Tampere University

DARANEE LEHTONEN

'Now I Get It!'

Developing a Real-World Design Solution for Understanding Equation-Solving Concepts

Tampere University Dissertations 538

Tampere University Dissertations 538

DARANEE LEHTONEN

‘Now I Get It!’

Developing a Real-World Design Solution for Understanding Equation-Solving Concepts

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Education and Culture
of Tampere University,
for public discussion in the Väinö Linna auditorium
of the Linna building, Kalevantie 5, Tampere,
on 11 February 2022, at 13 o’clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Education and Culture
Finland

Responsible supervisor and Custos	Docent Jorma Joutsenlahti Tampere University Finland	
Supervisor	Docent Päivi Perkkilä University of Jyväskylä Finland	
Pre-examiners	Docent Anu Laine University of Helsinki Finland	Associate Professor Glenn Smith University of South Florida The United States
Opponent	Docent Lasse Eronen University of Eastern Finland Finland	

Cover design: Roihu Inc.

ISBN 978-952-03-2249-6 (print)
ISBN 978-952-03-2250-2 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-2250-2

PunaMusta Oy – Yliopistopaino
Joensuu 2022

In loving memory of my father, Direk Anuyotha, who would have been proud.

ACKNOWLEDGEMENTS

Many people supported me during these past 7 years. I would like to thank the following people, without whom I would not have possibly completed this dissertation and would not have made it through this long, challenging journey.

My journey in educational sciences would not have begun without the Kuulumisia complementary education programme at the Faculty of Education and Culture, Tampere University. I am grateful to the Ministry of Education and Culture and the Finnish National Agency for Education for funding this education programme. I wish to thank everyone who made the programme possible, especially Outi Stüber, Tuula Pantzar, and Dr. Juha Merta. Thanks also to all my Kuulumisia fellows for entering the Finnish education world with me.

I would like to express my deepest gratitude to my supervisors, Docents Jorma Joutsenlahti and Päivi Perkkilä. Jorma, thank you for everything you did for me. Your belief in me and my research proposal allowed me to begin this scholarly journey. Without your invaluable guidance and endless encouragement, I would not have become the person I am today. You patiently listened and always promptly responded to my endless e-mails, WhatsApp messages, and phone calls. I promise there will be no more late night and weekend disturbances from now on. You also kindly provided me job opportunities, which financially secured my doctoral study. Päivi, thank you for not hesitating to be my supervisor in the middle of my journey. Your unparalleled support and thoughtful advice helped me to improve significantly not only my work but also my development as a researcher. Your countless reassurance also made me feel confident in my abilities to complete this dissertation.

I would like to express my appreciation to the pre-examiners of my dissertation, Associate Professor Glenn Smith and Docent Anu Laine, for their constructive comments and suggestions. I am also grateful to Docent Lasse Eronen for agreeing to be my opponent in my doctoral defence.

Many thanks to everyone at our doctoral seminars for their comments, advice, and encouragement throughout these years. Docent Pirjo Kulju, thank you for your kind support and warm hugs when I needed them. Adjunct Professor Reijo Kupiainen, thank you for your eye-opening comments during my years at EDU and your help with my new journey at ITC. Dr. Carita Kiili, thank you for your generous support from the moment we met. I had the great pleasure of sharing the doctoral journey with all my fellow doctoral researchers. Helen Alfaro Viquez, your work ethics urged me to work harder. Elina Viro, our WhatsApp chats comforted me in knowing I was not on this journey alone.

I was fortunate to be part of the Virta Building fifth floor community, where I always received intellectual and mental support. Jaska Poranen, thank you for getting me back to the amazing mathematics world. Dr. Ville Mankki, thank you for your generous support, which helped me through this doctoral journey. Vesna Holubek, thank you for our cheerful coffee break moments. Dr. Tahani Aldahdouh, thank you for your kind advice concerning statistical analysis. I must also thank Docent Kaisu Rättyä, Docent Jyri Lindén, Docent Vesa Korhonen, and Professor Kristian Kiili for their valuable comments and suggestions that helped improve my work.

Special thanks to Dr. Anne Jyrkiäinen and Lucas Machado for cowriting our publications. I gratefully acknowledge the contributions of Lucas and his team for software development and technology implementation. Thanks also go to Kari Kouhia, who kindly helped me with prototyping. I greatly appreciate the teachers and the students who participated in the research and invited me into their classrooms. I am also grateful for the financial support from the Finnish Cultural Foundation, Häme Regional Grant and the City of Tampere's Grant Committee.

Thanks to all my friends, particularly my Satit KKU classmates and Dr. Patarawan Chaowasakoo, for long-distance friendship and support. My wholehearted thanks go to my family in Thailand for their unconditional love. My parents, Direk and Warunee, thank you for teaching me the value of education. Without your wisdom, I would not have pursued my doctoral journey. To my siblings, Kop, Guide, and Kip; my bother-in-law, Ko; my sister-in-law, Ou; and my niece, Panwa, you have always been a part of my life, no matter where I am. Thanks also to my husband's family, my family in Finland for your heartfelt support.

Finally, my biggest thanks go to my husband, Tapio, and my kids Touko, Aamu, and Tomi; you earned this degree right along with me. To my beloved kids, I am sorry for being distracted by my study during all these years. I will now spend more quality time with you, I promise! And to my dear Tapio, you have been amazing. You have not only taken care of the housework, the kids, and me but also helped me with my dissertation. What would I do without you?

On a winter day in Hämeenlinna, December 9, 2021

Daranee Lehtonen

ABSTRACT

Strong conceptual understanding contributes to mathematics learning. Manipulatives (i.e., hands-on learning tools that allow for mathematical concept exploration through different senses) can facilitate students' understanding of mathematical concepts when used meaningfully. However, a body of research has demonstrated that although teachers have considered manipulatives pedagogically beneficial, when it comes to everyday classroom practice, they often prefer traditional teacher-centred and paper-and-pencil instruction.

This doctoral research aims to develop a manipulative and its appropriate use to promote not only students' understanding of mathematical concepts, but also classroom adoption. Solving one-variable linear equations in primary school classrooms was used as a case study. An educational design research (EDR) approach was used throughout three phases of a 6-year enquiry: initial research, concept development, and design development. Phase 1 (initial research) was undertaken to gain a theoretical and contextual understanding and investigate existing manipulatives. In Phase 2 (concept development), four manipulative concepts were generated based on the Phase 1 findings. Each concept was then evaluated in terms of its pedagogical benefits and compatibility with school and classroom practice. During Phase 3 (design development), informed by the Phase 2 findings, a design solution (i.e., a tangible manipulative allowing physical input and providing digital output, student worksheets, teacher guides, and class activities) was developed. The developed design solution was then implemented and evaluated in classrooms.

Empirical research was conducted in Finnish comprehensive schools. Altogether, 18 teachers, 98 primary school students, and 65 lower secondary school students took part in different phases of the research. The data were collected using mixed methods, including class interventions, paper-based tests, thinking aloud, questionnaires, and interviews. Qualitative and quantitative data collected from various methods and data sources were simultaneously analysed and then compared and combined to holistically understand the research results.

Together, multiple iterations (of investigation, design, and assessment) resulted in practical and theoretical outcomes. The research-based design solution, which promotes students' understanding of equation-solving concepts and classroom practice, is the practical outcome of this research to directly improve educational practice. Additionally, the research contributes to three types of theories: domain theories, design frameworks, and design methodologies.

The first theoretical outcome is a *domain theory* yielding two types of knowledge, that is, context and outcomes knowledge. The *context* knowledge describes the challenges and opportunities of using manipulatives in mathematics classrooms, as well as strengths and limitations of existing manipulatives. The *outcomes* knowledge describes outcomes of implementing the design solution: the developed tangible manipulative accompanied by the instructional materials enhanced students' understanding of equation-solving concepts through discovery learning, social interaction, and multimodal expression of mathematical thinking; the manipulative is likely to be adopted in the classroom because of its pedagogical benefits and compatibility with school and classroom practice. The second theoretical outcome is a *design framework* for real-world educational technologies. Content, pedagogy, practice, and technology should be taken into consideration when designing real-world educational technologies to ensure their educational benefits, utilisation, adoption, and feasibility. The third theoretical outcome is a *design methodology* built on firsthand experience from undertaking this EDR. The guidelines for conducting EDR guides how to embrace opportunities and overcome challenges that may emerge.

This research contributes to a link between research and practice in mathematics education. It provides researchers with knowledge of how multimodal interaction with manipulatives enhances mathematics learning and guidelines for conducting EDR. It guides educational designers to take various aspects into consideration when designing educational technologies to improve real-world practice. Moreover, this research also has practical implications. First, it encourages teacher educators to prepare pre- and in-service teachers for successful incorporation of manipulatives in their mathematics classrooms. Second, it guides practitioners on how to support their students to benefit from manipulatives. Third, it urges schools to support the acquisition and utilisation of manipulatives. Finally, it calls on school curricula to encourage the use of manipulatives in the mathematics classroom to promote students' conceptual understanding.

Keywords: linear equation solving, conceptual understanding, manipulatives, educational technologies, educational design research

CONTENTS

ABBREVIATIONS

ACARA	Australian Curriculum, Assessment and Reporting Authority
ALLEA	All European Academies
CCSSO	Council of Chief State School Officers
DBRC	Design-Based Research Collective
ECTS	European Credit Transfer and Accumulation System
EDR	Educational design research
EDUFI	Finnish National Agency for Education
ICT	Information and communication technology
MAL	Multimodal Algebra Learning
NCC	National core curriculum
NCTM	National Council of Teachers of Mathematics
NGA	National Governors Association Center for Best Practices
NLVM	National Library of Virtual Manipulatives
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
PM	Physical manipulative
RQ	Research question
TENK	Finnish National Board on Research Integrity
TM	Tangible manipulative
VM	Virtual manipulative

ORIGINAL PUBLICATIONS

Publication I	Lehtonen, D., Jyrkiäinen, A., & Joutsenlahti, J. (2019). A systematic review of educational design research in Finnish doctoral dissertations on mathematics, science, and technology education. *International Journal on Math, Science and Technology Education, 7*(3), 140–165. https://doi.org/10.31129/LUMAT.7.3.399
Publication II	Lehtonen, D., & Joutsenlahti, J. (2017). Using manipulatives for teaching equation concepts in languaging-based classrooms. In N. Pyyry, L. Tainio, K. Juuti, R. Vasquez & M. Paananen (Eds.), *Changing subjects, changing pedagogies: Diversities in school and education* (pp.164–185). Finnish Research Association for Subject Didactics.
Publication III	Lehtonen, D., Machado, L., Joutsenlahti, J., & Perkkilä, P. (2020). The potentials of tangible technologies for learning linear equations. *Multimodal Technologies and Interaction, 4*(4), Article 77. https://doi.org/10.3390/mti4040077
Publication IV	Lehtonen, D. (2021). Constructing a design framework and design methodology from educational design research on real-world educational technology development. *Educational Design Research, 5*(2), Article 38. https://doi.org/10.15460/eder.5.2.1680

AUTHOR'S CONTRIBUTION

I am the first and corresponding author of Publications I–III and the sole author of Publication IV.

I was responsible for the research design, data collection and analysis, results interpretation, discussion and conclusions, concept and design development of the educational solution, and writing of all manuscripts.

Docent Jorma Joutsenlahti supervised the doctoral research, contributed to the data analysis of Publication I, and commented on the manuscripts of all publications. Docent Päivi Perkkilä supervised the doctoral research and commented on the manuscripts of Publications III and IV. Dr. Anne Jyrkiäinen contributed to the data analysis of Publication I and commented on its manuscript. Lucas Machado contributed to the writing, particularly related to technology implementation and software development, and visualisation of Publication III. He and his development team (Fouzia Khan, Juho Korkala, Roni Perälä, Niko Sainio, and Krishna Bagale) were responsible for the technology implementation and software development of the educational solution under the supervision of Pekka Mäkiaho.

1 INTRODUCTION

1.1 Rationale

The current Finnish national core curriculum (NCC) for basic education 2014 was implemented in August 2016 in primary and lower secondary schools across the country (Finnish National Agency for Education [EDUFI], 2016). On a general level, basic education aims to develop students' transversal competences, including thinking and learning to learn, multiliteracy, and information and communication technology (ICT) competence in all grades and school subjects. Students are encouraged to learn through active learning, exploration, and discovery; collaboration and social interaction; diverse modes of meaning making and communication; and the use of ICT. Moreover, basic education seeks to promote equality by recognising students' diversity, including personal needs and developmental differences, and providing differentiation of instruction appropriately.

EDUFI conducted a longitudinal study on the mathematics learning outcomes of 3,502 Finnish comprehensive school students during their third, sixth, and ninth grades between 2005 and 2012. Although students' learning outcomes increase during the school years, the assessments indicate demands for developing students' conceptual understanding to promote their positive attitudes towards self-efficacy (Tuohilampi & Hannula, 2013) and provide a good foundation for their study in the upper grades (Joutsenlahti & Vainionpää, 2010).

Traditional school mathematics typically emphasises developing students' arithmetic skills (e.g., Groves, 2012; Kilpatrick et al., 2001). It has been shown that procedural knowledge and computational skills are not sufficient for successful mathematics education (Kilpatrick et al., 2001; Schoenfeld, 2007), and a sole emphasis on procedures and computation development has gradually shifted to also including other important mathematical proficiency.

It is now recognised that conceptual understanding is one of the key mathematical proficiencies. Previous research has shown that strong conceptual understanding has a number of advantages for mathematics learning, whereas inadequate conceptual understanding can hinder students' mathematics learning and performance (e.g., Andamon & Tan, 2018; Kilpatrick et al., 2001). The national mathematics curricula and standards of various countries, including Australia, Finland, and the United States, have recently paid more attention to students' conceptual understanding, instead of focusing solely on procedural knowledge and skills (see Australian Curriculum, Assessment and Reporting Authority [ACARA], 2016; EDUFI, 2016; National Governors Association Center for Best Practices [NGA] & Council of Chief State School Officers [CCSSO], 2010).

Mathematics textbooks have traditionally been used as the main instructional materials in schools around the world (see e.g., Alshwaikh & Morgan, 2013; Lepik et al., 2015; Neber, 2012). In Finland, it is common for each student to silently do exercises in the textbook during mathematics lessons (Joutsenlahti & Kulju, 2017), but recently, there has been a growing concern that students' low performance in mathematics could be partially a result of textbook-emphasised instruction (Alshwaikh & Morgan, 2013; Joutsenlahti & Vainionpää, 2010). The Finnish NCC 2014 (EDUFI, 2016) underlines the significant role of ICT and concrete hands-on learning tools in mathematics across all grade levels.

