N.C., Hanich, L.B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, *74*(3), 834–850. http://dx.doi.org/10.1111/1467-8624.00571.

Jordan, N.C., Kaplan, D., Nabors Oláh, L., & Locuniak, M.N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, *77*(1), 153–175. http://dx.doi.org/10.1111/j.1467-8624.2006.00862.x.

Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews*, (15), 60–68. http://dx.doi.org/10.1002/ddrr.46.

Käser, T., Baschera, G. -M., Kohn, J., Kucian, K., Richtmann, V., Grond, U., ... von Aster, M. (2013). Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Frontiers in Psychology*, *4* [Advance online publication. doi: 10.3389/fpsyg.2013.00489].

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology*, *105*(1), 162–175. http://dx.doi.org/10.1037/a0029285.

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology*, *103*, 516–531. http://dx.doi.org/10.1016/j.jecp.2009.03.009.

Kraus, W.H. (1981). Using a computer game to reinforce skills in addition basic facts in second grade. *Journal for Research in Mathematics Education*, *12*(2), 152–155 Retrieved from http://www.jstor.org/stable/748709.

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönman, C., Plangger, F., ... von Aster, M. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage*, *57*(3), 782–795. http://dx.doi.org/10.1016/j.neuroimage.2011.01.070.

McClelland, M.M., Acock, A.C., & Morrison, F.J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*, 471–490. http://dx.doi.org/10.1016/j.ecresq.2006.09.003.

McMaster, K.N., Fuchs, D., Fuchs, L.S., & Compton, D.L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*(4), 445–463.

Mevarech, Z.R., & Rich, Y. (1985). Effects of computer-assisted mathematics instruction on disadvantaged pupils' cognitive and affective development. *The Journal of Educational Research*, *79*(1), 5–11.

Morgan, P.L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*(4), 306–321. http://dx.doi.org/10.1177/0022219408331037.

Murphy, M.M., Mazzocco, M.M.M., Hanich, L.B., & Early, M.C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, *40*(5), 458–478. http://dx.doi.org/10.1177/00222194070400050901.

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, *23*, 125–135. http://dx.doi.org/10.1016/j.learninstruc.2012.08.004.

Okolo, C.M. (1992). The effect of computer-assisted instruction format and initial attitude on the arithmetic facts proficiency and continuing motivation of students with learning disabilities. *Exceptionality*, *3*(4), 195–211. http://dx.doi.org/10.1080/09362839209524815.

Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology*, *4*(141), 649–666. http://dx.doi.org/10.1037/a0027520.

Richardson, U., & Lyytinen, H. (2014). The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*, *10*(1), 39–60 [Downloadable from http://www.humantechnology.jyu.fi/articles/volume10/2014/Richardson_Lyytinen.pdf].

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York, NY: Academic Press.

Rubinsten, O., & Henik, A. (2009). Developmental dyscalculia: Heterogeneity may not mean different mechanisms. *Trends in Cognitive Science*, *13*(2), 92–99. http://dx.doi.org/10.1016/j.tics.2008.11.002.

Salminen, J., Räsänen, P., Koponen, T., & Aunio, P. (2008a). The number word — quantity correspondence test. Niilo Mäki Institute. Unpublished.

Salminen, J., Räsänen, P., Koponen, T., & Aunio, P. (2008b). The basic addition fluency test. Niilo Mäki Institute. Unpublished.

Salonen, P., Lepola, J., Vauras, M., Rauhanummi, T., Lehtinen, E., & Kinnunen, R. (1994). *Diagnostiset testit 3. Motivaatio, Metakognitio ja Matematiikka [Diagnostic Tests 3: Motivation, metacognition and mathematics]*. Turku, Finland: Centre for Learning Research, University of Turku.

Sarama, J., & Clements, D.H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Schoppek, W., & Tulis, M. (2010). Enhancing arithmetic and word-problem solving skills efficiently by individualized computer-assisted practice. *The Journal of Educational Research*, *103*, 239–252. http://dx.doi.org/10.1080/00220670903382962.

Seo, Y. -J., & Bryant, D.P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education*, *53*, 913–928. http://dx.doi.org/10.1016/j.compedu.2009.05.002.

Seo, Y. -J., & Woo, H. (2010). The identification, implementation, and evaluation of critical user interface design features of computer-assisted instruction programs in mathematics for students with learning disabilities. *Computers & Education*, *55*(1), 363–377. http://dx.doi.org/10.1016/j.compedu.2010.02.002.

Shin, N., Norris, C., & Soloway, E. (2006 June 27 — July 1). Effects of handheld games on students learning in mathematics. *Proceedings of the 7th international conference on learning science(pp. 702–708). Bloomington, IN* Retrieved from http://www.fisme.science.uu.nl/publicaties/literatuur/endnote_ecgbl_472_shin.pdf.

Slavin, R.E., & Lake, C. (2008). Effective programs in elementary mathematics: A best evidence synthesis. *Review of Educational Research*, *78*(3), 427–455. http://dx.doi.org/10.3102/0034654308317473.

Trifiletti, J.J., Frith, G.H., & Armstrong, S. (1984). Microcomputers versus resource rooms for LD students: A preliminary investigation of the effects on math skills. *Learning Disability Quarterly*, *7*(1), 69–76 Retrieved from http://www.jstor.org/stable/1510263.

Van Luit, J.E.H., & Schopman, E.A.M. (2000). Improving early numeracy of young children with special educational needs. *Remedial and Special Education*, *21*(1), 27–40. http://dx.doi.org/10.1177/074193250002100105.

Von Aster, M.G. (2000). Developmental cognitive neuropsychology of number processing and calculation: Varietes of developmental dyscalculia. *European Child and Adolescent Psychiatry*, *9*(II), 41–57. http://dx.doi.org/10.1007/s007870070008.

Von Aster, M.G., & Shalev, R.S. (2007). Number development and developmental dyscalculia. *Developmental Medicine and Child Neurology*, *49*(11), 868–873. http://dx.doi.org/10.1111/j.1469-8749.2007.00868.x.

Wilson, A.J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, *3*(4), 224–234. http://dx.doi.org/10.1111/j.1751-228X.2009.01075.x.

Wilson, A.J., Dehaene, S., Pinel, P., Revkin, S.K., Cohen, L., & Cohen, D. (2006a). Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*(19), 1–14. http://dx.doi.org/10.1186/1744-9081-2-19.

Wilson, A.J., Revkin, S.K., Cohen, D., Cohen, L., & Dehaene, S. (2006b). An open trial assessment of the Number Race, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*(20), 1–36. http://dx.doi.org/10.1186/1744-9081-2-20.

Wright, R.J. (2003). Mathematics recovery: A program of intervention in early number learning. *Australian Journal of Learning Disabilities*, *8*(4), 6–11. http://dx.doi.org/10.1080/1940415030954674.