A variety of technologies are being increasingly used in mathematics classrooms for different purposes, including conceptual understanding development, skills development, performance improvement, attitude change, collaboration and discussion support, and teacher support (Bray & Tangney, 2017). The use of educational technology to enhance mathematics education has been emphasised in the current Finnish NCC (EDUFI, 2016), as well as in other international curricula (e.g., ACARA, 2016; NGA & CCSSO, 2010). Education technology has played a significant role in distance learning during the COVID-19 pandemic. Despite the potential benefits of technology in mathematics education, technology is only a tool, not an end. The value of technology utilisation in mathematics education depends on how it is used by teachers and students (National Council of Teachers of Mathematics [NCTM], 2014; Tran et al., 2017; Warren et al., 2016).

Mathematical manipulatives, such as beads, geoboards, and educational apps, have been advocated as hands-on learning tools that enable students, particularly pre- and primary school students, to explore abstract mathematical concepts through different senses. While previous studies on the pedagogical effectiveness of manipulatives have yielded mixed results (e.g., Manches & O'Malley, 2012; Uribe-Flórez & Wilkins, 2017), there is evidence that proper use of manipulatives is likely to enhance students' understanding of mathematical concepts (e.g., Carbonneau et al., 2013; Kilpatrick et al., 2001; McNeil & Jarvin, 2007).

Most research (e.g., Carbonneau et al., 2013; Moyer-Packenham & Westenskow, 2013; Uribe-Flórez & Wilkins, 2017; Vessonen et al., 2020) has focused on examining the impact of manipulatives on students' mathematics learning and achievement. Some studies (e.g., Marshall & Swan, 2008; Moyer-Packenham et al., 2013) have investigated the use of manipulatives in the classroom, including commonly used manipulatives, grade level utilisation, frequency of use, advantages and disadvantages, and hindrances to their utilisation. Research findings reveal disagreement between pedagogical benefits and the classroom utilisation of manipulatives. Although primary and lower secondary school teachers have considered manipulatives to be beneficial to mathematics learning, they usually prefer to use traditional teacher-centred and paper-and-pencil instruction in their classrooms (e.g., Joutsenlahti & Vainionpää, 2010; Marshall & Swan, 2008; Toptaş et al., 2012). This finding signals that the classroom adoption of pedagogically sound manipulatives may be hindered by day-to-day practice-related factors.

To date, a considerable number of studies have been conducted to gain a better understanding of manipulatives, particularly regarding their benefits to mathematics learning and classroom utilisation. Little attention has been paid to holistically integrating the research findings regarding these aspects. Moreover, most research has failed to directly link its findings on manipulatives and the actual utilisation of manipulatives in real educational settings, apart from providing implications for practice.

This indicates a need for research that holistically investigates how to benefit from the use of manipulatives in classrooms and then utilises the derived knowledge to develop manipulatives that enhance students' conceptual understanding as well as classroom practice. Such research would simultaneously contribute to both theory and practice in mathematics education.

1.2 Objectives and research questions

This doctoral research was conceived during my time as a primary school remedial teacher. I witnessed an incident in which mathematical manipulatives were buried under a pile of dust in the school copy room. The incident left me wondering why potentially useful learning tools bought with the school's tight budget were abandoned. Having a background in educational design and primary school education, I intend to utilise my expertise to answer this question and find a better solution.

This study used an educational design research (EDR) approach to bridge the research on manipulatives and its direct benefits to real-world educational practice. Learning to solve one-variable linear equations in primary schools was used as a case study because this important area in algebra has typically been taught regarding rules and procedures instead of the concepts contributing to those rules and procedures (Figueira-Sampaio et al., 2009; Kilpatrick et al., 2001). Research has indicated that students' inadequate understanding of equation-solving concepts hinders their equation-solving learning and performance (e.g., Booth & Koedinger, 2008; Knuth et al., 2006).

This study aims to develop a manipulative and its appropriate use to promote not only students' understanding of mathematical concepts, but also classroom adoption. It also intends to possibly take advantage of technologies that can contribute to this aim. This study was undertaken through three phases of a 6-year EDR enquiry: initial research, concept development, and design development (Figure 1). First, the challenges and opportunities of using manipulatives in primary school classrooms were investigated. Then, four manipulative concepts that help students understand linear equation-solving concepts were generated and evaluated. Finally, a manipulative and its accompanying instructional materials and class activities were developed. The manipulative was developed by taking into account both pedagogy and practicality to ensure its successful utilisation and adoption in the classroom. It was then implemented and evaluated in the classroom.

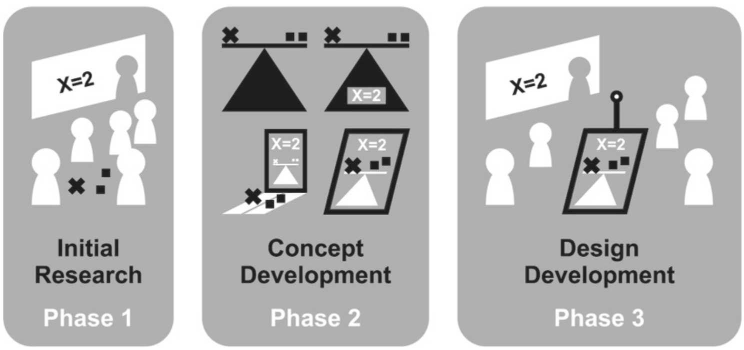


Figure 1. Three phases of doctoral study, illustrated by T. Lehtonen

The research-based design solution that directly improves educational practice is the practical contribution of this study. Additionally, this study intends to advance three types of EDR theories proposed by Edelson (2002): (1) *domain theories*, knowledge about a real-world educational problem to be solved (*context theory*), and outcomes of implementing a design solution to solve that particular problem (*outcomes theory*); (2) *design frameworks*, knowledge about the characteristics of a successful design solution to a particular educational problem in a particular setting; and (3) *design methodologies*, knowledge about successfully conducting EDR (for more details, see Chapter 3.1). The study strives to answer the following research questions (RQs):

1. Domain theory is divided into two sub-questions:

    1.1 Context theory: What are the needs, challenges, and opportunities of using manipulatives in primary school classrooms? What are the strengths and limitations of existing manipulatives? (Publications II and IV)

    1.2 Outcomes theory: How does the developed design solution help students understand equation-solving concepts and encourage teachers to adopt it in their classrooms? (Publications III and IV)

2. Design framework: What key aspects should be taken into account when developing a manipulative to ensure its successful classroom utilisation and adoption? (Publications III and IV)

3. Design methodology: What guidelines for conducting successful EDR can be drawn from the lessons learnt from undertaking this study? (Publications I and IV)

The RQs were addressed in four publications. The contributions of the publications to the RQs are presented in parentheses after each RQ.

1.3 Summary of the publications

Publication I aims to better understand how EDR, a research approach of this dissertation, has been used, and what challenges it has encountered during the last two decades in similar research environments. This publication systematically reviewed 21 Finnish EDR doctoral dissertations on mathematics, science, and technology education published between 2000 and 2018, particularly in terms of their EDR process, research methodology, contributions, and challenges. Although the publication did not directly answer RQ 3, its results guided the planning and implementation of this study and provided fundamental methodological knowledge for Publication IV.

Publication II seeks to advance the context knowledge regarding the use of manipulatives in classrooms to support students' understanding of equation-solving concepts (partially RQ 1.1). It investigated pedagogical challenges and opportunities when learning to solve equations with two existing manipulatives in real classrooms compared to traditional paper-and-pencil instruction. The publication reports on the pedagogical benefits and limitations of both manipulatives. The findings provided initial ideas for how to design potential solutions that could better enhance students' understanding of equation-solving concepts.

Publication III reports the design principles guiding design solution development as well as the development, implementation, functions, and features of the design solution, its classroom evaluation regarding pedagogical benefits and usability, and its future development. This publication provides outcome knowledge about how the developed manipulative enhanced students' understanding of equation-solving concepts and how usable it was in real educational settings (partially RQ 1.2). Additionally, the established design principles contribute to the design framework (RQ 2) by providing key characteristics for a design solution to successfully improve educational practice in this particular context.

Publication IV summarises the overall EDR process of this doctoral study. This publication highlights various practical factors as well as the strengths and limitations of existing manipulatives that were found to have an effect on manipulative utilisation and adoption in classrooms (partially RQ 1.1). It continues to describe how these findings informed the design development and how the incorporation of practical considerations into the design contributed to the positive design evaluation results (partially RQ 1.2). Additionally, based on the experience of the doctoral study, the publication proposes a design framework that underlines key aspects that should be taken into account when developing design solutions to a similar educational problem in other contexts (RQ 2). Built on the findings in Publication I, Publication IV highlights the lessons learnt from undertaking this EDR and then provides guidelines for conducting successful EDR (RQ 3).

1.4 Structure of the dissertation

This dissertation is divided into seven chapters. The introduction (Chapter 1) is followed by Chapter 2, which highlights the theoretical background of the research.

Chapter 3 outlines the research methodology: research approach, context, process, design, as well as ethics and integrity. Chapter 4 describes the research process and key findings of all three research phases: initial research, concept development, and design development. Chapter 5 reflects on my experience in developing the design solution and undertaking this EDR. Chapter 6 evaluates the research quality of this dissertation. Chapter 7 summarises the theoretical outcomes of the research, explores research contributions, reflects on research limitations, and finally provides suggestions for future research.

2 THEORETICAL BACKGROUND

This doctoral research aims to develop a manipulative and its appropriate use to promote not only students' understanding of equation-solving concepts, but also classroom adoption. In this chapter, I outline the theoretical background necessary for the development of such manipulatives and how to use it meaningfully. Learning mathematics with understanding is presented first, followed by mathematics manipulatives and linear equation solving.

2.1 Learning mathematics with understanding

2.1.1 Defining conceptual understanding

There has long been a debate in mathematics education regarding the competence needed to succeed in learning mathematics, particularly procedural knowledge (e.g., Hiebert & Carpenter, 1992; Kilpatrick et al., 2001). Traditional school mathematics typically emphasises developing students' procedural knowledge and computational skills (e.g., Groves, 2012; Kilpatrick et al., 2001), but a growing body of literature and research has acknowledged that knowledge and skills in handling computational procedures alone are insufficient to succeed in learning mathematics (Kilpatrick et al., 2001; Schoenfeld, 2007). The need for a shift from this focus to also including other important mathematical proficiency has increasingly gained the attention of mathematics education scholars since the late 1900s.

Skemp (1976) identified two types of mathematical understanding: instrumental understanding and relational understanding. He described *instrumental understanding* as knowing how to perform mathematical procedures, and *relational understanding* as knowing both how to perform them and why (p. 2). He argued that learning instrumental mathematics teaches a learner step-by-step instructions on how to complete a particular task, so the learner usually needs new guidance for completing a new task. On the contrary, relational understanding enables learners to construct their own paths to execute a certain task and be able to execute novel tasks.

According to Hiebert and Lefevre (1986), mathematical knowledge comprises the primary relationship between two types of knowledge: conceptual and procedural knowledge. They defined *conceptual knowledge* as a network that links all pieces of information (pp. 3–4), and *procedural knowledge* as a familiarity with mathematical symbols as well as rules and procedures used to execute mathematical tasks (p. 6). They emphasised that both types of knowledge were important and closely related. They also cautioned that rote learning resulted in knowledge that was closely tied to a specific mathematical task and not connected with other knowledge. It is difficult to apply such knowledge to other tasks that are different from those for which it was initially intended. Therefore, to be competent in mathematics, students need to meaningfully learn sufficient conceptual and procedural knowledge, as well as connect them.

Kilpatrick et al. (2001) proposed a model of mathematical proficiency—knowledge, skills, abilities, and beliefs—for today's students to learn mathematics successfully. Their concept of mathematical proficiency consists of five intertwined strands: *conceptual understanding* (understandings of mathematical concepts, operations, and relations), *procedural fluency* (ability to perform mathematical procedures flexibly, accurately, efficiently, and appropriately), *strategic competence* (capability to formulate, represent, and solve mathematical problems), *adaptive reasoning* (competence in logical thinking, reflection, explanation, and justification), and *productive disposition* (perception of mathematics as sensible, useful, and worthwhile, and belief in their own diligence and efficacy; p. 5). Kilpatrick et al. emphasised that each component of mathematical proficiency was important, and interwoven and interdependent. Their (2001) model of mathematical proficiency has been adopted by mathematics curricula and standards in various countries, including Australia (ACARA, 2016) and the United States (NGA & CCSSO, 2010), and endorsed by various mathematics education scholars (e.g., Moschkovich, 2015; Schoenfeld, 2007).

Within this dissertation, conceptual understanding (used interchangeably with mathematical understanding and understanding of mathematical concepts) is defined based on Skemp's (1976) relational understanding, Hiebert and Lefevre's (1986) conceptual knowledge, and Kilpatrick et al.'s (2001) conceptual understanding. Although the main interest of this study is understanding mathematical concepts, an understanding of operations and relations (e.g., why the procedure is appropriate for a particular task) was also included in the study due to their interconnection.

There has been growing recognition that students' inadequate understanding of mathematical concepts can result in their low performance in mathematics

(Kilpatrick et al., 2001). Thus, mathematics education must foster students' mathematical understanding, along with other components of mathematical proficiency (e.g., Barmby et al., 2010; Kilpatrick et al., 2001; Moschkovich, 2015). According to the literature and previous studies, strong conceptual understanding has a number of advantages to mathematics learning, including:

- Promoting retention of what has been already learnt. Knowledge learnt with understanding is well connected, and therefore easier to remember (Hiebert & Carpenter, 1992; Kilpatrick et al., 2001; Skemp, 1976), long-lasting (Hiebert & Carpenter, 1992), and easier to recall when forgotten (Kilpatrick et al., 2001).
- Helping avoid errors. Knowledge that has been learnt with understanding is unlikely to be remembered incorrectly (Kilpatrick et al., 2001).
- Promoting learning transfer. Knowledge learnt with understanding enables students to adapt what they have learnt to solve novel problems (Hiebert & Carpenter, 1992; Kilpatrick et al., 2001; Skemp, 1976).
- Supporting the learning of new concepts. Prior knowledge learnt with understanding enables students to learn new related concepts more easily because they can see the relationship between learnt and to-be-learnt knowledge (Hiebert & Carpenter, 1992; Kilpatrick et al., 2001).
- Providing intrinsic rewards. Students can naturally get satisfaction from learning with understanding; therefore, there is no need for external rewards and punishments (Skemp, 1976).
- Enhancing self-discovery and self-directed learning. Knowledge learnt with understanding enables students to extend their knowledge independently (Kilpatrick et al., 2001), and satisfaction with learning with understanding motivates students to actively continue learning (Skemp, 1976).
- Enhancing self-confidence. The better understanding students have, the more confidence they have in their own competence to independently accomplish tasks (Kilpatrick et al., 2001; Skemp, 1976).
- Influencing the right belief in mathematics. Conceptual understanding enhances students' belief that mathematics is a network linking pieces of knowledge, that mathematics representation systems are consistent, and that correspondences between representation systems are predictable.