III

RESPONSE TO COMPUTERIZED EARLY NUMBER INTERVENTION: THE ROLE OF RISK-LEVEL

by

Jonna Salminen, Tuire Koponen, Riikka Sorvo, Pilvi Peura, & Mikko Aro

(Manuscript submitted for publication)

Response to computerized early number intervention: The role of skill-level

Jonna B. Salminen, M.Ed., University of Jyvaskyla, Department of Education, Special Education, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland, jonna.salminen@jyu.fi;

Tuire K. Koponen, PhD, University of Jyvaskyla, Department of Education, Special Education, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland, tuire.k.koponen@jyu.fi;

Riikka Sorvo, M.Ed., University of Jyvaskyla, Department of Teacher Education, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland, riikka.sorvo@jyu.fi; and

Pilvi Peura, M.Ed., University of Jyvaskyla, Department of Education, Special Education, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland, pilvi.peura@jyu.fi;

Mikko T. Aro, PhD, University of Jyvaskyla, Department of Education, Special Education, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland, mikko.t.aro@jyu.fi

Correspondence concerning this article should be addressed to
Jonna Salminen, University of Jyvaskyla, Department of Education, Special Education,
phone number: +358 40 805 3027
e-mail address: jonna.salminen@jyu.fi
postal address: P.O. Box 35, FI-40014 University of Jyvaskyla, Finland

Acknowledgements and ethical conducts
The development of GraphoGame Math and the data collection of the study have been administered during the LukiMat -project (started in 2007). The project is funded by the Ministry of Education and Culture in Finland. The authors declare no conflicts of interest with respect to the authorship or publication of the data. The ethical guidelines of the Finnish Advisory Board on Research Integrity were followed during the study. We wish to thank the participating day care centers, the children, and their teachers, as well as the parents.

Response to computerized early number intervention: The role of risk-level

The poorest-performing children in early number skills (those performing below the 10th percentile) seem to have even higher risk for mathematics difficulties than their low-achieving age-peers (those performing between the 11th and 25th percentiles). However, little is known of how this skill-level group responds to early intervention. In this study, originally of *278* children, the intervention effects of an adaptive computer game were examined among the low-performing children, and then separately among the poorest-performing children. After two pre-tests, intervention participants (6–7-year-olds) were randomized in two groups, first practicing either number concept skills or number sequence knowledge, and thereafter basic addition skills. Among low-performing children (*33* out of *278*), the results indicated similar benefits in both conditions. However, among the poorest-performing children (*14*), the improvements were dependent on the training content. The children who focused on exact symbolic processing before basic addition practice showed better improvement in their basic addition skills.

Keywords: early number skills; mathematics difficulties; computer-assisted intervention (CAI); adaptation; feedback

Response to computerized early number intervention

Children with the most severe difficulties in early number skills (performance below the 10th percentile) differ from their low-achieving age-peers (performance between the 11th and 25th percentiles) by having even more persistent deficits (Chong & Siegel, 2008; Geary, Hoard, & Bailey, 2012). This difference can be seen in processing numbers, learning arithmetic procedures, retrieving arithmetic facts, in working memory (Geary, 2011), and in achievement growth rates from kindergarten to later primary grade levels (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Morgan, Farkas, & Wu, 2009; Murphy, Mazzocco, Hanich, & Early, 2007). However, very little is known about how the poorest-performing children benefit from intensified early number skill support.

In order to understand relationships between quantities, to compose and decompose numbers, and to learn basic arithmetic, as well as eventually to be fluent in calculations, children first need to learn basic numerical skills (Geary, 2013; Krajewski & Schneider, 2009; Salaschek, Zeuch, & Souvignier, 2014). The abilities to detect and compare numerosities (i.e., magnitude comparison skills), to match corresponding number words, quantities and number symbols together (i.e., number concept skills), and to recite number words and set them in the correct order (i.e., number sequence skills) are all necessary for learning the aforementioned number relations and arithmetic skills.

Children with atypical development seem to have deficits particularly in processing numerical information (e.g., De Smedt, Noël, Gilmore, & Ansari, 2013; Geary, 2011). Two hypotheses for factors underlying mathematics difficulties (MD) have been proposed (see De Smedt, & Gilmore, 2011). The first hypothesis is based on the theory that an innate ability to understand quantities affects later learning of number symbols and basic arithmetic skills (Butterworth, 2005; defective number module hypothesis). The second hypothesis is based on the theory that children with MD have deficits in "accessing" numerical meaning from number symbols (De Smedt & Gilmore, 2011; Rousselle & Noël, 2007; access deficit hypothesis). If the first hypothesis receives support, the early intervention should focus on supporting non-symbolic quantity discrimination (De Smedt, & Gilmore, 2011). If the second hypothesis receives support, the early intervention should focus on symbolic number processing (i.e., mapping between number symbols and their corresponding quantities; De Smedt, & Gilmore, 2011; Schneider et al., in press). In this study the access deficit hypothesis was selected for examination because symbolic magnitude processing strongly predicts basic arithmetic skills (recently Bartelet, Vaessen, Blomert, & Ansari, 2014; for review Schneider et al., in press).

Besides the two former hypotheses, explicit number system knowledge (cf. Geary, 2013) requires mastering the counting procedure. Thus, another potential factor underlying MD could be a deficit in learning number sequence knowledge. For memorizing basic arithmetic combinations, the development typically proceeds from counting to reasoning strategies, and thereafter automatically retrieving the response from long-term memory (Baroody, 1987). In particular, children with MD (or those who are at risk) seem to struggle in their strategy development; they typically have difficulty to use shortened counting strategies, such as counting on from the first addend instead of counting all items of both addends one by one (Wylie, Jordan, & Mulhern, 2012). In addition to this, verbal counting skills have been found to strongly predict calculation fluency (Koponen, Salmi, Eklund, & Aro, 2013). Therefore, the effects of early number sequence training on the development of arithmetic skills should also be explicitly assessed.

One potential method for supporting early number skills is computer-assisted intervention (CAI) (cf. Butterworth, Varma, & Laurillard, 2011). For example, by linking verbal counting, object counting, numerical relations, written numbers, and arithmetic (Baroody, Eiland, Purpura, & Reid, 2012, 2013); locating the number of dots, written numbers, or sums and differences on a number line (Kucian et al., 2011); or training ordering and number line skills with arithmetic operations (Käser et al., 2013) encouraging findings have been reported in children having (or at risk for) MD. The present study extends previous research by examining the effects of CAI on very basic number skills (i.e., dot counting, verbal counting, and basic addition) among low- and the poorest performing children, and whether the number concept or number sequence knowledge training is more beneficial for learning basic arithmetic.

Even though theory-based CAIs can provide the possibility to individualize and intensify early number skill practice, there are also other pedagogical aspects worth remembering when developing CAI methods. Via parameters related to instructions, cues and stimulus presentation, the actual content of practice, the requirement of cognitive effort, and the possibilities for relational learning (i.e., linking numerical sub-skills) could be modified (cf. Hunt, Valentine, Bryant, Pfannenstiel, & Bryant, 2015; Fuchs et al., 2008). Furthermore, explicit immediate, continuous, and corrective feedback types are all recommended for supporting learning in children having (or at risk for) MD (Baker, Gersten, & Lee, 2002; Fuchs et al., 2008; Gersten et al., 2009; Moreno, 2004). Clear, specific, and explicit task-focused feedback is required when enhancing learning and performance (Shute, 2008). Cues should be offered for facilitating performance and supporting response process instead of terminating the task, as direct dichotomous feedback (i.e., correct or incorrect) can easily do (Shute, 2008). Therefore, it might be meaningful to scaffold the response process so that the feedback system guides a child towards reaching the correct response (cf. Finn & Metcalfe, 2010; Hunt et al., 2015). The feedback system of the intervention method used in this current study was based on these suggestions.