> Such beliefs can promote the further development of mathematical knowledge (Hiebert & Carpenter, 1992).

Conceptual understanding is one of the key mathematical proficiencies that the current Finnish NCC (EDUFI, 2016) intends to develop among students. Mathematics instruction must assist students in developing their understanding of mathematical concepts.

2.1.2 Developing conceptual understanding

Understanding is gradually built as networks of *internal representations* (i.e., an individual's cognitive constructs, such as verbal, imagistic, and symbolic; Goldin, 2014) of mathematical concepts or procedures become larger and more cohesive (Hiebert & Carpenter, 1992). A number of scholars (e.g., Barmby et al., 2010; Kilpatrick et al., 2001; Lesh et al., 1987) argue that *external representations* (i.e., visible or tangible productions, such as mathematical expressions, number lines, depictions on a computer screen, and manipulatives that stand for mathematical ideas or relationships; Goldin, 2014) provide students access to abstract mathematical concepts, and thus help them understand these concepts. Representations can be classified into different categories regarding their attributes. For example, Bruner (1966) argued that a person develops their own understanding through three modes of representations: *enactive* (e.g., combining two and three candy bars), *iconic* (e.g., seeing a picture of two and three candy bars), and *symbolic* (e.g., reading 'two candy bars plus three candy bars' or '2 + 3'). Lesh et al. (1987) identified five external representational systems commonly used in mathematics: real scripts, spoken language, written symbols, static pictures, and manipulatives. These external representations embody students' internal conceptualisations and are observable.

To develop mathematical understanding, students need to develop their *representational fluency* (also known as representational competence, representational flexibility, and representational thinking), which is the ability to represent mathematical concepts and procedures in multiple forms, interpret these multiple representations, and make connections between and within them (e.g., Ceuppens et al., 2018; Lesh et al., 1987; Zbiek et al., 2007). Studies during the last decades (e.g., Brenner et al., 1999; Moyer-Packenham et al., 2021; Suh & Moyer, 2007) have indicated that representational fluency plays an important role in mathematics learning. Hiebert and Carpenter (1992) agree with constructivist perspectives, particularly discovery learning (see Bruner, 1961), that students actively develop their own networks of mathematical representations by constructing internal representations of their direct interactions with the environment, rather than having them provided by teachers or textbooks. Informed by research, the NCTM (2014) calls for the use of multiple representations as one of the mathematics teaching practices for strengthening mathematics education.

Pape and Tchoshanov (2001) found that the theory of cognition and brain investigation provides support for how individuals develop their mathematical understanding through representations. The brain functions when constructing representational patterns for internalising (i.e., encoding) external representations and externalising (i.e., decoding) mental images; an external representation reduces a person's cognitive load, thereby enabling them to devote their cognitive resources to learning (Pape & Tchoshanov, 2001). Representations are not inherently meaningful on their own; thus, they need to be manipulated with reflection to lead to mathematical understanding (Hiebert & Carpenter, 1992). Pape and Tchoshanov (2001) argued that representations must be seen as cognitive tools to assist students in thinking, explaining, and justifying an argument instead of the end result of a task. Likewise, drawing upon Sierpinska's (1994) process of understanding, Barmby et al. (2010) proposed a 'representational–reasoning' model of understanding. The model provides implications for developing students' mathematical understanding in the classroom, stating that students should be provided with various external representations of a concept, as well as opportunities for them to develop their reasoning between the representations. When using multiple representations, students examine a mathematical concept through various lenses, with each lens offering a different viewpoint that enriches and deepens the concept (Tripathi, 2008). Similarities and differences between multiple representations of the same concept enable students to see the relationships between these representations (Hiebert & Carpenter, 1992).

Individually working with representations does not guarantee that students will eventually develop their understanding (Hiebert & Carpenter, 1992; Pape & Tchoshanov, 2001). From a social constructivist perspective, learning is constructed during a collaborative process; learners co-construct their knowledge through social interaction (Vygotsky, 1978). In this light, it has been recommended that students' internalisation of external representations and externalisation of their mental images should take place through social activity (i.e., interaction with peers and teachers) to support the construction of their understanding. Thus, external representations work not only as access to abstract mathematical concepts (Barmby et al., 2010), but also as vehicles for co-constructing their understanding with peers and teachers (Pape &

Tchoshanov, 2001). Classroom discussions (in pairs, small groups, and the whole class) enable students to explain, construct arguments, justify, and reason about possible relationships between multiple representations of a mathematical concept, thereby helping them develop their conceptual understanding accordingly (e.g., Hiebert & Carpenter, 1992; Moschkovich, 2015; Pape & Tchoshanov, 2001). In particular, peer discussion yields various benefits for learning. For example, explaining their own thoughts to classmates helps students remember that information and relate it to information previously kept in their memory (Slavin, 2010); presenting their own thoughts and listening to others allows students to examine and reflect upon their own reasoning (Barmby et al., 2010).

2.1.3 Assessing conceptual understanding

There are a number of complications concerning how to assess students' mathematical understanding. Hiebert and Carpenter (1992) pointed out that because understanding is a mental process of gaining knowledge, it is not directly observable and thus not easy to assess. Consequently, students' understanding is indirectly measured through teachers' or researchers' inferences, drawn from observable evidence. They also noted that students can perform any individual task correctly without understanding, and students' understanding normally cannot be concluded from their single response on a single task. Rather, a variety of tasks are required to create a profile of students' understanding. Barmby et al. (2007) also emphasised that because understanding is a complex network, quite often, the assessment of students' mathematical understanding can offer insight into only a small part of such a complex network.

Hiebert and Carpenter (1992) suggested that assessment of students' understanding could be thought of as a reverse process of how students construct their networks of internal representations through interaction with external representations. Thus, students' internal networks of mathematical concepts can be determined by asking them to communicate their understanding through connections between external representations (Barmby et al., 2007; Hiebert & Carpenter, 1992). Nevertheless, it is difficult to determine whether students' manipulations of external representations indicate their understanding, or whether they simply manipulate the representations mechanically (Hiebert & Carpenter, 1992). Students' reasoning behind their mathematical solutions—for example, why a certain result is the correct answer—can demonstrate their conceptual understanding (Moschkovich, 2015). In light of this, to ensure that the assessment actually reflects students' mathematical understanding, it should measure students' ability to make links between multiple external representations and their ability to reason (Barmby et al., 2010). Unlike standard assessments, the assessment of students' understanding should take into account both the content and the process that students use to arrive at their solutions (Hiebert & Carpenter, 1992).

Students can be asked to represent a particular mathematical concept or procedure in multiple ways (e.g., using concrete objects, pictures, or mathematical expressions) and then talk about, for example, the links between the external representations or their approach to the solution (Barmby et al., 2007; Hiebert & Carpenter, 1992; Moschkovich, 2015). Opportunities to explain their own reasoning in several ways, including verbally and in writing, while tackling mathematical tasks, should be provided to students (Barmby et al., 2010). To analyse students' understanding, all responses and explanations provided by students can be categorised based on the perceived level of their understanding; a mark can be given to each category of the given explanations (Barmby et al., 2007).

There are some difficulties with this assessment method. Unless students are familiar with the external representations provided, they may not be able to work with the representations, even though they understand the mathematical concepts or procedures (Hiebert & Carpenter, 1992). Moreover, it may not be possible to look for evidence of students' mathematical understanding from their explanations. Students usually construct their understanding before they can actually verbalise what they understand (Kilpatrick et al., 2001) and often have difficulties explaining what they think in a written form (Laine et al., 2018). There are other possible methods for assessing students' mathematical understanding that are beyond the scope of this study, for example, analysis of students' errors (Hiebert & Carpenter, 1992) and the use of concept maps and mind maps (Barmby et al., 2007).

2.1.4 Multimodality in mathematics

Mathematics is by nature multimodal; that is, different semiotic resources (cf. modes of representations and representational systems) are used in mathematical texts and in the mathematics classroom to make meanings (e.g., Jewitt et al., 2016; Morgan et al., 2021; O'Halloran, 2015a). It should be noted that *meaning making* involves only one person's expression and/or interpretation, while *communication* involves one person's expression and another person's interpretation (Jewitt et al., 2016).

O'Halloran (2004, 2015a) distinguishes three semiotic systems used for meaning making in mathematics: linguistic (i.e., natural language), symbolic (i.e., numbers and symbols), and visual (e.g., pictures and graphs) forms of representation. While the three semiotic resources are interwoven as a whole, each has its own characteristics, affordances, constraints, and roles in the construction of mathematical knowledge (e.g., Morgan et al., 2021; O'Halloran, 2015b; Schleppegrell, 2010). Natural language facilitates rationally thinking about mathematical processes and their results; symbols precisely and unambiguously describe mathematical relations and assist in arriving at mathematical solutions; and images visualise and concretise abstract mathematical concepts or process (O'Halloran, 2015a, 2015b). According to Duval (2006), a mathematical idea cannot be directly sensed, but can be experienced through its representations. However, a representation (e.g., a mathematical expression, a graph, or a verbal description) is never identical to the mathematical idea that it represents and cannot portray aspects of the mathematical idea completely due to the specific characteristics of each representational system.

Because a separate semiotic resource is not sufficient for meaning making, multimodality plays a crucial role in the mathematics classroom. To learn mathematics with understanding, students need to be able to interpret and benefit from multisemiotic resources simultaneously (O'Halloran, 2004; Schleppegrell, 2010). O'Halloran (2015a) uses a term *multimodal literacy* for this mastery of meaning making with natural language, mathematical symbols, and visual images. Morgan et al. (2021) noted that recent technological advancements have offered novel semiotic resources, such as dynamic, manipulable, and linked images, that enable new methods of interaction in mathematics education. They also highlighted the growing interest in the roles of embodied forms of interaction (e.g., gesture and gaze) in the mathematics classroom. Similarly, drawing on an embodied cognition perspective, Alibali and Nathan (2012) argue that mathematical knowledge is embodied (i.e., based in perception and action, and grounded in the physical environment; p. 247). They also presented evidence for their claim using gestures that teachers and students produced when explaining mathematical concepts or ideas.

In this study, I employed a *languaging* model (also known as languaging mathematics and multimodal languaging) to guide design solution development, for example, what modes of representations the solution should provide and how it should be used in the classroom to foster students' mathematical understanding. The languaging concept was previously used in mathematics education (see Bauersfeld, 1995) and second language education (see Swain, 2006) in relation to verbal communication in classrooms. Joutsenlahti and Rättyä (2015) extended the concept of language in mathematics as a process in which students express their own mathematical thinking using four different languages (i.e., semiotic systems): natural, mathematical symbolic, pictorial, and tactile. While the first three languages are commonly used in mathematics texts to make meanings of mathematical concepts and procedures (see e.g., O'Halloran, 2004), tactile language (e.g., manipulating concrete objects) is used, particularly in primary school mathematics classrooms. The four languages used in this model align with Bruner's (1966) three modes of representation (tactile language/enactive representation, pictorial language/iconic representation, and natural and mathematical symbolic language/symbolic representation). It is noteworthy that Bruner's (1966) representation modes are typically applied as sequence-based learning from the enactive to iconic to symbolic (i.e., from concrete to abstract representations); for example, in Fyfe et al.'s (2015) and Warren and Cooper's (2009) interventions. However, in the languaging model, the use of different languages happens simultaneously.

According to Joutsenlahti and Kulju (2017) and Joutsenlahti and Rättyä (2015), languaging plays an important role in the mathematics classroom. When using different languages to solve mathematical tasks or presenting their own solutions to others, students organise their own thinking, and thus gradually construct their understanding of mathematical concepts and procedures. When listening to or looking at peers' solutions, students reflect on their own thinking or solutions, and therefore develop better understanding. Moschkovich's (2015) *academic literacy in mathematics* provides support for the languaging model. Using multiple representational systems to think and communicate (e.g., to talk, listen, write, or draw) helps students make connections between different ways of representing mathematical concepts, and enhances their conceptual understanding (Moschkovich, 2015). Besides supporting the development of students' understanding, languaging makes students' thinking observable, and can be used as a tool for assessing how students understand mathematics concepts or procedures (Joutsenlahti & Kulju, 2017; Joutsenlahti & Rättyä, 2015). Findings from research on languaging-based mathematics instruction at different educational levels (e.g., Alfaro & Joutsenlahti, 2020; Joutsenlahti & Kulju, 2017) suggest that the use of different languages for expressing mathematical thinking can promote students' conceptual understanding and support the assessment of their understanding.

Multimodality and languaging have been emphasised in the Finnish NCC 2014 (EDUFI, 2016) as important teaching and learning methods in the mathematics classroom. Students are guided to make use of concrete tools, mathematical symbols, natural language, visuals, and ICT in developing their mathematical thinking and presenting it to their teachers and classmates.

2.2 Mathematics manipulatives

2.2.1 Manipulatives in mathematics education

Manipulatives are hands-on learning tools that are used to concretely represent an abstract mathematical concept, and can be manipulated by students through various senses to construct their understanding of the concept to be learnt (e.g., McNeil & Jarvin, 2007; Moyer, 2001). Manipulatives can be categorised into three types based on forms of interaction between students and manipulatives: physical, virtual, and tangible.

Physical manipulatives (PMs) are physical objects, such as beads, geoboards, and base-10 blocks, which have long been used as hands-on learning tools in the mathematics classroom. PMs provide unique benefits, including concretising abstract concepts (Bruner, 1966), encouraging physical action to promote learning (Martin & Schwartz, 2005), offloading cognition (Manches & O'Malley, 2012), improving memory through physical action (McNeil & Jarvin, 2007), assisting embodied cognition (Pouw et al., 2014), serving as tools for reflection and communication (Kilpatrick et al., 2001), and making students' thinking visible to teachers (Marshall & Swan, 2008).

Despite their contributions to learning, PMs have possible limitations. Students may not be able to manipulate manipulatives as intended and might need guidance on how to work with them meaningfully (Carbonneau et al., 2013). Even if students manipulate manipulatives in an appropriate manner, they may have difficulty extracting an underlying mathematical concept from its external representation, such as a physical object (e.g., Clements, 1999; Kilpatrick et al., 2001; Pape & Tchoshanov, 2001). Therefore, they need support in interpreting certain physical representations, that is, how to induce mathematical meanings from their interactions with PMs (e.g., Clements, 1999; Kilpatrick et al., 2001; Manches & O'Malley, 2012). Moreover, students may not see the connection between the manipulatives-based and symbolic representations of the concept (e.g., Kilpatrick et al., 2001; Pape & Tchoshanov, 2001; Uttal et al., 2013). In this case, they may also need support in connecting the concrete representation with the symbolic representation (e.g., Kilpatrick et al., 2001;

Manches & O'Malley, 2012; Uttal et al., 2013). With typically only one teacher in the classroom, it is challenging to help individual students interpret a physical representation as intended and connect it with the symbolic counterpart. PMs are also typically expensive to purchase, which limits their availability (Magruder, 2012).