The Study

The purpose of this study was to examine the effects of an educational computer game (GraphoGame Math) on early number skills among children in Finnish pre-primary education

(6–7-year-olds) who were identified as needing extra support in math before starting formal schooling the following autumn. The participants were divided into two groups: one practiced first number concept skills and thereafter basic addition skills (the NC-group), and the other practiced first number sequence knowledge and thereafter basic addition skills (the NS-group). Both conditions aimed to increase conceptual and procedural knowledge, as well as to build links between the numerical sub-skills needed for learning basic arithmetic (cf. Dowker, 2009; Geary, 2013; Krajewski & Schneider, 2009; Salaschek et al., 2014).

There were two research questions:

1) Does a computerized early number intervention produce immediate, condition-specific effects for dot counting fluency and verbal counting skills within a group of children in pre-primary education who are in need of extra support for their early number skills?

2) To what extent do the two intervention conditions used in the current study improve basic addition skills within a group of children in pre-primary education in need of extra support for their early number skills?

It was expected that a targeted (Dowker & Sigley, 2010; Geary, 2011; here number concept and number sequence knowledge trainings), intensive (Cheung & Slavin, 2013), and individually adaptive training (Fuchs, Fuchs, & Compton, 2012) with supportive feedback (cf. Hunt et al., 2015; Fuchs et al., 2008) would produce content-specific effects after the first intervention period. The effects of the two conditions on basic addition skills over the study (i.e., after the second intervention period) were tested. Both the content of the three training components and the feedback system are described in more detail in the “Intervention Method” section.

For the purpose of this study, the analyses were first conducted with all participants nominated by pre-primary education teachers and special educators as being in need of extra support for early number skills (Examination 1). The analyses were then conducted in a sub-sample of nominated children who were most at risk for MD on the basis of individual pre-assessments (performance below the 10th percentile as compared to the age-level normative data; Examination 2). For ethical reasons, a business-as-usual control group of low- or the poorest-performing children was not set for our relatively long intervention study of 16 weeks.

Method

Participants

Voluntary teachers, who had observed children in pre-primary education through the autumn semester (*n* = 278), nominated children who, in their opinion, were most in need of extra support for early number skills. At the beginning of the spring semester, the teachers selected a total of 33 candidates (6–7-year-olds) for the study. All candidates had parental consent to participate in individual assessments and computerized interventions. Parents were informed of their right to cancel participation in the study at any time. Participants came from 14 public day care centers in central Finland, and were native speakers of Finnish. None of them had severe visual, hearing, motor, or intellectual impairments.

The participating children were randomized into two experimental groups: the NC-group (11 girls and 6 boys) and the NS-group (11 girls and 5 boys). The analyses were carried out separately for the whole sample (*n* = 33; 12 % out of *278* children; Examination 1) and a sub-sample of the poorest-performing children (*n* = 14; 5 % out of *278* children; the NC-group: 7 girls and 1 boy, and the NS-group: 4 girls and 2 boys; Examination 2). The inclusion criterion for the latter group was performance below the 10th percentile in at least two of the three core early number skill measures: dot counting fluency, verbal counting, and basic addition. The population-based normative data used for determining the age-referenced level of performance were collected in separate studies carried out within a two-month period of the pre-tests of this study. The descriptive statistics for both the intervention and the reference groups are presented later in Tables 1, 2, 3, and 4.

Measures

Background measures. The following background measures were administered to assess the potential initial differences between the two intervention groups. Peabody Picture Vocabulary Test (PPVT) was conducted in the first pre-test, and The Number Sets Test, non-symbolic story problems, as well as Rapid Automatized Naming were conducted in the first and the second pre-tests. Two pre-tests were carried out to control for the stability of initial performance-level among the poorest-performing children. This level was defined by the sum score of the PPVT and by the average scores of the other background measures.

The Peabody Picture Vocabulary Test. A shortened version (30 items) of the Peabody Picture Vocabulary Test (PPVT − Revised; Dunn & Dunn, 1981; for the experimental shortened version, see Lyytinen et al., 2004) was administered to control for vocabulary. The test consists of pages containing four outline drawings. The child is presented with a set of pictures and is given a word describing one of the pictures. The child has to point out the picture that the word describes. Before the test, two sets of pictures were presented for practice. The score was the sum of the correct responses (maximum of 30 points).

The Number Sets Test. The Number Sets Test (Geary, Bailey, & Hoard, 2009) was used to control for the understanding of decomposing and composing principles (i.e., part-whole schema). In this test, the child was asked to determine as quickly and accurately as possible if pairs or trios (presented as sets of objects, and/or Arabic numbers) matched the target number (here number 5). In the first part, only sets of objects were presented; in the second part, sets of objects and numbers were mixed. The Cronbach’s alpha values for correctly identified pairs or trios have been found to be .88 (cf. Geary et al., 2007). Before starting the test, the target number 4 was used to demonstrate the task, and the target number 3 was administered as practice. Of 36 items per part, 18 matched the target number and 18 did not match. The score was the number of the correct responses minus the incorrect responses within the time limit of one minute, scored separately for both parts.

Non-symbolic story problems. Non-symbolic story problem solving skills were assessed to control for the concepts of adding and taking away. The task included three simple addition (2+1, 3+2, and 2+2), and subtraction problems (3-1, 4-2, and 5-1) (Salminen, Räsänen, Koponen, & Aunio, 2008). The tester took beads (so the child could see them) and put them under a scarf, and simultaneously said aloud the amount. After that, the tester took another number of beads, added them under the scarf, and again simultaneously said aloud the amount. The child was asked to say how many beads there were under the scarf in total. In subtraction tasks, the tester took beads and put them under a scarf while saying aloud the amount. Next, the tester took beads from under the scarf, again saying aloud the amount. The child was asked to tell how many beads remained under the scarf. The score was the sum of the correct responses (maximum of 6 points). The Spearman's correlation coefficient for test-retest was .68.

The Rapid Automatized Naming. The Rapid Automatized Naming (RAN) of objects (Denckla & Rudel, 1974; standardized Finnish version by Ahonen, Tuovinen, & Leppäsaari, 2006) was administered in order to assess the initial level of processing speed. This test included five different objects (car, house, fish, pen, and ball) each presented several times in pseudorandom order, with no consecutive presentations of the same stimulus. Altogether 50 stimuli were arranged in five rows. Before the test, practice items were presented to ensure that the child knew the names of the objects. The child was instructed to name all stimuli as quickly and accurately as possible. A stopwatch was used to measure the time of completion, which was used in the analyses. The Spearman's correlation coefficient for test-retest was .74.

Outcome measures. To evaluate the potential initial differences between the two intervention groups and to examine the effects of GraphoGame Math intervention dot counting fluency, verbal counting accuracy, and basic addition skills were assessed at each of the five time points. The average scores of the two pre-tests were calculated for the tasks to define the initial performance-level.

Dot counting fluency. In the dot counting fluency test, the child was asked to rapidly name sets of dots with quantities varying from 1 to 5 (Salminen & Koponen, 2010). The child was shown a matrix (5 x 8) with a total of 40 stimuli of varying sets of randomly arranged dots. The child was asked to say aloud the number of dots in each set as quickly as possible. Before the actual test, three sets were presented for practice. Dot counting fluency was scored as the total time (used for naming all sets of dots) divided by the number of the correct responses. The Spearman's correlation coefficient for test-retest was .70.