During the last few decades, the use of *virtual manipulatives* (VMs, also called *computer manipulatives* by Clements, 1999) has become established in mathematics classrooms due to the development of digital technologies. VMs are interactive and dynamic technology-enabled visual representations in the form of computer applets or mobile apps (e.g., Manches & O'Malley, 2012; Moyer-Packenham & Bolyard, 2016). Research has reported various advantages of VMs that address the limitations of PMs. VMs benefit learning by providing step-by-step guidance, immediate feedback, and scaffolding (e.g., Magruder, 2012; Suh & Moyer, 2007); drawing students' attention to what is relevant; promoting students' creativity and increasing the diversity of their solutions; providing precise representations; simultaneously linking pictorial and symbolic representations with students' interactions; recording and tracking students' actions for reflection (Martin, 2008) and assessment (Clements, 1999; Marichal et al., 2017); and motivating students (Moyer-Packenham & Westenskow, 2013), particularly older students (Vessonen et al., 2020). Practical benefits of VMs include flexibility, manageability, and ease of sharing (Clements, 1999; Manches & O'Malley, 2012), ease of cleaning up (Magruder & Mohr-Schroeder, 2013), ease of storing and retrieving configurations (Clements, 1999), accessibility (Magruder, 2012), availability (Marichal et al., 2017; Vessonen et al., 2020), and affordability (Manches & O'Malley, 2012).

Nevertheless, interaction with VMs has caused concern about replacing rich physical interaction (with PMs) with mouse-keyboard clicking or touch screen tapping and scrolling (Manches et al., 2010; Pires et al., 2019). Other possible disadvantages of VMs include rote learning, distraction (Magruder, 2012), losing mathematics learning time to learning how to operate VMs or solving technical issues, and inaccessibility and unavailability of necessary technology (Magruder, 2012; Magruder & Mohr-Schroeder, 2013).

Previous studies (e.g., Magruder, 2012; Manches et al., 2010; Suh & Moyer, 2007) indicate that combining advantageous properties of PMs and VMs may better enhance learning and classroom practice than using PMs or VMs alone. More recently, *tangible manipulatives* (TMs), also called *digital manipulatives* by Resnick et al. (1998), have begun to emerge as a result of the advancement of tangible technologies. Tangible technologies embed digital technology into physical objects, which serve as both external representations and controls (e.g., Ishii & Ullmer, 2012).

TMs enable a novel form of interaction in which students directly operate digital information via the natural manipulation of physical objects, rather than through the cumbersome manipulation of typical input devices (e.g., mouse or keyboard). Since the pioneering work (i.e., blocks, beads, balls, and badges) of Resnick et al. (1998), a growing number of TMs have been developed (e.g., Ceibal Tangible; Marichal et al., 2017, and Owlet manipulatives; Zito et al., 2021) or fully commercialised (e.g., OSMO's math games; https://www.playosmo.com).

By combining the best of PMs and VMs (i.e., physical and digital worlds), TMs provide various advantages for learning. One benefit of physical representations is that the manipulation of physical objects enables natural bodily interaction and creates a sense of physicality and embodiment (Ishii & Ullmer, 2012). Moreover, according to Manches and O'Malley (2012), physical manipulation can free up students' cognitive resources for learning mathematical concepts. Instead of learning how to manipulate manipulatives, as in the case of VMs, students are able to intuitively manipulate physical objects and can concentrate on the results of their actions. In contrast to VMs, TMs provide tactile information, which enables students to offload their cognitive demands. For example, students are able to identify some information by touching an object, and can thus use their visual attention for other information. Tactile information also enables accessibility for students with visual impairments.

Digital representations provide various benefits for TMs. Manches and O'Malley (2012) argue that digital effects can be used to change different perceptual properties of TMs, such as colour, size, and sound, which are not easy to change in the case of PMs. This provides the opportunity to create a more dynamic manipulative material that changes its perceptual properties, representing certain mathematical concepts as students interact with it. Digital representations can also provide immediate feedback and scaffolding, which promotes autonomous learning (Price, 2013). Additionally, digital representations enable the recording of students' activities and therefore allow students' work to be traced later (Manches & O'Malley, 2012) for reflection and assessment.

The integration of physical and digital representations of TMs also benefits learning. Physical actions with objects and corresponding digital effects together can provide a conceptual metaphor for concepts to be learnt, thereby fostering students' knowledge structures (Manches & O'Malley, 2012). TMs facilitate linking between physical representations (e.g., touch and gestures) and digital representations (e.g., pictorial, symbolic, and other representations; Manches & O'Malley, 2012) and thus contribute to students' conceptual understanding. Moreover, TMs attract students' multiple senses (Zhou & Wang, 2015). TMs also enable parallel multi-user interaction, which promotes co-located and distanced collaborations (Ishii & Ullmer, 2012; Price, 2013); encourage facial, gestural, and verbal communication (Price, 2013); enable accessibility to different learners; and motivate learning (Zhou & Wang, 2015).

Previous research has provided empirical evidence of TM benefits for mathematics education in various content domains and educational levels. For example, primary school number composition (Pires et al., 2019), primary school fractions (Pontual Falcão et al., 2007), pre-primary and primary school geometry (Salvador et al., 2012; Starcic et al., 2013), undergraduate school trigonometry (Zamorano Urrutia et al., 2019), and upper and lower secondary school algebra (Reinschlüssel, Döring, et al., 2018).

2.2.2 Theoretical underpinnings for the use of manipulatives

The use of manipulatives in mathematics classrooms has long been endorsed based on Piaget's, Bruner's, and Montessori's work that physical interactions with concrete materials, such as (physical) manipulatives, help children construct their understanding of abstract mathematical concepts (e.g., Carbonneau et al., 2013; McNeil & Jarvin, 2007; Uttal et al., 2013). To date, there have been various theoretical explanations that manipulatives support learning by providing opportunities for students to learn mathematical concepts through exploration, supporting their real-world knowledge, and enhancing their cognition.

Bruner's (1961) discovery learning provides a theoretical underpinning for using manipulatives to facilitate mathematics learning through hands-on experience and reflection. As opposed to rote learning, active interaction with the environment (e.g., manipulatives) and inductive reasoning help students construct their own meaningful knowledge of mathematical concepts (Bruner, 1961). During discovery learning, students should be appropriately supported with guidance and scaffolding rather than left unaided (Carbonneau et al., 2013; Neber, 2012). Research (e.g., Suh & Moyer, 2007; Zamorano Urrutia et al., 2019) reveals that VMs and TMs can provide students not only with hands-on experience, but also with guidance and/or scaffolding during their discovery learning.

Cognitive perspectives, particularly embedded embodied cognition, also account for the potential advantages of using manipulatives in classrooms. From the embedded cognition perspective, manipulatives can benefit mathematics learning by providing students with an additional representation of mathematical concepts (e.g., McNeil & Jarvin, 2007; Tran et al., 2017) and reducing students' cognitive load, and thus allow students to devote their cognitive resources to learning (e.g., Pouw et al., 2014; Tran et al., 2017). Martin and Schwartz's (2005) theory of physically distributed learning proposes that physical action, such as interaction with manipulatives, enables students to rearrange their environment and gradually develop their thinking, thereby benefiting their learning of abstract ideas. From the embodied cognition perspective, humans have a natural desire to connect the body and mind (e.g., the use of fingers to count; Tran et al., 2017), and cognitive processes are often composed of mental simulations that are founded on modalities (Marley & Carbonneau, 2014). Interactions with manipulatives allow physical enactment of mathematical concepts, and thus increase forms of representations (e.g., audiovisual, tactile-kinesthetic, and symbolic in the case of TMs) for encoding the concepts (Carbonneau et al., 2013; Tran et al., 2017). Having access to multiple representations enhances students' understanding (by allowing them to connect different representations) and their memory (by helping them easily retrieve knowledge stored in their minds later; Carbonneau et al., 2013; Tran et al., 2017). Students' memories of interacting with manipulatives can promote their learning transfer to other situations (Alibali & Nathan, 2012; Pouw et al., 2014). Usually when students' internal cognition increases, their dependence on manipulatives decreases (Pouw et al., 2014). Eventually, the interaction with manipulatives is no longer needed (Tran et al., 2017).

2.2.3 Impacts of manipulatives on mathematics learning

Despite the theoretical support for using manipulatives in classrooms, empirical findings on the effectiveness of manipulative use on students' mathematics learning and achievement are mixed (e.g., Manches & O'Malley, 2012; McNeil & Jarvin, 2007; Uribe-Flórez & Wilkins, 2017). The use of manipulatives has been found to benefit, not benefit, and even hinder mathematics learning.

Recent meta-analyses by Domino (2010), Carbonneau et al. (2013), and Moyer-Packenham and Westenskow (2013) and a longitudinal analysis by Uribe-Flórez and Wilkins (2017) provide evidence of positive impacts of manipulatives on mathematics learning. Domino's (2010) meta-analysis examined results from 31 empirical studies with 35 effect sizes that compared mathematics instruction, in which 5,288 students from kindergarten to Grade 6 learnt either with or without

PMs. Their statistically significant results indicated that using PMs for mathematics instruction had a moderate-to-large effect on students' mathematics achievement compared to traditional teaching methods without manipulatives. Findings also revealed that other instructional variables, including grade levels, instructional duration, and learning ability, increased the efficacy of manipulatives. The effect sizes increased from kindergarten to Grade 4, and then started to decrease slightly. PMs were beneficial to students at all attainment levels, particularly students with learning disabilities and with high attainments. Domino used previous studies to explain her findings that repetitive hands-on experiences with manipulatives helped students with learning disabilities learn mathematics; manipulatives assisted students with high attainments in quickly understanding mathematical concepts and then transferring their understanding to other problems.

Carbonneau et al.'s (2013) meta-analysis examined results from 55 studies that compared instruction, in which 7,237 students from kindergarten to college level either interacted with PMs or only learnt with mathematical symbols. Similar to Domino (2010), Carbonneau et al. (2013) found statistically significant results identifying small to moderate effect sizes in favour of instruction with manipulatives compared to instruction with mathematical symbols. Additionally, they found that the degree of this effect depended on other instructional variables, such as the cognitive development status of the learner and the level of instructional guidance.

According to a cognitive developmental perspective (see e.g., Piaget, 1965, 1970), using manipulatives to concretely represent abstract mathematical concepts should provide students at the preoperational stage (ages 2–7) and concrete operational stage (ages 7–11) with cognitive benefits, whereas it may not provide students at the formal operational stage (adolescence to adulthood) with compatible benefits. Carbonneau et al.'s (2013) findings partially support these predictions about cognitive development. At the aggregated level, studies with students aged 7–11 years old showed larger effect sizes than studies with students aged 12 and older. Nevertheless, when investigating retention of learning outcomes, studies conducted with preoperational-aged students showed a statistically lower and negative mean effect size compared to studies with concrete or formal operational students. A possible explanation for the lower effectiveness of manipulative use with younger students could be that younger students may have difficulties with manipulative dual representation (i.e., as an object itself and simultaneously as a representation of a mathematical concept; see e.g., Uttal et al., 2009). The meta-analysis results also provide support for conflicting recommendations for the level of guidance provided to students during their learning with manipulatives. On the one hand, the findings from the aggregated, retention, and problem-solving data affirm the perspective that high instructional guidance on how to use manipulatives promotes students' learning. On the other hand, the results from the transfer of learning outcome data support the perspective that low instructional guidance allows greater conceptual understanding, flexibility, and transfer of learning to novel circumstances.

Moyer-Packenham and Westenskow's (2013) meta-analysis examined results from 32 empirical studies with 82 effect cases that compared the effects of various instructional treatments on student achievement from the preschool to university level. Consistent with two other meta-analyses, they found moderate effects in favour of using VMs alone or in combination with PMs, compared to instruction from textbooks. The analysis of the effect of VMs by instructional duration provides support for the prolonged use of manipulatives for a positive impact on student performance. In general, the use of VMs for short treatment duration yielded no effect or small effects, while longer treatment duration yielded moderate effects. The analysis by grade level revealed mixed effects. When compared to other instructional treatments, VMs yielded small to moderate effects for preschool–Grade 6, negative effects for Grades 7–8, and large effects for Grades 9–12 and upwards. McNeil and Uttal's (2009) perspective provides a possible explanation for the mixed effects. Based on Bruner's (1966) view and previous studies, they argued that the benefits of learning through concrete objects were not limited to learners of a particular age, but in fact, applicable to any age. Unlike Piaget's (1970) view, this suggests that manipulatives can be used to introduce novel concepts to students of all ages.

Uribe-Flórez and Wilkins (2017) examined the relationship between mathematics learning of 10,673 students from kindergarten to Grade 5 and their manipulative use. When conducting a cross-sectional correlational analysis, they found no relationship between students' mathematics achievement at each specific year and their manipulative use. Moreover, no considerable relationship between students' achievement and their use of manipulatives was found at any grade level. Nevertheless, when conducting a longitudinal analysis over an extended period of kindergarten–Grade 5, they found a positive relationship between students' mathematics learning (i.e., growth and change in mathematics achievement) and their manipulative use across grade levels. Their findings affirmed Sowell's (1989) meta-analysis results, which provided evidence for a positive effect of long-term manipulative use on students' mathematics learning. Research results showing no relationship between manipulative use and students' learning may be due to too short a period of interaction time with the manipulatives.

Several studies have suggested that manipulative use may not assist or may even harm mathematics learning and performance (e.g., McNeil & Jarvin, 2007; Uribe-Flórez & Wilkins, 2017). For example, manipulatives tend to benefit only young students, and students may have difficulties transferring their knowledge (gained from manipulative use) to contexts without the presence of manipulatives (McNeil & Jarvin, 2007) or applying their knowledge to other contexts (Fyfe et al., 2015). McNeil and Jarvin (2007) provided two possible reasons for the inefficacy of manipulative use: manipulatives are beneficial but may not be used properly, or manipulatives themselves could cause the problem regarding dual representation.

For the first case, teachers may use manipulatives only for enjoyment or diversion in their classrooms rather than for facilitating students' construction of mathematical understanding. For the second case, there are several interferences in the dual representation of manipulatives; for example, students cannot possibly understand the mathematical concepts represented only by interacting with manipulatives, their restricted cognitive resources may be overloaded by the dual representation, and it may be difficult for students to perceive familiar objects differently. Goldin (2002) argued that manipulatives may be useful tools for supporting the development of students' conceptual understanding, but the contextualised understanding built concretely though manipulatives tends to pose a cognitive obstacle to the later abstraction of mathematical understanding, which is necessary for mathematics learning. Likewise, findings from Kaminski et al.'s (2009) experiments with 11-year-old (concrete operational) and undergraduate (formal operational) students demonstrated that whereas relevant concrete instantiations can potentially enhance learning, this contextual bound concreteness tends to hinder knowledge transfer.