Verbal counting accuracy. To assess verbal counting skills, three different counting tasks were adapted from Diagnostic Tests 3 (Salonen et al., 1994). In the counting forward subtest, the child was asked to count forward starting from one. If the child reached the number words between 2 and 9, a score of one point was given. Correctly reaching the number words between 10 and 19 yielded two points; 20−29 three points; and 30 four points. In the counting forward from a given number subtest, the child was asked to count four steps forward from the four given number words (3, 8, 12, and 19). One point was given for each correct response. In the counting backward from a given number subtest, the child was asked to count four steps backward from the four given number words (4, 8, 12, and 23). One point was given for each correctly given sequence. A total score (maximum of 12 points) was used for the analyses. The Cronbach's alpha in the first pre-test was .74, and in the second pre-test it was .79. The Spearman's correlation coefficient for test-retest was .91.

Basic addition. Basic addition skill was measured by a time-limited test of 3 minutes consisting of “a + b = __” -type addition problems, in which the sum was 10 or less (cf. Salminen, Koponen, Leskinen, Poikkeus, & Aro, in press). The tasks were randomly ordered with two exceptions: the problems in which the sum was five or less were presented first, and problems with the same or highly similar addends (e.g., 2 + 2 = __ and 2 + 3 = __) did not follow each other. “Add zero” problems were excluded. The task “1 + 1 = __” was administered as practice, and the tasks “3 + 3 = __” and “1 + 5 =__” were used for demonstrating the vertical list presentation. The child was instructed to go through the list of problems as quickly as he/she could and to provide the responses orally. The score was the sum of the correct responses given within the time limit. The Spearman’s correlation coefficient for test-retest was .94.

Intervention Method

GraphoGame Math. GraphoGame Math (GGM; specific intervention software for children in need of preventive support for early number skills; see Salminen et al., in press) includes three training components: number concept (NC), number sequence (NS), and basic addition (BA). The NC-component aims to support number concept skills with the following tasks: approximate comparison tasks (40 % of all number concept tasks), number word-quantity-number symbol mapping tasks (10 %), and exact comparison tasks (50 %). Over 70 % of the tasks require symbolic number processing skills. In such tasks, the child has to match or discriminate the numerical value of (and between) non-symbolic and symbolic stimuli which are simultaneously presented among alternatives in each trial (cf. access deficit hypothesis; De Smedt & Gilmore, 2011; Rousselle & Noël, 2007).

The NS-component requires continuing and completing number sequences forward and backward (80 % of all number sequence tasks), and recognition of number neighbors (i.e., number-after and number-before rules; 20 %). Since the correct sequences are repeated aloud by GGM, the tasks support both verbal counting and number sequence knowledge, which are supposedly skills important for arithmetic strategy development (cf. Baroody, Bajva, & Eiland, 2009; Wylie et al., 2012).

The BA-component consists of composing and decomposing numbers (30 % of all basic addition tasks), mental addition (10 %), non-symbolic addition (10 %), and symbolic addition (50 %). The tasks are supposed to support basic arithmetic (here, addition) skills, in which the deficit is argued to be the core manifestation of mathematics disabilities (cf. Geary, 2011).

Each of the three GGM components starts with tutorials that introduce the forthcoming mathematical concepts and contents, and provide two trials for practicing the related tasks. The purpose of the tasks is explained auditorily by an agent (presented visually as a pelican figure). After the tutorial, each component starts with content-specific assessments (maximum of eight trials for each of the four number ranges: 0−5, 3−7, 6−10, and 10−20) to select an appropriate individual training level. After training within each addressed number area, GGM ensures the child's accuracy level by reassessing the trained content before letting the child proceed to the more demanding training level. Through the GGM training, the child either selects the correct stimulus (or stimuli) among the alternatives or organizes the alternatives to match an auditory instruction by clicking the mouse. The child can return the selection(s) back to the original places by clicking the mouse on the stimulus. The child confirms the response by clicking the "check" button on the screen.

The adaptation algorithm was based on the principle of gradually increasing difficulty levels by changing (1) the numerical representation (concrete objects, dots, and number symbols), (2) the distance between correct and incorrect response alternatives (numerical distance between the correct and incorrect stimuli was initially large, but through progress in training the distance diminished gradually), (3) the amount of (in)correct response alternatives presented and correct stimuli required per individual trial (both varied from one to three), and (4) the number range (cf. Dehaene, 2011). The adaptation algorithm aimed to keep the individual success rate at 80 percent.

All GGM components include the “scaffolding respond-until-correct” feedback system. The explicit immediate, continuous, and corrective feedback (Baker et al., 2002; Fuchs et al., 2008; Gersten et al., 2009), was given in three ways: visually only, visually combined with an auditory signal, or visually with an extra instruction given by the animated pedagogical agent. Rewarded feedback (Fuchs et al., 2008) was offered using a visual success bar presented at the bottom of the screen. On the basis of performance, the child was also awarded stickers at the end of each level (i.e., a set of 20 tasks focusing on certain content, e.g., mapping between different numerical representations). Cues were offered (cf. Finn & Metcalfe, 2010; Hunt et al., 2015) after an incorrect response. For example, if the child completed a number sequence such as “3, __, __, __, 7” incorrectly (e.g., “3, 5, 4, 6, and 7”), the incorrect stimuli were removed and the first correct stimulus (here, number symbol 4) was moved to the correct place in the sequence. This was a cue for the correct solution (corrective and scaffolding feedback). If the child made a comparable error again, the incorrect stimulus was removed from the sequence (immediate feedback), and the next correct stimulus (number symbol 5) was moved to its place (corrective, scaffolding feedback) before the child could continue further. If the child made a third comparable error, the third (and final) correct stimulus (number symbol 6) was moved to the right place (corrective feedback), and GGM repeated aloud the resulting correct number sequence (restorative feedback).

Fidelity to intervention

The game log-data were used for analyzing the number of sessions played, and the time used for active practice in order to control for intervention fidelity and exposure to intervention. Furthermore, the pre-primary education teachers were instructed to keep a diary for practice sessions and their lengths; they were also contacted by telephone to ensure that both intervention periods were started as instructed, and to offer guidance if needed. During the intervention periods of three weeks, the participants followed their pre-primary education curriculum (for five days a week, and for 2–3 hours per day). Typically, pre-primary education activities focus on supporting children’s learning-to-learn skills, and positive self-image, as well as basic skills, knowledge, and capabilities that are needed for forthcoming formal schooling (Finnish National Board of Education, 2010; downloadable in English). No formal instruction in reading, spelling, or arithmetic is given.

Procedure

The current study included five assessment points and two cycles of intervention. After the two pre-tests (weeks 1−2), the two study groups were formed via randomization (week 2). During weeks 3−5, one group participated in number concept training and the other one in number sequence training. After this, the children participated in the third assessment (week 6). After a pause of one week (week 7; holiday season), both groups participated in basic addition training (weeks 8−10). The fourth assessment (week 11) was conducted after this second three-week intervention period, and the fifth assessment was administered five weeks later (week 16). The study design is presented in Figure 1.

Participants were assessed individually at each of the five time points (two times in January, then February, March, and May) by the third and the fourth author trained for the assessment procedure. All assessment sessions took place in a quiet room and lasted for approximately 20 minutes. The assessment tasks were presented in a fixed order: PPVT, verbal counting skills, the Number Sets Test, story problems, basic addition, dot counting fluency, and RAN of objects.

During both three-week intervention periods, the pre-primary education teachers were asked to arrange individual practice sessions with a frequency of three to five sessions per week. Each session was instructed to last approximately 10 to 15 minutes. Teachers helped the children with the headphones (equalizing audio levels), and in starting and ending the intervention sessions.

Insert Figure 1 here.

Data Analysis

The analyses were carried out with SPSS 22. Non-parametric methods were used because the distributions were positively skewed and the sample sizes were small. The between-group differences at the initial performance level were analyzed by the Mann-Whitney U-test. The results were interpreted with Exact, two-tailed p values.

To analyze the within-group intervention effects, the Wilcoxon signed-rank test was used. The results were interpreted with Exact, one-tailed p values for the condition-specific variables, which were dot counting fluency in the NC-group and verbal counting ability in the NS-group during the first intervention period, and basic addition in both groups over both intervention periods. To analyze the between-group effects, the differences in gain scores were tested by the Mann-Whitney U- test. The test results were interpreted with Exact, one-tailed p values over the first intervention period, and with Exact, two-tailed p values over both intervention periods because there was no exact hypothesis of one of the two conditions producing larger gain scores in basic addition than the other. To calculate both the within-group effect sizes of the Wilcoxon signed-rank test, and the between-group effect sizes of the Mann-Whitney U -test, the following formula was used: ES (r) = $Z / \sqrt{N}$, where N is the number of observations (Field, 2013).

One child took part only in the first pre-test, which was used as the child's initial performance level. Another child in the NC-group (among the poorest performers) missed the dot counting fluency test in both pre-tests. In this case, the missing initial level value was calculated by subtracting the mean gain (that the poorest-performing NC-group received in dot counting fluency) from the case's own reaction time measured in the third assessment.

Response to computerized early number intervention

The participants reached both the targeted number of practice sessions and the total time of practice as instructed, with the exception of two children. One child did not reach the amount of sessions required, and the other one did not reach the minutes that were instructed even though the amount of sessions was fulfilled. However, there was no significant correlation (Spearman's rho) between the number of intervention sessions or training times and the initial-level skill variables or the intervention gain scores in either Examination 1 or Examination 2.

Results

Initial performance levels

To determine the initial performance levels, the average scores of the two pre-tests were calculated for Number Sets Test (A), story problems, RAN of objects, dot counting fluency, verbal counting, and basic addition. PPVT was measured once in the first pre-test and was used as an initial level in vocabulary. In Examinations 1 and 2, there were no significant group-level differences between the two conditions in any of the background (PPVT; Number Sets Test; story problems; RAN) or outcome variables (dot counting; verbal counting; basic addition) before the intervention. The descriptive statistics for the background variables with the normative values (reference data collected outside the current study) concerning the whole sample (Examination 1) are shown in Table 1, and those concerning the sub-sample of the poorest performers (Examination 2) are shown in Table 2.

Insert Table 1 here.

Insert Table 2 here.

Intervention effects in Examination 1

After the first intervention period (T1–T2; Table 3), the Wilcoxon signed-rank test showed a significant within-group improvement for dot counting fluency, verbal counting, and basic addition skills in the NC-group ($Z = -3.62$, $p < .001$; $Z = -1.84$, $p = .040$; $Z = -2.91$, $p = .001$, respectively), and for dot counting fluency and verbal counting in the NS-group ($Z = -3.00$, $p = .001$; $Z = -1.97$, $p = .026$, respectively). The Mann-Whitney U -test showed a statistically significant between-group difference for gain scores of dot counting fluency (G1: $U = 68.0$, $Z = -2.45$, $p = .007$, $r = .43$; Table 3), indicating greater improvement in the NC-group. After the second intervention period during which both groups received basic addition training (T1–T3; Table 3), the Wilcoxon signed-rank test showed a significant within-group improvement in basic addition in both the NC- and the NS-groups ($Z = -3.30$, $p < .001$; $Z = -2.40$, $p = .006$, respectively). The effect remained stable in the follow-up assessment in both groups (T1–T4: $Z = -3.41$, $p < .001$; $Z = -2.58$, $p = .003$, respectively; Table 3). There were no differences in gain scores of basic addition between the two conditions either after the second intervention period or the follow-up assessment. Both the raw scores of each assessment point and the intervention gain scores in both intervention groups of Examination 1 are shown in Table 3.

Insert Table 3 here.

Intervention effects in Examination 2

After the first intervention period (T1–T2; Table 4), the Wilcoxon signed-rank test showed a significant within-group improvement in dot counting fluency among the poorest performers of the NC-group ($Z = -2.52$, $p = .004$), and in dot counting fluency and verbal counting in the NS-group ($Z = -1.78$, $p = .047$; $Z = -2.23$, $p = .016$, respectively). The Mann-Whitney U -test showed a close to significant between-group difference for gain scores of dot counting fluency (G1: $U = 11.0$, $Z = -1.68$, $p = .054$, $r = .45$; Table 4), favoring the NC-group. After the second intervention period (T1–T3; Table 4), the Wilcoxon signed-rank test showed a significant within-group improvement in basic addition only in the NC-group ($Z = -2.20$, $p = .016$). The effect remained stable in the follow-up assessment (T1–T4: $Z = -2.39$, $p = .008$;

Table 4). The Mann-Whitney U -test showed a statistically significant between-group difference for gain scores of basic addition after the follow-up assessment (G3: $U = 9.0$, $Z = -1.98$, $p = .046$, $r = .53$; Table 4), indicating larger improvements in the NC-group. Both the raw scores of each assessment point and the intervention gain scores in both intervention groups of Examination 2 are shown in Table 4.

Insert Table 4 here.

Discussion

This study aimed to examine to what extent an intensive, adaptive, and theory-based CAI (here GraphoGame Math) produces immediate, condition-specific effects for dot counting fluency when training number concept skills, or verbal counting skills when training number sequence knowledge. Furthermore, the study aimed to evaluate how these two different intervention conditions would support basic addition skills that were practiced during the second intervention cycle in both intervention conditions. The improvements were assessed first within the group of low-performing children (identified by teachers), and second, within the group of poorest-performing children (identified by teachers and individual assessments). All participants were 6–7-year-olds.

After the first three-week intervention period, the results indicated that the low-performing children improved significantly in both dot counting fluency and verbal counting accuracy despite the intervention conditions. The NC-group also showed gains in basic addition even though it was not yet practiced. The only significant between-group difference in gain scores was seen in dot counting fluency, where the NC-group gained more than the NS-group. In contrast, among the children most at risk for MD (performance below the 10th percentile as compared to age-peers; 5 % of the original sample), the first three-week intervention resulted in condition-specific effects showing that the NC-group improved in fluency in dot counting, and the NS-group improved in verbal counting accuracy. Even though the NS-group also improved in dot counting, the gain in the NC-group was larger than in the NS-group (Table 4).