Manipulatives have received approval as well as criticism. Clearly, merely using manipulatives does not necessarily lead to meaningful mathematics learning (e.g., Baroody, 1989; Kilpatrick et al., 2001; Manches et al., 2010; McNeil & Jarvin, 2007). Manipulatives as learning materials are only one of many factors that influence students' learning (Carbonneau et al., 2013; Manches & O'Malley, 2012). The benefits of manipulatives for learning depend on the learning context in which they are used, for example, how students work with manipulatives and for what purposes (Ball, 1992; Clements, 1999; Manches et al., 2010). Thus, contextual variables that may have an impact on the efficacy of manipulatives must be considered when planning mathematics instruction (Carbonneau et al., 2013).

Classroom activities and the teacher's role in promoting the meaningful use of manipulatives in mathematics learning have been recommended. Instead of using manipulatives mechanically just to get the correct answers, students should use them to meaningfully learn mathematics by thinking and reflecting on what they have experienced (Baroody, 1989; Clements, 1999) and by discussing with others what they have discovered (Ball, 1992; Marshall & Swan, 2008; Pape & Tchoshanov, 2001). The teacher's pedagogical approach and attitude to the use of manipulatives play a crucial role in what and how manipulatives are used in the classroom, thereby affecting the success of manipulative utilisation (Manches & O'Malley, 2012; Uribe-Flórez & Wilkins, 2010). Teachers should help students make a connection between various representations, which students construct when using manipulatives, and symbolic representations of the underlying concepts (e.g., Clements, 1999; McNeil & Jarvin, 2007) and map between the external and internal representations (Pape & Tchoshanov, 2001). However, according to the literature and empirical research, teachers should not direct students' actions with manipulatives, but rather allow students to direct and regulate their own activities (Hatfield, 1994; McNeil & Uttal, 2009).

2.2.4 Challenges to the use of manipulatives in the classroom

Although the benefits of meaningfully using manipulatives have been proposed in the literature, supported by research, and recommended by mathematics curricula, including the current Finnish NCC (EDUFI, 2016), much research indicates the limited use of manipulatives in the classroom. There is evidence that manipulatives are most often used in kindergarten, and the use of manipulatives usually declines as grade level increases (e.g., Carbonneau et al., 2013; Hatfield, 1994; Marshall & Swan, 2008; Uribe-Flórez & Wilkins, 2010). Teachers play a significant role in deciding whether and when to use manipulatives in their class. Their beliefs about manipulatives have been found to be an important predictor of their use (Uribe-Flórez & Wilkins, 2010), and may explain possible reasons for the discouragement to manipulative use. Teachers who believe in the efficacy of manipulative use are likely to use them more often in their class (Uribe-Flórez & Wilkins, 2010), whereas teachers who doubt their benefits are less likely to incorporate them in their mathematics instruction (Marshall & Swan, 2008). The decline in manipulative use in upper grades may be because teachers believe that manipulatives do not benefit older students (Uribe-Flórez & Wilkins, 2010; Marshall & Swan, 2008) or that using manipulatives is childish (Marshall & Swan, 2008).

Previous studies demonstrate disagreement between teachers' positive attitudes towards manipulative use and their classroom practice. Although teachers consider manipulatives to be pedagogically beneficial, they typically prefer traditional teacher-centred and paper-and-pencil instruction to manipulatives (e.g., Joutsenlahti & Vainionpää, 2010; Marshall & Swan, 2008; Toptaş et al., 2012). This finding suggests that the classroom adoption of pedagogically sound manipulatives may be hindered by their incompatibility with classroom practice.

A body of research reveals various possible hindrances to teachers' utilisation of manipulatives in their classrooms. Common hindrances related to day-to-day practice include manipulative availability, manipulative organisation, classroom management, and time factors (e.g., Bedir & Özbek, 2016; Hatfield, 1994; Marshall & Swan, 2008). Regarding availability, there is often a lack of money to acquire manipulatives, or there are limited numbers of manipulatives to go around. Typical problems with manipulative organisation include difficulty in borrowing and returning, setting up and packing away, sorting and storing, and damaging and losing manipulatives. The classroom management issues include crowded classes, students not listening to instructions, misuse of manipulatives, noisiness, and messiness. A lack of time is also an issue in terms of not having enough time to use manipulatives as well as to organise, set up, and pack them away.

2.3 Linear equation solving

Algebra plays a significant role in advanced mathematics learning (e.g., Kilpatrick et al., 2001; Knuth et al., 2006; Warren et al., 2016). The transition from arithmetic to algebra is usually challenging for students (Kilpatrick et al., 2001) due to a shift from concrete to more abstract concepts (Poon & Leung, 2010; A. Watson, 2009). To learn algebra, students need to shift from thinking about computing to arrive at concise answers to thinking about mathematical relations and their representations using letters and symbols (Kilpatrick et al., 2001; A. Watson, 2009). School algebra has traditionally focused on rules and procedures, so students mostly try to memorise rules and manipulate symbolic expressions and equations according to the rules without constructing their understanding (e.g., Bogomolny, 2007; Kilpatrick et al., 2001). Research has indicated that students who rely on rote learning often get confused, forget rules, and misapply them (Kilpatrick et al., 2001; A. Watson, 2009). Linear equation solving is one of the foundational domains within algebra curriculum, but it is often challenging for students to master equation solving (e.g., Kilpatrick et al., 2001; McNeil et al., 2019; Poon & Leung, 2010), particularly due to their inadequate understanding of concepts essential for equation solving (e.g., Booth & Koedinger, 2008; Knuth et al., 2006).

2.3.1 Key concepts in equation solving

Students with inadequate or incorrect conceptual understanding usually apply taught strategies to solve equations superficially (e.g., Booth & Koedinger, 2008). A good example is solving equations by performing the same operation on both sides of the equation. When applying this strategy, students usually solve equation $x + 2 = 6$ correctly by removing (i.e., subtracting) 2 from both sides of the equation: $x + 2 – 2 = 6 – 2$. Nevertheless, when solving equations such as $x – 2 = 6$, students who fail to recognise the meaning of the minus sign in front of number 2 are likely to remove (i.e., subtracting) 2 from both sides of the equation, resulting in $x = 6 – 2$. Previous studies have provided evidence that students' inadequate mathematical understanding and misconceptions hinder their learning and performance in equation solving (e.g., Booth & Koedinger, 2008; Knuth et al., 2006). To master linear equation solving, students need to understand various mathematical concepts, including equations, equivalence, different terms in an equation, and equation solving.

An equation is a mathematical statement in which two expressions are equal to each other. An equation has two sides (i.e., left and right) connected by an equal sign, which expresses an equivalence relation between its two sides. Although mathematical equivalence is an important concept for equation solving, it is difficult for students to understand (e.g., Knuth et al., 2006; A. Watson, 2009). Equivalence relations are usually represented in traditional school contexts as an arithmetic expression followed by an equal sign and then the answer to the arithmetic operation, for example, $5 + 1 = 6$ (e.g., Kilpatrick et al., 2001; McNeil et al., 2019). Students who emerge from school arithmetic usually view the equal sign as a signal to calculate rather than as a symbol of an equivalence relation between two sides of an equation, and thus write an answer to the calculation after the equal sign (e.g., Kilpatrick et al., 2001; Knuth et al., 2006; Sherman & Bisanz, 2009). Such misconceptions impede students' equation-solving competence (e.g., Kilpatrick et al., 2001; Warren et al., 2016; A. Watson, 2009) and usually persist as students advance in school (e.g., Knuth et al., 2006; Warren & Cooper, 2009). Students who interpret the equal sign as *operational* instead of *relational* tend to solve equivalence problems (also known as nonstandard equations), such as $5 + 1 = \_ + 2$ incorrectly by adding all the numbers before the equal sign (5 + 1 = 6 + 2) or adding all the numbers (5 + 1 = 8 + 2; McNeil et al., 2019). Research indicates that an understanding of equivalence not only promotes students' performance in algebra, including equation solving (Driver & Powell, 2015; A. Watson, 2009), it also assists them in learning new mathematical concepts (McNeil et al., 2019).

Understanding mathematical equivalence means that students not only have a relational understanding of the equal sign, but are also able to encode equations, identify both sides of the equation, view them as manipulatable entities, and understand that numbers and mathematical expressions can be represented in different comparable ways (e.g., McNeil et al., 2019; Rittle-Johnson et al., 2011). For example, the equation $5 + 1 = 6$ can also be represented as $6 = 6$, $5 + 1 = 4 + 2$, or $5 + 1 = x + 2$. Because equations usually include various terms (i.e., constants, unknowns, and coefficients), it is important to recognise the meanings of the mathematical symbols and letters that represent those terms (Poon & Leung, 2010; A. Watson, 2009). Particularly, the presence of the negative sign (e.g., $10 – 2x – 4 = 4$) and the presence of the unknown on both sides of the equation (e.g., $3x + 4 = 5x$) require a good understanding of equivalence and different terms in an equation (A. Watson, 2009). It is also essential to understand that the goal of solving an equation is to find the solution(s)—the value(s) of the unknown(s) that yields equivalence between two sides of an equation (Otten et al., 2019)—or to demonstrate that there is no real number-value solution to the equation.

2.3.2 Equation-solving approaches

Linear equations can be solved using *nonalgebraic* or *informal* approaches with no emphasis on the equivalence relation of both sides of the equation or using *algebraic* or *formal* approaches with their emphasis on mathematical equivalence (Knuth et al., 2006). Nonalgebraic approaches, such as trial-and-error substitution, the cover-up method, and the undoing method, are typically used to start equation solving (Kilpatrick et al., 2001). *Trial and error* or *guess and check* is an approach in which students try to substitute different values for the unknown in the equation until arriving at the value(s) that makes the equation true. In using the *cover-up* method to solve an equation, for example, $3x + 4 = 5x$, students start by covering up 4 and then thinking that $3x$ plus the cover up (i.e., 4) equals $5x$; therefore, 4 equals $2x$. After that they cover up x and then think that 2 times the cover up (i.e., x) equals 4; thus, x equals 2. In solving an equation using the *undoing* or *unwinding* method, students work backward by inversing mathematical operations and performing arithmetic calculations. For example, to determine the value of x in the equation $3x + 4 = 10$, students subtract 4 from 10 to get $3x = 6$ and then divide 6 by 3; so, x equals 2. Students who emerge from arithmetic often undo operations in reverse order to solve word problems, such as 'When 3 is added to 5 times a certain number, the sum is 38; find the number'; however, in algebra classes, they will be guided to first represent the relationships stated in the word problem as the equation $5x + 3 = 38$ (Kilpatrick et al., 2001, p. 262). Research has also found that students tend to use inverse operations to solve equations without understanding, and are thus likely to make errors (A. Watson, 2009).

Performing the same operation on both sides of the equation is an important algebraic equation-solving method that emphasises the equivalence of equations. In using this method to solve an equation, students make the same change (i.e., mathematical operation) to both sides of the equation to maintain equivalence between the two sides. For example, to solve the equation $5x + 3 = 38$, students first subtract 3 from both sides. This makes $5x + 3 – 3 = 38 – 3$, which after simplifying results in $5x = 35$. Then, students divided both sides by 5, that is, $5x/5 = 35/5$. After simplifying, it results in $x = 7$. Despite its contribution to equation solving, performing the same operation on both sides is usually not the first method taught to students (Kilpatrick et al., 2001).

The *canonical* method (i.e., a set of algebraic transformational rules to be carried out in a particular order) for solving equations has dominated algebra textbooks and gained special status in school algebra (Buchbinder et al., 2015). Buchbinder et al.'s (2015) online survey of teachers' perspectives on equation solving indicated that most teachers preferred canonical equation-solving solutions over noncanonical ones and perceived them as evidence of students' knowledge and mastery of equation solving. *Change side change sign* is one of the transformational rules used for solving equations. Students typically learn this rule by heart, that when a term is moved to the other side of the equal sign, its operation changes to the opposite. For example, in solving the equation $x + 2 = 6$, students move 'add 2' to the right side of the equation, in which it becomes 'subtract 2', resulting in $x = 6 – 2$. Students may be able to solve equations correctly by solely memorising and applying this 'magic' rule without necessarily understanding the inverse operation concept on which the rule is built (de Lima & Tall, 2008). However, without understanding why the inverse operation is valid (i.e., why the sign changes when the term moves to the other side), they usually encounter difficulties in solving more challenging equations (Capraro & Joffrion, 2006) and tend to misapply the rule (e.g., de Lima & Tall, 2008; Figueira-

Sampaio et al., 2009; A. Watson, 2009). For example, when solving equation $2x = 6$, students make variations of common errors: moving '2' to the right side of the equation, where it becomes 'subtract 2' resulting in $x = 6 – 2$ or correctly moving 'multiply by 2' to the right side, where it becomes 'divided by 2', but also changing a plus sign of the term to a minus sign resulting in $x = 6/(–2)$.

Otten et al. (2019) made the observation that although the change side change sign rule may appear to differ from performing the same operation on both sides, both methods rather resemble each other. For example, solving the equation $x – 4 = 6$ with the change side change sign rule means moving 'subtract 4' to the right side, in which it becomes 'add 4,' resulting in $x = 6 + 4$. When solving the equation by performing the same operation on both sides, 4 is added to the left and right sides of the equation, so that the equation becomes $x – 4 + 4 = 6 + 4$ and then $x = 6 + 4$. Thus, the key difference between these two methods is that the change side change sign rule takes a shortcut by employing the inverse operation of 'subtract 4' instead of adding 4 to both sides.

It has been argued that if students are taught to focus on the equivalence relation of both sides of the equation, they will learn to intuitively apply different methods to solve equations more effectively (A. Watson, 2009). Despite the pedagogical advantages of performing the same operation on both sides, empirical evidence shows that students typically use other equation-solving methods, including trial-and-error substitution, undoing, and change side change sign (e.g., Kilpatrick et al., 2001; A. Watson, 2009). Possible explanation for this could be that the equations they encounter can be easily or conveniently solved by other methods (Kilpatrick et al., 2001; A. Watson, 2009), or their previous knowledge of equation solving has an impact on how they learn and understand new methods (Kilpatrick et al., 2001). Kilpatrick et al.'s (2001) observation indicates that students who prefer the undoing method tend to experience difficulties in understanding the concept of performing the same operation on both sides, while students who prefer trial-and-error substitution and view equations as entities with a balance between both sides are likely to learn to solve equations easily by performing the same operation on both sides. Thus, trial-and-error substitution can be used to assist students in understanding the meaning of mathematical expressions and equations (A. Watson, 2009).