We were also interested in seeing whether there are differences between the number concept and number sequence knowledge training groups regarding to the improvement of basic addition skills. After the second three-week intervention period, during which both groups received basic addition intervention, both the low-performing NC- and the NS-groups improved similarly in their basic addition skills. The effect remained stable over the follow-up period, and there were no group-differences. This finding is comparable with recent CAI studies. For example, Baroody et al. (2013, 2012) found improvement in basic arithmetic skills in children at risk for MD when quantity discrimination, numeral recognition, verbal counting, transcoding between different numerical representations, and basic addition trainings were mixed.

Interestingly, among the poorest-performing children, number concept training provided transfer effect to basic addition skills but number sequence training did not. Moreover, after the second three-week intervention period, improvement in basic addition was seen only in the NC-group, even though both groups received similar training in basic addition. This within-group effect remained stable over the follow-up, and the NC-group showed larger improvement than the NS-group in gain scores of basic addition. Because there were no between-group differences in the initial level of vocabulary, rapid naming, or numerical skills, or in sessions/minutes related to interventions, and the intervention periods were relatively short, these general factors or differences in curricular activities do not seem to explain the observed differences between the two intervention groups.

The findings of this study show that the theory-based practice targeted to number concept or number sequence skills can produce effects within rather short training periods. The intervention was individually adaptive, explicit, and repetitive, which have all been found to be vital for supporting children having (or at risk for) MD (e.g., Fuchs et al., 2012). Furthermore, explicit corrective, continuous, task-focused, and scaffolding feedback has been suggested as important elements of support (Baker et al., 2002; Gersten et al., 2009; Hunt et al., 2015; Shute, 2008).

The results are in line with the access deficit hypothesis, which states that children with MD face difficulties accessing the numerical meaning from number symbols, which *per se* would explain deficits in learning arithmetic (De Smedt & Gilmore, 2011; Rousselle & Noël, 2007). The NC-group benefited more from the training than the NS-group did, with regard to basic addition skills. The NC-component of the training required symbolic number processing skills in matching and discriminating the numerical values of (as well as between) the alternatives (i.e., potential correct stimuli) since the stimuli were mixed to be either non-symbolic or symbolic within most of the trials (73 % of the total amount of trials). Respectively, NS-component required completing and continuing the number sequences forward and backward without similar requirement for symbolic number processing. Non-symbolic and symbolic stimuli were rarely presented simultaneously (only in 18 % of the total amount of trials). Furthermore, the NC-component of the training (e.g., comparison and mapping) with the number processing requirement might have supported understanding of the exact relations between numerals. Therefore, the NC-practice might have created a stronger basis for the practice of composing, decomposing, and simple addition during the second intervention period as compared to the NS-group practicing number sequence skills (cf. Geary, 2013; Krajewski & Schneider, 2009).

Even though the study design met many of the criteria often listed for qualified intervention studies (cf. Cook & Cook, 2013), the sample sizes were rather small. It would also be important to control more extensively for domain-general deficits (cf. Kaufmann et al., 2013) in order to better identify children with specific problems, or to be able to interpret individual variations in responsiveness to intervention. Although we assessed the initial level of vocabulary (PPVT) and processing speed (RAN) in our study, we did not include an assessment of working memory, which *per se* is strongly associated with arithmetic skills (Geary, 2011). Since there seem to be qualitative differences in early number skills development between low-performing and the poorest-performing children (Geary, 2011, 2013; Geary et al., 2007; Kaufmann et al., 2013; Morgan et al., 2009; Murphy et al., 2007), there is a need for further knowledge on the potential differences in response to interventions among these two skill-level groups. The replications concerning this should be done with larger sample sizes.

Implications for practice

CAI could offer cost-effective support allowing high training intensity, repetition of basic number concepts and skills, as well as continuous feedback and individualization of the intervention according to the performance-level. The findings of this study indicated that the low-performing children (12 % of the original sample) showed generalized positive intervention effects regardless of the content that was intensively practiced. In contrast, the poorest-performing children (performance below the 10th percentile; 5 % of the original sample) improved mainly in the skills that were effectively intervened. Furthermore, the observed effects of exact symbolic processing requirement in intervention as a prerequisite for significant gain in basic addition in the NC-group offer a theoretically and practically relevant input toward developing early interventions for children most at risk for MD (cf. access deficit hypothesis; De Smedt & Gilmore, 2011; Rousselle & Noël, 2007). In sum, with theory-based training content, CAI could be a meaningful method for identifying needs, monitoring progress during practice, and preventing cumulative deficits in learning.

References

Ahonen, T., Tuovinen, S., & Leppäsaari, T (2006). *Nopean sarjallisen nimeämisen testi* (2nd ed.). Jyväskylä, Finland: Niilo Mäki Institute.

Baker, S., Gersten, R., & Lee, D-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103*(1), 51−73. Retrieved from http://www.jstor.org/stable/1002308

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education, 18*(2), 141−157. http://www.jstor.org/stable/749248

Baroody, A. J., Bajva, N. P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews 15*, 69–79. doi: 10.1002/ddrr.45

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering at-risk kindergarten children's number sense. *Cognition and Instruction, 30*(4), 437–470. doi: 10.1080/07370008.2012.720152

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal, 1*, 1−41. doi: 10.3102/0002831212473349

Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D., (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *Journal of Experimental Child Psychology, 117*, 12–28. doi: 10.1016/j.jecp.2013.08.010

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46*(1), 3–18. doi: 10.1111/j.1469-7610.2005.00374.x

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science, 332*, 1049−1053. doi: 10.1126/science.1201536

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

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology, 33*(3), 300–317. doi: 10.1080/87565640801982387

Cook, B. G., & Cook, S. C. (2013). Unraveling evidence-based practices in special education. *The Journal of Special Education, 47*(2), 71–82. doi: 10.1177/0022466911420877

Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* Oxford, UK: University Press.

Dencla, M. B., & Rudel, R. (1974). Rapid “automatized” naming of pictured objects, colours, letters and numbers by normal children. *Cortex, 10*(2), 186−202.

De Smedt, B., & Gilmore C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*, 278–292. doi:10.1016/j.jecp.2010.09.003

De Smedt, B., Noël, M-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children’s mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*, 48–55. doi: 10.1016/j.tine.2013.06.001

Dowker, A. (2009). Use of derived fact strategies by children with mathematical difficulties. *Cognitive Development, 24*(4), 401−410. doi: 10.1016/j.cogdev.2009.09.005

Dowker, A., & Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *The British Psychological Society, Monograph Series 2*(7), 65–81. doi:10.1348/97818543370009X12583699332492

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test − Revised.* Circle Pines, MN: American Guidance Service.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE Publications Ltd.

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory & Cognition, 38*(7), 951−961. doi:10.3758/MC.38.7.951

Finnish National Board of Education. (2010). Core curriculum for preschool education in Finland. Downloadable (in English) from http://www.oph.fi/download/153504_national_core_curriculum_for_pre-primary_education_2010.pdf

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*(3), 257–269. doi: 10.1177/0022219412442167

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31*(2), 79–92. doi: 10.2307/20528819

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics, 32*(3), 250–263. doi:10.1097/DBP.0b013e318209edef

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23–27. doi: 10.1177/0963721412469398

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment, 27*(3), 265–279. doi: 10.1177/0734282908330592 [Finnish version: Numerosetit (Koponen, T., Salminen, J., Räsänen, P., & Manninen, J. 2008.) Niilo Mäki Institute. Unpublished.]

Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities, 45*(4), 291–307. doi: 10.1177/0022219410392046

Geary, D., C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343−1359. doi: 10.1111/j.1467-8624.2007.01069.x

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. R. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202–1242. doi: 10.3102/0034654309334431

Hunt, J. H., Valentine, C., Bryant, D. P., Pfannenstiel, K. H., & Bryant, B. R. (2015). Supplemental Mathematics Intervention: How and why special educators intensify intervention for students with learning disabilities. *Remedial and Special Education.* doi: 10.1177/0741932515597293

Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Göbel, S. M., & Grabner, R. H., … Nuerk, H-C. (2013). Dyscalculia from a developmental and differential perspective. *Frontiers in Psychology, 4*(516), 1−5. doi: 10.3389/fpsyg.2013.00516

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*(1), 162–175. doi: 10.1037/a0029285

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*, 516–531. doi:10.1016/j.jecp.2009.03.009

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönman, C., Plangger, F., … von Aster, M. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage, 57*(3), 782–795. doi: 10.1016/j.neuroimage.2011.01.070.

Käser, T., Baschera, G-M., Kohn, J., Kucian, K., Richtmann, V., Grond, U., … von Aster, M. (2013). Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Frontiers in Psychology, 4*, Advance online publication. doi: 10.3389/fpsyg.2013.00489.

Lyytinen, H., Ahonen, T., Eklund, K., Guttorm, T., Kulju, P., Laakso, M-L., … Torppa, M. (2004). Early development of children at familial risk for dyslexia − Follow-up from birth to school age. *Dyslexia, 3*, 146–178. doi:10.1002/dys.274

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*(1−2), 99−113. doi: 10.1023/B:TRUC.0000021811.66966.1d

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*(4), 306−321. doi: 10.1177/0022219408331037

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*(5), 458−478. doi: 10.1177/00222194070400050901

Rousselle, L. & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*, 361–395. doi:10.1016/j.cognition.2006.01.005

Salaschek, M., Zeuch, N., & Souvignier, E. (2014). Mathematics growth trajectories in first grade: Cumulative vs. compensatory patterns and the role of number sense. *Learning and Individual Differences, 35*, 103–112. doi: 10-1016/j-lindif.2014.06.009

Salminen, J. & Koponen, T. (2010). The Rapid Naming of Small Quantities Test. Niilo Mäki Institute. Unpublished.

Salminen, J., Koponen, T., Leskinen, M., Poikkeus, A-M., & Aro, M. (in press). Individual variance in responsiveness to early computerized mathematics intervention. Learning and Individual Differences. doi: 10.1016/j.lindif.2015.09.002

Salminen, J., Räsänen, P., Koponen, T., & Aunio, P. (2008). The Non-Symbolic Story Problems Test. niilo Mäki Institute. Unpublished.

Salonen, P., Lepola, J., Vauras, M., Rauhanummi, T., Lehtinen, E., & Kinnunen, R. (1994). *Diagnostiset testit 3. Motivaatio, Metakognitio ja Matematiikka* [Diagnostic Tests 3: Motivation, metacognition and mathematics]. Turku, Finland: Centre for Learning Research, University of Turku.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (in press). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. doi: 10.3102/0034654307313795

Wylie, J., Jordan, J. A., & Mulhern, G. (2012). Strategic development in exact calculation: Group and individual differences in four achievement subtypes. *Journal of Experimental Child Psychology, 113*(1), 112–130. doi: 10.1016/j.jecp.2012.05.005

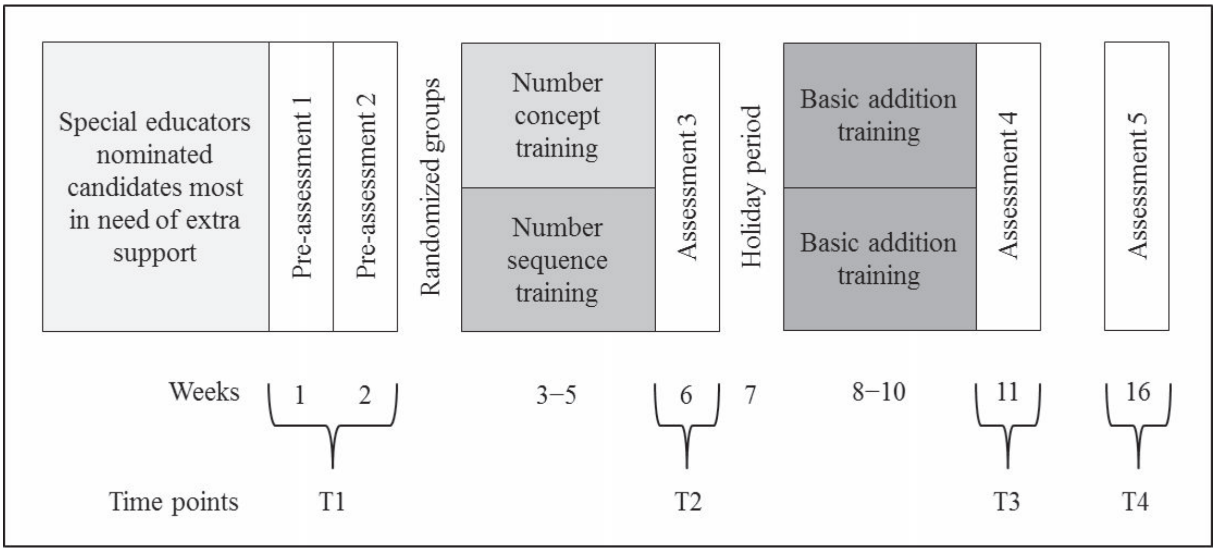
Special educators nominated candidates most in need of extra support
Pre-assessment 1
Pre-assessment 2
Randomized groups
Number concept training
Number sequence training
Assessment 3
Holiday period
Basic addition training
Basic addition training
Assessment 4
Assessment 5
Weeks
1
2
3–5
6
7
8–10
11
16
Time points
T1
T2
T3
T4

Figure 1 title:

The study design

Figure 1 note:

Figure 1. Time points = Assessment points. The average of the two pre-assessments was used as the initial performance-level in all skill variables (Time point 1) except the PPVT which was conducted once (in the first pre-test).

Table 1

Group Means, Standard Deviations, and Medians in Background Variables in the Examination 1

			Intervention condition			
	REF		NC (*n* = 17)		NS (*n* = 16)	
Variable	*M (SD)*	TP	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
PPVT (30)	19.82 (3.38)	T1[a]	16.24 (2.86)	16.00	15.69 (3.44)	15.50
Number Sets A (18)		T1	3.12 (2.79)	3.00	3.81 (2.12)	3.50
Story problems (6)		T1	5.00 (1.09)	5.00	5.16 (1.00)	5.50
Rapid naming (s)	72.60 (15.30)	T1	87.00 (21.88)	80.00	88.59 (23.38)	84.00

Note. REF = Reference data (population-based normative data; collected in separate studies administered within two month period of the two pre-assessment time points of the current study); NC = Number concept training first, then basic addition; NS = Number sequence training first, then basic addition. PPVT = Peabody picture vocabulary test. TP = Time point. T1 = initial level = the average of the two pre-assessment scores.

[a] Measured only in the first pre-assessment.