Equation solving is one of the key content areas of third-to-ninth-grade mathematics in the current Finnish NCC (EDUFI, 2016). Third-to-sixth graders should get to know the concept of the unknown and learn to solve equations by trial-and-error substituting values and reasoning for the unknown. Seventh-to-ninth graders should develop the competence to form and solve equations algebraically.

2.3.3 Learning equation solving

Research has demonstrated that students' misconceptions, particularly about the equal sign, can be altered, and as a result, their performance in equation solving is improved (Booth & Koedinger, 2008). To support equation-solving learning that enhances students' understanding and performance, mathematics education scholars have recommended various pedagogies, including learning through multiple representations (e.g., manipulatives), mathematical models, reasoning, reflection, and social interaction.

There has been empirical evidence that the use of multiple representations enhances students' understanding of algebraic concepts, such as equation solving (e.g., Warren et al., 2016; A. Watson, 2009). Warren et al. (2016) conducted a review of proceedings of the Conference of the International Group for the Psychology of Mathematics Education during 2005–2015, particularly studies regarding equality and inequality. They found that the types of representations that students experience can influence how they model equations and recognise equivalence relations. Algebraic relationships represented in nonsymbolic forms (e.g., with manipulatives, graphs, or diagrams) are usually easier for students to understand compared to the same relationships represented in symbolic forms (A. Watson, 2009). Sherman and Bisanz (2009) reported positive results from using nonsymbolic representations for solving a variety of equations. Second-grade students who solved equations represented with manipulatives (i.e., blocks and bins) had higher performance and were more likely to give relational justifications for their solutions than their peers who solved the same equations represented with mathematical symbols. Similarly, Driver and Powell's (2015) research results showed that second-grade students with and without mathematics difficulty performed much better in solving nonstandard equations presented with pictures and stories compared to the same equations presented with mathematical symbols. Sherman and Bisanz (2009) also discovered that second-grade students' experience of equations represented with manipulatives facilitated their understanding, and thus enhanced their performance on solving subsequent symbolic equations. Their findings support mathematics educators' (e.g., Goldin, 2002; A. Watson, 2009) view that learning through multiple representations helps students to overcome difficulties in understanding the symbolic expressions of abstract algebraic notions.

Goldin (2002) gives an example that familiar concrete objects, such as a bag of not-yet-counted objects, can be used to introduce students to the concept of using the letter x to stand for a specific unknown number. Therefore, $x + 5$ refers to the results of counting the objects inside the bag and adding five more, while $5x$ refers to the number of objects in five uniform bags, and so on. When using a representation, the meaning of the mathematical ideas under investigation needs to be emphasised to students (Foster, 2007). Students should be encouraged to compare and connect different representations of algebraic structures (A. Watson, 2009), and later, to generalise and move away from nonsymbolic representations (Goldin, 2002). The goal is that eventually, students can work with symbolic representation of algebraic structures (e.g., $x + 5$ vs. $5x$) flexibly and spontaneously (Goldin, 2002).

New technologies can play a significant role in the learning of algebra (Warren et al., 2016; A. Watson, 2009). Technologies make multiple representations available to students and provide bridges between the representations and the underlying algebraic concepts (A. Watson, 2009). Technologies help students not only to investigate, test, reason, and understand the meaning of algebraic expressions, but also to solve equations and work in different interesting and motivating settings (Warren et al., 2016). Substantial evidence demonstrates that students who have used technological tools for their algebra learning develop a stronger understanding of algebraic expressions, equations, and equation solutions compared to their peers who have used only paper and pencil (A. Watson, 2009). Thus, technologies should be used in an active and participative way as learning tools to support students in learning algebra with understanding (Warren et al., 2016; A. Watson, 2009).

Manipulatives can be used as hands-on learning tools to make algebra more concrete (Foster, 2007). Research indicates that students do not automatically develop algebraic understanding by merely using manipulatives; rather, they need to use manipulatives as support for constructing their understanding through experiences with algebraic concepts (Foster, 2007). Working with manipulatives over time can enable students to confront their misconceptions (Foster, 2007) and build knowledge of algebraic relationships and structures (A. Watson, 2009).

A body of research reports positive results of using manipulatives for learning linear equations, particularly by students without prior knowledge of equation solving. High performance and relational justifications of second-grade students who solved equations represented with PMs in Sherman and Bisanz's (2009) study (mentioned earlier) support the use of manipulatives in learning algebra. The findings suggest that due to not being associated with misconceptions (e.g., the equal sign) as mathematical symbols are, the manipulatives may allow students to concentrate on the relationships between quantities on both sides of the equations, thereby helping them develop their equivalence understanding. In Suh and Moyer's (2007) study, third-grade students worked with either PM (i.e., Hands-On Equations®; Borenson & Associates, n.d.) or VM (i.e., Algebra Balance Scales applet; National Library of Virtual Manipulatives [NLVM], n.d.) to solve simple equations using nonalgebraic strategies for one week. Their research results revealed that students in both instructional groups gained significant achievement between pretest and posttest, and showed representational fluency, which indicated their algebraic, particularly equivalence, understanding.

Figueira-Sampaio et al. (2009) developed a virtual balance to replace traditional physical balances, which usually pose practical challenges in Brazilian classrooms. Physical balances often require much preparation for their balance mechanism and also prohibit students' exploration and interaction due to difficulty in maintaining equilibrium, and a lack of physical balances in schools. After theoretical classes related to linear equations, sixth-grade students learnt to solve five equations by either using the developed virtual balance by themselves or observing the physical balance manipulated by their teacher. Their research findings indicate that virtual balance successfully resolves practical classroom challenges because it does not require any mechanical preparation, enables students' direct interaction, and promotes their social interaction in small groups.

Magruder and Mohr-Schroeder (2013) studied seventh-grade students using NLVM Algebra Balance Scales applet (also used in Suh and Moyer's [2007] study) to solve 40 equations within a week at their own pace. Their research findings indicate that the VM not only enhances students' understanding of equal signs and algebraic symbols, but also promotes their procedural knowledge. In another study, Magruder (2012) studied three groups of sixth-grade students learning to solve equations either without manipulatives, with PM (i.e., algebra tiles), or with VM (i.e., NLVM Algebra Balance Scales applet, also used in her and Mohr-Schroeder's study) over 10 instructional days. She found that PM and VM appeared to be effective tools for solving equations. Nevertheless, on the posttest, students learning equation solving without manipulatives outperformed those learning with manipulatives. According to Magruder, the cognitive overload that PM and VM students possibly experienced when learning with manipulatives may explain their underperformance on the posttest. She suggested that enough time should be reserved for students to develop their conceptual understanding through manipulatives.

Based on her research literature review on how children learn algebra, A. Watson (2009) argued that the use of concrete materials (e.g., lengths of rods or areas of tiles to represent unknown values) can bridge students' prior experience and abstract relationships, and thus enable them to focus on relations instead of numbers. Nevertheless, most manipulatives have limitations, particularly with negative numbers, fractional values, and division operations that are difficult to represent concretely. When values and operations cannot be represented with manipulatives, students have to detach themselves from manipulatives and move towards abstraction. Their realisation of the limitations of manipulatives actually encourages their development of algebraic thinking.

Mathematical models can be used to help students gain access to abstract concepts, such as algebra. A balance model has commonly been used for teaching and learning linear equations (e.g., Figueira-Sampaio et al., 2009; Otten et al., 2019; Warren & Cooper, 2009). An in-balance stage of a scale represents the rational view of the equal sign; weights on both sides represent mathematical expressions on both sides of the equation (e.g., Otten et al., 2019; Warren & Cooper, 2009). To benefit from the model, students must understand the shared principle of an equation and a balance scale (i.e., both sides are equal) and the similarity between mathematical operations (i.e., addition, subtraction, and division) and physical actions (i.e., adding, removing, and partitioning objects on both sides of the scale; Foster, 2007). Experiences with the balance model help students understand the equal sign as indicating equivalence (Foster, 2007), view both sides of the equation as two equal entities (rather than an instruction to find an answer; Warren & Cooper, 2005), and develop strategies for solving equations by doing operations to the equation while preserving its equality (Foster, 2007).

Otten et al. (2019) conducted a systematic literature review of 34 peer-reviewed journal articles in which the balance model was utilised for linear equation lessons. They found that the most common rationale for using the balance model was to support students' understanding of the equality concept, followed by enabling students' learning through physical experiences and the use of models and representations. The model was used for students ranging from kindergarten to Grade 9 to learn to solve equations. More positive effects on learning outcomes were reported for using physical or virtual balances for young students (from kindergarten to Grade 6) with no previous algebra experiences to explore the equality concept through balances and solve simple equations with mostly positive values and additions. In contrast, more mixed and negative effects on learning outcomes were reported on using drawn balances for students with previous algebra experiences (generally from Grade 7 upwards) to model, transform, and solve equations with negative values and subtraction—in other words, to revitalise their foundation of equation solving.

While the balance model appears to be beneficial for learning equation solving, the model is limited by its inability to represent equations with negative integers, such as $x = -1$ (simplified from $x + 3 = 2$), and with subtraction, such as $x - 3 = 2$ (e.g., Otten et al., 2019; Warren & Cooper, 2009). It is difficult to construct a meaningful representation of such equations due to the balance embodiment involving physical weights (de Lima & Tall, 2008; Otten et al., 2019). Some solutions to the limitations of the balance model have been proposed. For example, taking off 3 on the left side of the equation $x - 3 = 2$ can be represented by adding 3 to the right side of the balance scale, or by using helium balloons to lift up the left side of the scale, thereby acting as a subtraction of 3 (de Lima & Tall, 2008). Another solution is that a diagram representing a balance scale can include all mathematical operations and numbers (Warren & Cooper, 2009). Rather than finding ways to represent the negative sign, as concrete models no longer work for this, it may be time to shift to abstract meanings of operations and relations (A. Watson, 2009). Therefore, the models should be viewed as analogues for students to generalise their understanding to more abstract situations, for example, solving equations containing negative quantities and subtraction (Warren & Cooper, 2005).

Overall, multiple representations and models can be used to assist students' learning of equation solving. However, what is meant to be learnt through representations and models may not be self-evident to students. Moreover, relying heavily on concrete representations and models may hinder students from generalising (i.e., transferring knowledge to other contexts). It has been recommended that self-explanation and reasoning (Sherman & Bisanz, 2009), social interaction (Foster, 2007; Warren & Cooper, 2009), and reflection (McNeil et al., 2019; Warren & Cooper, 2009) can encourage students to develop their algebraic understanding, detach from the representations and models, and generalise what they have learnt.

3 METHODOLOGY

This doctoral study employed an EDR approach to bridge research on manipulatives and its direct benefits to real-world educational practice. This study was conducted in Finnish comprehensive schools through a 6-year iterative enquiry. The empirical data were collected and analysed using mixed methods research. After presenting the research methodology of this study, I discuss how research ethics and integrity were considered throughout the study.

3.1 Educational design research approach

Since being introduced to educational research in the 1990s (see Brown, 1992; Collins, 1992), EDR has become recognised as a research approach seeking to bridge the separation between theoretical research and practice in education (e.g., Anderson & Shattuck, 2012; Collins et al., 2004; McKenney & Reeves, 2019). It is placed in Stokes' (1997) *Pasteur's quadrant* for use-inspired basic research aiming to advance both theoretical understanding and practical application (e.g., Barab, 2006; Phillips, 2006). Starting out with complex educational problems in practice, EDR attempts not only to develop design solutions to improve educational practices, but also to better understand teaching and learning through iterative processes in actual educational settings (e.g., McKenney & Reeves, 2019; Plomp, 2013). Therefore, EDR appeared to be the right research approach for this doctoral study to resolve shortcomings regarding the current research on manipulatives and its benefits for real-world practice.

McKenney and Reeves (2019) used the term *EDR* for a family of approaches that has the dual goal of contributing to both educational theory and practice. In addition to EDR, other commonly used terms, including *design-based research*, *design research*, *design experiments*, and *development(al) research*, also belong to this family. They are, however, not entirely synonymous and have somewhat dissimilar characteristics (for examples of how different scholars have characterised EDR, see Publication I). According to the observation in Publication I, key characteristics of EDR commonly defined among EDR scholars include (1) development of practical solutions to educational problems, (2) development of theories, (3) evolution through multiple iterations, (4) undertaken in real-world educational settings, and (5) utilising various research methods, often mixed methods.

Practical solution development. EDR intends to improve educational practice directly by developing research-based practical solutions (also known as *interventions*) to educational problems in the real world. In this dissertation, I use the term *(design) solution* when referring to the practical outcome of EDR to avoid ambiguity with *class interventions* (i.e., experimental designs) used in this study for collecting empirical data in classrooms. Common types of solutions include educational products and environments (e.g., textbooks and educational games), processes (e.g., class activities and teaching approaches), programmes (e.g., courses and learning units), and policies (e.g., curricula and assessment protocols; see e.g., Anderson & Shattuck, 2012; McKenney & Reeves, 2019; Plomp, 2013). The solution development is informed by theoretical understanding, problem analysis, and solution testing in real-world educational settings (McKenney & Reeves, 2019). In this present study, a TM, student worksheets, teacher guides, and class activities were developed as a solution for promoting students' understanding of equation-solving concepts and classroom practice. Its design and development are described in Chapters 4.2–4.3 (for more details, see Publications III and IV).

Theory development. EDR also contributes to the research community by advancing usable and generalisable knowledge constructed during iterative empirical investigation (e.g., Edelson, 2006; Kelly, 2006). According to Edelson (2002), EDR can assist in developing three types of theories: domain theories, design frameworks, and design methodologies. A *domain theory* yields two types of knowledge: a *context theory* describing the education problem to be solved and the real-world educational setting of interest, and an *outcomes theory* describing outcomes of implementing the developed solution. Informed by the outcomes theory, a *design framework* describes the characteristics of a successful solution in a particular educational setting. A *design methodology* provides guidelines for undertaking EDR to achieve the research objectives. In short, design frameworks inform others on how to develop a solution to a similar educational problem in other contexts, while design methodologies advise others on how to successfully conduct EDR.

Although Edelson's (2002) three types of theories were proposed almost two decades ago, they still inform recent EDR (e.g., Alshabeb, 2020; Fox, 2018; Kerslake, 2019). McKenney and Reeves (2019) argued that the theoretical contributions of EDR are used for various purposes: descriptive, explanatory, predictive, and prescriptive/normative understanding. *Descriptive* understanding is often developed at the beginning of EDR to describe real-world phenomena (i.e., the educational problem and its situated context), while *explanatory* understanding explains why or how particular phenomena occur (cf. context theory; Edelson, 2002). *Predictive* understanding can be pursued by testing whether, how, and to what extent the developed solution produces desirable outcomes (cf. outcomes theory; Edelson, 2002). *Prescriptive* understanding combines descriptive, explanatory, and predictive understanding to direct the design of successful solutions in similar contexts (cf. design framework; Edelson, 2002). Not all EDR intends to contribute to all purposes. This study seeks to develop, particularly descriptive, explanatory, and prescriptive understandings that result in theoretical outcomes. The theoretical outcomes of this study are presented according to Edelson's (2002) three types of theories in Chapter 7.1. The theoretical outcomes were also published as follows: context theory in Publications II and IV, outcomes theory in Publications II–IV, and design framework and design methodology in Publication IV.

Multiple iterations. The EDR process is naturally iterative (e.g., Anderson & Shattuck, 2012; McKenney & Reeves, 2019; Plomp, 2013). Similarly to design practice, during EDR, a design solution gradually evolves through iterative cycles of investigation, design, and evaluation until the research goals are achieved. Concurrently, usable and generalisable knowledge is gradually constructed through cyclical investigation and reflection. Phases addressed in different EDR models, such as those of Easterday et al. (2017) and Reeves (2006), may vary in details. However, according to McKenney and Reeves (2019), three main phases can be drawn from most models: (1) initial phase (i.e., analysing and investigating the current situation, cf. preliminary research; Plomp, 2013), (2) design phase (i.e., designing and prototyping the solution, cf. development or prototyping phase; Plomp, 2013), and (3) evaluation (i.e., assessing the solution and reflecting on refinement of the solution, cf. assessment phase; Plomp, 2013). McKenney and Reeves (2019) described EDR as an overall study consisting of at least one subcycle from each of the following: analysis and exploration, design and construction, and evaluation and reflection. Most EDR, including the majority of the dissertations reviewed in Publication I, involves multiple subcycles of design and construction, and evaluation and reflection. The iterations of this dissertation are described in Chapters 3.3 and 4 (for more details, see Publication IV).

Undertaken in real-world educational settings. An empirical enquiry in naturalistic real-world (as opposed to controlled laboratory) settings is an essential part of EDR (e.g., McKenney & Reeves, 2019; Sandoval & Bell, 2004). Findings from a lab-like setting inadequately contribute to a holistic understanding of classroom reality, in which learning actually takes place in a dynamic and rich social context (e.g., Brown, 1992; Collins, 1992; Juuti & Lavonen, 2006; McKenney & Reeves, 2019). Thus, being conducted in an authentic environment ensures that the solution developed during EDR takes into account the complexity and messiness of the educational practice it is intended to improve (e.g., Barab, 2006; Plomp 2013). As a result, solution utility in the real world is promoted (Anderson & Shattuck, 2012). Moreover, different stakeholders should be involved in EDR to inform the development of the design solution towards the desirable outcomes (e.g., Anderson & Shattuck, 2012; Barab, 2006; Ørngreen, 2015). In this study, an empirical investigation and evaluation was conducted in Finnish comprehensive schools (see Chapters 3.2, 4.1.3, 4.2.2, and 4.3.2). The data were collected from both the teachers and their students, who were the target users of the design solution. In particular, class interventions were implemented in classroom situations in which teaching and learning took place through social interactions between teachers and students, as well as among peers. In addition to the direct involvement of teachers and students, the perspectives of policymakers, that is, the current Finnish NCC (EDUFI, 2016), also informed the solution development.

Utilising various research methods. EDR should involve the triangulation of data sources (e.g., students, teachers, and policymakers), data collection methods, data types (i.e., qualitative and quantitative), theories, and researchers/investigators (e.g., Design-Based Research Collective [DBRC], 2003; McKenney & Reeves, 2019; Plomp 2013). The triangulation not only helps to better understand open real-world problems (Kelly, 2013), it also enhances the validity and reliability of EDR (e.g., McKenney & Reeves, 2019; Plomp, 2013). EDR is not methodologically bound. The choice of research methods is based on which method is accurate and productive for addressing the research questions (McKenney & Reeves, 2019). EDR often requires mixed methods combining the benefits of qualitative and quantitative methods (e.g., Anderson & Shattuck, 2012; Cobb et al., 2003), as in the case of this study (see Chapters 3.4, 4.1.3, 4.2.2, and 4.3.2).

Publication I systematically reviewed 21 Finnish EDR doctoral dissertations on mathematics, science, and technology education published during 2000–2018. The aim was twofold: (1) to investigate how other Finnish researchers who conducted their dissertations in a similar field as this doctoral study employed EDR and what challenges they encountered, and (2) to draw implications for conducting this study. Common EDR characteristics (e.g., research in actual educational contexts, evolution through multiple iterations, and yielding both practical and theoretical contributions) were discovered in all the dissertations. Furthermore, in line with those previously described by others (e.g., Kelly, 2013; McKenney & Reeves, 2019), most researchers were faced with challenges, such as conducting EDR with limited resources and collaborating with multi disciplines. The implications from the review (e.g., use of both classic and up-to-date EDR literature, research triangulation, multidisciplinary collaboration, working with alternative designs, and emphasis on design activities and processes) were used to guide this study (for more details on the implications, see Publication I).

3.2 Research context

This study was conducted in comprehensive schools (i.e., Grades 1–9 of basic education) in southern Finland. Third- to sixth-grade mathematics classrooms were purposefully chosen as the primary research context for two reasons. First, according to Flyvbjerg (2011), a *critical case* (i.e., a most-likely or least-likely case) usually reveals rich and insightful information related to the phenomenon of interest, which a representative (i.e., an average or typical) case may not be able to provide. A critical case is likely to generalise that 'if it is valid for this case, it is valid for all (or many) cases' (p. 307). In this study, with limited resources, it was expected to be productive to focus on a critical case, the upper primary grades, where manipulative use declines remarkably (e.g., Marshall & Swan, 2008; Moyer-Packenham et al., 2013; Uribe-Flórez & Wilkins, 2010). The investigation of this particular setting would contribute to an in-depth and thorough understanding of what causes the decline in manipulative use in primary school classrooms. This holistic understanding would then help to develop a design solution to the problem. Moreover, if the developed manipulative could promote students' understanding of equation-solving concepts and classroom practice in this particular setting, then it is likely that the manipulative would succeed in the lower primary grades, where the use of manipulatives is typically higher.

Second, in Finland, the equity and quality of the education system and the homogeneity of teacher quality made it possible to carry out research in any school. According to the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA) 2015 and 2018 data, Finnish students' socioeconomic background has a low impact on their mathematics performance (OECD, 2017, 2020). Qualified class teachers in Finnish primary schools, who were 93% of the teachers employed, according to principals' reports in PISA 2018, are required to have a master's degree in education (OECD, 2020). It could therefore be claimed that Finnish schools have a high level of uniform quality across the country.

In Finland, class teachers teach all school subjects, including mathematics at the primary level (Grades 1–6), while mathematics teachers provide instruction for mathematics at the lower secondary level (Grades 7–9). According to the Teaching Qualifications Decree 986/1998, to work as a qualified teacher in basic education, both class and mathematics teachers must have completed a master's degree, at least 60 credits under the European Credit Transfer and Accumulation System (ECTS credits) of pedagogical studies, and either at least 60 ECTS credits of multidisciplinary studies in school subjects taught in primary schools for class teachers or at least 60 ECTS credits of mathematics studies for mathematics teachers (Finlex, 2021).

The current Finnish NCC (EDUFI, 2016) has become less detailed and prescriptive, functioning as a framework instead of a roadmap. It leaves teachers with a considerable degree of autonomy to interpret the NCC and decide what and how they will teach. For example, teachers can freely select their textbooks and other instructional materials and plan their lessons.

3.3 Multiple-phase research process

This study employed EDR as a research approach to resolve the separation between research on manipulatives and its educational practice. EDR contributed not only to a better understanding of how manipulatives could be used to promote students' equation-solving concepts understanding and classroom practice, but also to the development of such a manipulative. Figure 2 outlines how the study was undertaken through multiple iterations of investigation, design and construction, and evaluation and reflection, as recommended by McKenney and Reeves (2019). The overall EDR process of the study consisted of three phases (i.e., initial research, concept development, and design development), which were divided into six subcycles. Although the process flow in Figure 2 moves from left to right, the actual process progressed in a nonlinear manner, which, according to Plomp (2013), resulted from one element repeatedly fed into others (i.e., iterative), and some subcycles were revisited (i.e., flexible). The overall process of the study is also summarised in Publication IV. The process, research methods, and results of each phase are presented in Chapter 4.

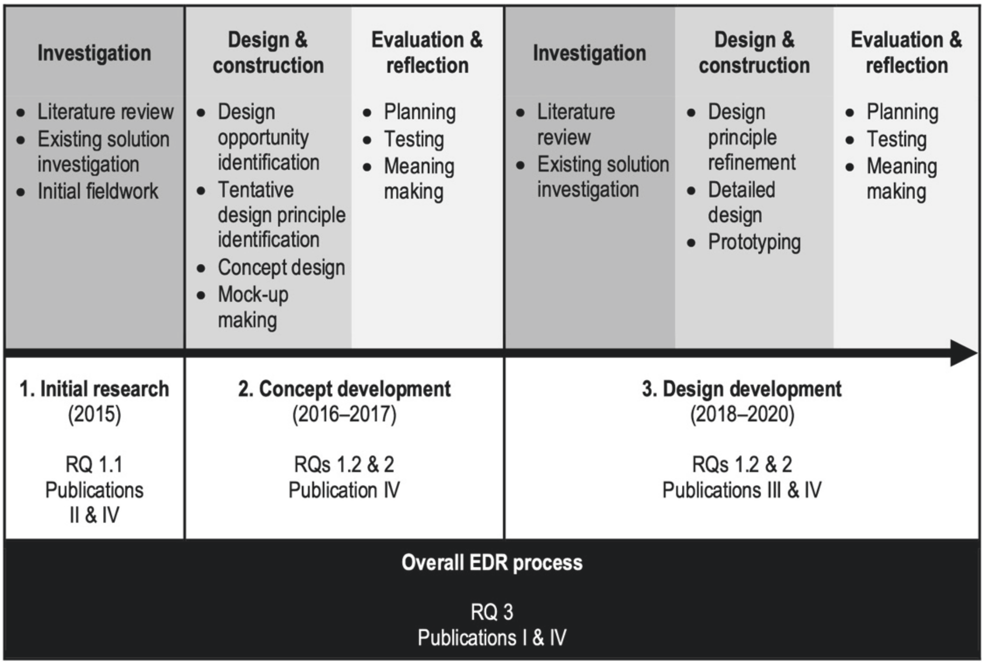


Figure 2. Overall EDR process, research time frame, and corresponding RQs and publications. Adapted from 'Constructing a Design Framework and Design Methodology from Educational Design Research on Real-World Educational Technology Development,' by D. Lehtonen, 2021, *Educational Design Research,* *5*(2), Article 38, p. 4 (https://doi.org/10.15460/eder.5.2.1680). CC BY 4.0.

During Phase 1, initial research was conducted to theoretically and empirically construct a context understanding of the educational problem and the target real-world setting. The focus was on what should be considered when developing a solution to the educational problem. The areas of investigation included the needs, challenges, and opportunities of using manipulatives for learning equation solving in classrooms and the strengths and limitations of existing manipulatives (RQ 1.1). The investigation started with a literature review to establish a theoretical background understanding (see Chapter 4.1.1), followed by an analysis of existing manipulatives and educational games for equation solving (see Chapter 4.1.2). Based on the findings gained from both investigations, initial fieldwork was conducted in real classrooms with teachers and their students (see Chapter 4.1.3).

In Phase 2, the context knowledge derived from Phase 1 was used to identify design opportunities and tentative design principles, which initially addressed RQ 2. Based on that, various potential design solutions to the educational problem were explored and developed into four manipulative concepts (see Chapter 4.2.1). A nonfunctional mock-up of each concept was constructed and then evaluated by teachers (see Chapter 4.2.2). The most promising concept was selected for further development based on the teachers' responses: how each concept could promote students' understanding of equation-solving concepts and encourage teachers to adopt it in their classrooms (an initial outcome knowledge answering RQ 1.2).

At the beginning of Phase 3, another literature review and investigation of existing educational products relevant to design development were conducted. The knowledge derived from the literature review and the educational product investigation was used together with Phase 2 findings (i.e., initial outcome knowledge) to refine the design principles, which helped address RQ 2. The design principles then informed the development of the design solution: a TM, student worksheets, teacher guides, and class activities as a solution to the identified problem (see Chapter 4.3.1). The next step was prototyping the design solution. The TM, particularly its technological parts, was developed and prototyped in collaboration with a team of computer science students and their supervisor. Finally, the developed solution was tested by teachers and students (see Chapter 4.3.2) to answer RQ 1.2 and inform the refinement of the design solution.

3.4 Mixed methods research design

Mixed methods research (also known as *integrated research*, *combined research*, *hybrid research*, and *mixed research*) utilises two types of data—qualitative and quantitative—and integrates them in meaningful ways to rigorously investigate the phenomenon of interest (e.g., Creamer, 2018; Creswell & Plano Clark, 2017; Teddlie & Tashakkori, 2012). Throughout the three phases of this doctoral study, mixed methods research was used as a strategy of enquiry for collecting, analysing, and interpreting the empirical data. The combined qualitative and quantitative research is an effective enquiry for tackling complex real-world problems (Creamer, 2018). The mixed methods design used in this study was a *convergent design,* sometimes known as a *concurrent*, *parallel*, or *simultaneous design*. It is a type of design in which both qualitative and quantitative data are simultaneously collected and analysed, then compared and/or combined to obtain a better understanding of the findings than those provided by qualitative or quantitative findings separately (Creswell & Plano Clark, 2017). The intent of convergent design is that the advantages of one approach complement the shortcomings of the other.

As mentioned in Chapter 3.1, EDR is not bound to any specific methodology. The choice of a mixed methods research design was underpinned by the *pragmatism* paradigm. To determine the research methods employed, pragmatism places its priority on the question asked, rather than the method or the philosophical worldview underlying the method, and what works best for addressing the RQ is used (Creamer, 2018; Creswell & Plano Clark, 2017; Teddlie & Tashakkori, 2012). Among other paradigmatic stances (e.g., postpositivism, constructivism, and transformism) that provide a foundation for the use of mixed methods, pragmatism has been endorsed by scholars as the best worldview for mixed methods research (Tashakkori & Teddlie, 2003), particularly in applied disciplines like educational sciences (Creamer, 2018).

In this EDR, qualitative and quantitative data were used together over time to better understand the educational problem and context and to support the development and evaluation of the design solution. Each phase emphasised different research objectives, and the data were collected and used accordingly. The qualitative data, consisting of interviews, observations, and thinking aloud, were intended for various purposes in different research phases. During Phase 1, they were used to address the challenges and opportunities of using manipulatives in primary school classrooms and to explore the strengths and limitations of existing manipulatives. Later, the qualitative data were utilised to explain why and how the developed manipulative succeeded or failed in enhancing students' conceptual understanding and classroom practice. The quantitative data, including paper-based tests, questionnaires, and self-evaluation, were also used for several purposes. During Phase 1, they were intended to justify whether it was worth adopting manipulatives into the mathematics classroom. Later, they were used to inform the determination of the generated manipulative concepts during Phase 2 and to determine the success of the developed manipulative in real-world settings during Phase 3.