Table 2

Group Means, Standard Deviations, and Medians in Background Variables in the Examination 2

			Intervention condition			
	REF		NC ($n = 8$)		NS ($n = 6$)	
Variable	*M (SD)*	TP	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
PPVT (30)	19.82 (3.38)	T1[a]	16.13 (3.36)	15.50	15.33 (2.94)	16.00
Number Sets A (18)		T1	2.69 (3.15)	2.50	3.17 (1.72)	2.50
Story problems (6)		T1	4.38 (1.16)	4.75	4.75 (1.08)	4.50
Rapid naming (s)	72.60 (15.30)	T1	95.13 (23.44)	91.25	93.42 (17.97)	91.50

Note. REF = Reference data (population-based normative data; collected in separate studies administered within two month period of the two pre-assessment time points of the current study); NC = Number concept training first, then basic addition; NS = Number sequence training first, then basic addition. PPVT = Peabody picture vocabulary test. TP = Time point. T1 = initial level = the average of the two pre-assessment scores.

[a] Measured only in the first pre-assessment.

Table 3

Group Means, Standard Deviations, and Medians in Four Assessment Points, and Intervention Gain Scores in Outcome Variables in the Examination 1

			Intervention condition			
	REF		NC (*n* = 17)		NS (*n* = 16)	
Variable	*M (SD)*	TP	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Dot counting (s)	1.89 (0.55)	T1	2.77 (0.66)	2.93	2.49 (0.84)	2.33
		T2	2.04 (0.60)	1.92***	2.12 (0.72)	1.88**
		G1[c]	-0.73 (0.38)	-0.71	-0.37 (0.40)	-0.36
		T3	2.00 (0.57)	2.00	2.31 (0.96)	1.93
		G2	-0.77 (0.50)	-0.75	-0.17 (0.54)	-0.23
		T4	1.94 (0.67)	1.75	2.00 (0.74)	1.75
		G3	-0.83 (0.48)	-0.89	-0.49 (0.43)	-0.42
Verbal counting	< 20[b]	T1	6.62 (3.46)	6.00	6.59 (3.05)	5.75
		T2	7.41 (3.84)	8.00*	7.38 (2.85)	6.00*
		G1	0.79 (1.52)	0.00	0.78 (1.49)	0.50
		T3	8.18 (3.38)	9.00	7.94 (3.02)	7.00
		G2	1.56 (1.45)	1.00	1.34 (1.58)	1.50
		T4	8.41 (3.68)	10.00	7.94 (3.36)	8.00
		G3	1.79 (1.82)	2.00	1.34 (2.01)	1.00
Basic addition[a]	18.36 (10.42)	T1	5.79 (6.44)	4.00	6.75 (7.60)	4.50
		T2	8.35 (6.96)	9.00**	7.44 (8.50)	3.50
		G1	2.56 (3.18)	3.00	0.69 (5.96)	0.00
		T3	12.47 (8.50)	11.00***	10.81 (10.78)	8.00**
		G2	6.68 (6.45)	5.00	4.06 (7.12)	1.00
		T4	13.00 (8.62)	12.00***	12.38 (12.26)	12.00**
		G3	7.21 (6.15)	7.00	5.63 (7.82)	4.00

Note. REF = Reference data (population-based normative data; collected in separate studies administered within two month period of the two pre-assessment time points of the current study); NC = Number concept training first, then basic addition; NS = Number sequence training first, then basic addition. TP = Time point. T1 = initial level = the average of the two pre-assessments; T2 = after first intervention; G1 = gain scores for the first intervention (T2-T1); T3 = after second intervention; G2 = gain scores for both interventions (T3-T1); T4 = delayed assessment; G3 = gain scores over the study (T4-T1).

[a] Time limited task of 3 minutes. [b] Failure in “Counting on up to 20” sub-task was used as a criterion for belonging to the intervention group (performance below the 10th percentile) since 89,8 % of Finnish kindergarteners can count verbally until 20 at the same curriculum time point (Jan-Feb). [c] NC > NS** (Mann-Whitney U-test; Wilcoxon *ES* (*r*) = .43).

* *p* < .05. ** *p* < .01. *** *p* < .001.

Table 4

Group Means, Standard Deviations, and Medians in Four Assessment Points, and Intervention Gain Scores in Outcome Variables in the Examination 2

			Intervention condition			
	REF		NC (*n* = 8)		NS (*n* = 6)	
Variable	*M (SD)*	TP	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Dot counting (s)	1.89 (0.55)	T1	3.32 (0.44)	3.18	3.08 (1.10)	2.93
		T2	2.46 (0.59)	2.49**	2.62 (0.94)	2.78*
		G1[c]	-0.86 (0.45)	-0.88	-0.46 (0.51)	-0.45
		T3	2.32 (0.49)	2.32	3.07 (0.94)	3.01
		G2	-1.00 (0.43)	-0.91	-0.02 (0.35)	-0.11
		T4	2.35 (0.68)	2.30	2.56 (0.86)	2.57
		G3	-0.97 (0.62)	-1.05	-0.52 (0.32)	-0.41
Verbal counting	< 20[b]	T1	4.31 (2.43)	4.50	4.58 (1.80)	4.50
		T2	5.13 (3.56)	5.00	5.67 (1.75)	5.00**
		G1	0.81 (1.77)	0.00	1.08 (0.66)	1.00
		T3	6.63 (3.25)	7.00	6.83 (1.72)	7.00
		G2	2.31 (1.49)	2.50	2.25 (0.99)	2.00
		T4	6.38 (3.82)	6.00	6.17 (3.76)	6.50
		G3	2.06 (2.18)	2.25	1.58 (2.94)	1.50
Basic addition[a]	18.36 (10.42)	T1	1.38 (2.07)	0.00	2.83 (2.93)	2.50
		T2	4.50 (5.07)	3.00	2.50 (2.35)	2.50
		G1	3.13 (4.19)	2.00	-0.33 (1.37)	-0.50
		T3	8.63 (8.25)	7.50**	3.67 (3.72)	3.00
		G2	7.25 (8.61)	5.50	0.83 (2.99)	0.00
		T4	8.88 (7.34)	9.00**	4.17 (5.00)	3.00
		G3[d]	7.50 (7.35)	7.00	1.33 (3.98)	0.00

Note. REF = Reference data (population-based normative data; collected in separate studies administered within two month period of the two pre-assessment time points of the current study); NC = Number concept training first, basic addition after; NS = Number sequence training first, basic addition after. TP = Time point. T1 = initial level = the average of the two pre-assessments; T2 = after first intervention; G1 = gain scores for the first intervention (T2-T1); T3 = after second intervention; G2 = gain scores for both interventions (T3-T1); T4 = delayed assessment; G3 = gain scores over the study (T4-T1).

[a] Time limited task of 3 minutes. [b] Failure in “Counting on up to 20” sub-task was used as a criterion for belonging to the intervention group (performance below the 10th percentile) since 89,8 % of Finnish kindergarteners can count verbally until 20 at the same curriculum time point (Jan-Feb). [c] NC > NS^{ns} (Mann-Whitney U-test showed close to significant between-group difference, *p* = .054; Wilcoxon *ES* (*r*) = .45). [d] NC > NS* (Mann-Whitney U-test; Wilcoxon *ES* (*r*) = .53).

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