In short, the quantitative and qualitative strands were separately implemented according to their contributions to the RQs. Then, the findings from the separate strands were converged and compared to obtain a more complete understanding of the phenomena under study. Table 1 outlines the mixed methods research design and corresponding RQs of three empirical subcycles: initial fieldwork, concept evaluation, and design evaluation.

Table 1. The mixed methods research design and corresponding RQs of the study's empirical sub-cycles. Adapted from 'Constructing a Design Framework and Design Methodology from Educational Design Research on Real-World Educational Technology Development,' by D. Lehtonen, 2021, *Educational Design Research, 5*(2), Article 38, p. 5 (https://doi.org/10.15460/eder.5.2.1680). CC BY 4.0.

Empirical sub-cycles	**RQs**	**Data**	**Analysis**
Phase 1: Initial fieldwork	1.1 & 2	• Teacher interviews (*N* = 4)	• Inductive content analysis
		• Class intervention observations (teachers *N* = 4, students in paper-and-pencil group *n* = 25, in PM group *n* = 25, in VM group *n* = 24)	• Inductive content analysis
		• Student paper-based tests (paper-and-pencil group *n* = 25, PM group *n* = 25, VM group *n* = 24)	• Descriptive statistical analysis • Inferential statistical analysis (95% confidence intervals)
		• Student self-evaluations (paper-and-pencil group *n* = 25, PM group *n* = 25, VM group *n* = 24)	• Descriptive statistical analysis
Phase 2: Concept evaluation	1.2 & 2	• Teacher questionnaires (*N* = 12)	• Descriptive statistical analysis
		• Teacher interviews (*N* = 12)	• Inductive content analysis
Phase 3: Design evaluation	1.2 & 2	• Class intervention observations (teachers *n* = 2, students in paper-and-pencil group *n* = 12, in TM group *n* = 12)	• Inductive content analysis • Deductive content analysis • Descriptive statistical analysis • Inferential statistical analysis (Pearson's chi-squared test)
		• Student paper-based tests (paper-and-pencil group *n* = 12, TM group *n* = 12, comparison group without participating in class intervention *n* = 65)	• Descriptive statistical analysis • Inferential statistical analysis (Mann–Whitney U test)
		• Thinking aloud sessions of students in TM group (*n* = 12)	• Inductive content analysis
		• Questionnaires of students in TM group (*n* = 12)	• Descriptive statistical analysis
		• Interviews of students in TM group (*n* = 12)	• Inductive content analysis
		• Teacher questionnaires (*N* = 6, two participated in the class interventions)	• Descriptive statistical analysis • Inferential statistical analysis (Wilcoxon matched-pairs signed-rank test)
		• Teacher interviews (*N* = 6, two participated in the class interventions)	• Inductive content analysis

It is noteworthy that the empirical research directly addressed RQ 1 and assisted in building the design framework (RQ 2). Guidelines for conducting successful EDR

(RQ 3) were constructed based on lessons I learnt from undertaking the entire EDR (see Chapter 5).

Altogether, 18 basic education teachers (teaching experience 3–27 years), 98 primary school students (aged 9–12), and 65 lower secondary school students (aged 13–16) participated in different research phases. The research design of the initial fieldwork, concept evaluation, and design evaluation are separately described in Chapters 4.1.3, 4.2.2, and 4.3.2, respectively.

Conducting mixed methods research is challenging, particularly for novice researchers (Creamer, 2018; Creswell & Plano Clark, 2017). Mixed methods research requires researchers to have sufficient skills in quantitative, qualitative, and mixed methods research, as well as research software for each type of research data. It is also resource intensive: qualitative and quantitative data collection and analysis require extensive time and resources. Nevertheless, there is evidence that it is feasible for a graduate student to undertake mixed methods research independently (Creamer, 2018; Teddlie & Tashakkori, 2012). This was the case with this doctoral study.

As a qualitatively orientated researcher acquainted with mixed methods research, I was required to develop a better understanding of quantitative research and the necessary skills to conduct this study independently. I took various courses in quantitative and mixed methods research and engaged in the literature. I also consulted my supervisors and experienced quantitative researchers in the faculty about subjects, such as measurement instruments, statistical analyses, and rigour in quantitative research.

I conducted this doctoral study alongside my full-time work. The study was not part of any research project and thus did not have any specific scope and time constraints. At the beginning, the scope of the study was planned to be manageable within a reasonable time frame and resources. As the study progressed, the scope was altered due to the evolutionary nature of EDR (McKenney & Reeves, 2019): the results from the previous phase informed how the subsequent phase was conducted. Consequently, the scope became rather overambitious for a solo doctoral researcher. The multiple-phase mixed methods design required considerable time to obtain approval for conducting research in classrooms, recruit participants, collect qualitative and quantitative data, process (e.g., transcribe and translate intervention videos) and analyse both data types, and integrate them. Fortunately, the study was still feasible because qualitative and quantitative data could be conveniently collected simultaneously during school visits. There were also some overlaps of participants in the qualitative and quantitative research, and some RQs enabled a comparison of the overlapping qualitative and quantitative data analysis. Nevertheless, the study spanned 6 years.

Through this doctoral study, I broadened my research skillsets of quantitative, qualitative, and mixed methods towards becoming a versatile researcher. Although I am still an early-staged quantitative and mixed methods researcher, my expertise in both research methods has gradually developed over the course of this study (see e.g., Publication II vs. Publication III). I have become confident in flexibly using various research methods to best address different RQs.

3.5 Research ethics and integrity

Research ethics and integrity were taken into account throughout the research. The research followed the ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review of the Finnish National Board on Research Integrity (TENK, 2009), the responsible conduct of research and procedures for handling allegations of misconduct in Finland (TENK, 2012), and the European code of conduct for research integrity of All European Academies (ALLEA, 2017).

At the time the research was planned and conducted, the university was committed to TENK's ethics guidelines (2009). Thus, the research was required to comply with three areas of ethical principles in the humanities and social and behavioural sciences stated in the guidelines: (1) respecting the autonomy of research subjects, (2) avoiding harm, and (3) privacy and data protection. The research was planned according to the guidelines and then checked to determine whether an ethical review by the Ethics Committee of the Tampere Region was needed. According to the national guidelines and the committee's instructions, an ethical review was not required because the research did not deviate from the principle of informed consent, it was conducted as part of normal classroom activities with minors' parental consent, and it did not involve an intervention in the physical integrity of the research participants or expose them to mental harm beyond the risks encountered in daily life.

Prior to the study, research permission was obtained from the town's children and youth service director and school principals. Research participation was voluntary and based on informed consent of the participants or their legal guardians in the case of students aged under 15. An information sheet containing clear and sufficient information (e.g., research topic and goals, data collection methods and procedure, estimated duration of the data collection, participants' rights, data storage and utilisation, confidentiality, data protection, research dissemination, and the researcher's contact information) was provided to the participants and their guardians.

During the data collection, the research sites and participants were treated politely and with respect and care. The research data were collected responsibly by minimising interruptions in normal classroom activities; for example, the students took turns participating in the class interventions. While their teacher and classmates took part in the class intervention, the rest of the class studied independently according to their normal class plan. The content covered during the class interventions was also designed to align with the Finnish NCC (EDUFI, 2016). Classroom interventions were undertaken only in Phases 1 and 3 to avoid unnecessary classroom interruptions. The Phase 1 interventions were important for constructing an understanding of the educational problem and contexts, and in Phase 3, the design solution was mature enough for classroom implementation and evaluation. In contrast, in Phase 2, the design concepts were underdeveloped, and thus were evaluated only through teacher questionnaires and interviews instead of class interventions.

The data collection was designed to avoid mental and social harm to the participants. For example, pretests were excluded from the research because completing a pretest in a new content area is usually unfamiliar to Finnish primary school students and can cause them some stress and frustration. Moreover, primary schoolers in Finland typically do not obtain knowledge regarding untaught content (in this case, equation solving) from outside the school. It could thus be assumed that the students in this study had no/low prior knowledge of equation solving. For this reason, only posttests were sufficient for evaluating students' learning achievement of new knowledge after the class interventions. Another example was that all students were individually interviewed; however, their own teacher's presence during the interviews was allowed upon request for students' mental support. No unnecessary personal data from the participants, such as their racial or ethnic origin, was collected. The participants were also provided with the opportunity to ask questions concerning the research or withdraw from the research at any point.

During data processing, analysis, and interpretation, the participants' privacy was protected responsibly. Their personal data, unnecessary for the research, were removed from the stored data. The research materials containing their identifiers were carefully and confidentially stored and will be destroyed after the dissertation.

Additionally, the data were used only for research purposes, as stated in the informed consent. No one besides me had access to the data.

During research dissemination, the research publications sought to balance the participants' confidentiality with the openness of science and research. The research results were presented with respect for the participants and without bias. The participants' privacy was ensured through the anonymity and nonidentifiability of participants in all research publications. For example, quotations from the data were published anonymously; the participants were presented in an unidentifiable way; their identifiers (e.g., school name and location, age or teaching experience, and grade levels) were reported at a general level, such as in range; and their identifiers irrelevant to the research results were neither used for the data analysis nor presented in any publications. Moreover, the quantitative data were statistically analysed, and the results were then reported. It is not possible for the audience to identify any individual participants, even if the published findings are based on research data containing their identifiers.

Additionally, the research was conducted according to TENK's (2012) and ALLEA's (2017) guidelines to comply with the responsible conduct of the research and to avoid violations of research integrity. Good research practices are based on four principles of research integrity: reliability, honesty, respect, and accountability (ALLEA, 2017). Various good research practices are the same as those mentioned in the previous paragraphs regarding research ethics. Apart from that, various practices were undertaken throughout the research; for example, the research followed institutional, national, and international codes and regulations associated with educational sciences. Other researchers' work and achievements relevant to the research were taken into account and cited appropriately. The results were analytically interpreted and published accurately, honestly, and transparently. My possible biases and conflicts of interest, such as taking a dual role as a researcher and designer, were clearly stated. All partners in the research collaborations, all authors of research publications, and the research funding foundation were properly acknowledged.

I was also cautious about research misconduct and other unacceptable practices that might occur intentionally and unintentionally during the research, including fabrication, falsification, plagiarism and self-plagiarism, and misappropriation (see ALLEA, 2017; TENK, 2012). I endeavoured to prevent research misconduct and unacceptable practices, for example, by crediting the contributions of all parties involved accurately, citing others' and their own works appropriately, and avoiding redundant publication.

4 THREE PHASES OF THE RESEARCH

This EDR was conducted in three phases: initial research, concept development, and design development. During the multiple iterations, the design solution (i.e., a manipulative and its accompanying instructional materials and class activities) was developed, prototyped, and evaluated in classrooms.

4.1 Phase 1: Initial research

The goal of the initial research was to establish a context theory (RQ 1.1), which was later used to inform the design and development of the solution to the educational problem. During this phase, the needs, challenges, and opportunities of using manipulatives in primary school classrooms, particularly for promoting understanding of equation-solving concepts, were identified. Moreover, different manipulatives and educational games for solving equations were investigated.

4.1.1 Literature review

The literature review in Chapter 2 provides a foundation for contextual understanding. In Finland, third-to-sixth graders should learn the unknown concept and linear equation solving through reasoning and trial-and-error substitution of values for the unknown, whereas seventh-to-ninth graders should develop competence to form and solve equations algebraically. To become proficient in solving equations, students need to construct their conceptual understanding of equations, mathematical equivalence, different terms in an equation, and equation solving. Learning through multiple representations (e.g., manipulatives), mathematical models, reasoning, reflection, and social interaction, has been recommended for promoting students' understanding of key concepts in equation solving.

The balance model has commonly been used for students at different grade levels to learn linear equations. This model can intuitively and concretely illustrate the mathematical equivalence of two equal entities between the equal sign and support equation solving, particularly by doing the same operation on both sides of the equation. Nevertheless, it is difficult to represent negative integers and subtraction with the model. While some solutions to this shortcoming (e.g., using helium balloons to lift the scale) have been proposed, the model can be regarded as a foundation for students to construct their understanding of equation solving before moving to more abstract situations.

Manipulatives can be used to concretely represent abstract equation-solving concepts, thereby helping students construct an understanding of equation solving. To benefit from manipulatives, students should manipulate them, then think and reflect on their experience and discuss with others what they have discovered. Manipulatives can enhance learning through multiple representations, as well as discovery and social constructivist learning. There are three types of manipulatives (i.e., physical, virtual, and tangible), each of which has its own strengths and limitations, as outlined in Table 2.

Table 2. Summary of strengths and limitations of each type of manipulatives based on the literature review

Types of manipulatives	Strengths	Limitations
PMs	• Concretising abstract concepts • Encouraging physical action to promote learning • Offloading cognition • Improving memory through physical action • Assisting embodied cognition • Serving as reflection and communication tools • Making students' thinking visible to others	• Requiring considerable guidance and support to benefit from PMs • Typically, pricier than VMs
VMs	• Providing immediate guidance, feedback, and scaffolding • Providing precise representations • Linking pictorial and symbolic representations with students' interactions • Drawing students' attention to what is relevant • Motivating students • Promoting students' creativity • Increasing students' solution diversity • Recording and tracking students' actions for reflection and assessment • Ease of sharing, cleaning up, and storing and retrieving configurations • Accessibility, availability, and affordability	• Replacing rich physical interaction with mouse-keyboard clicking or touch screen tapping and scrolling • Potential rote learning • Distracting students' attention from learning • Losing mathematics learning time to learning how to operate VMs or solving technical issues • Requiring accessibility and availability of the necessary technology
TMs	**Physical representations:** • Creating a sense of physicality and embodiment • Enabling natural bodily interactions • Intuitive interaction enabling allocation of cognitive resources to mathematics learning • Offloading students' cognitive demands • Suitable for students with visual impairment	• More components and setup than PMs or VMs, decreasing practicality • Advanced technology means high prices compared to PMs and VMs
	Digital representations: • Providing immediate feedback and scaffolding • Recording and tracking students' actions for reflection and assessment • Changeable perceptual properties for representing certain mathematical concepts	
	Physical and digital representations together: • Providing a conceptual metaphor for to-be-learnt concepts • Connecting physical, pictorial, symbolic, and other representations • Attracting students' multiple senses • Enabling accessibility for different learners • Allowing parallel multi-user interactions • Encouraging facial, gestural, and verbal communication • Motivating learning	

Despite the benefits of meaningfully using manipulatives, their use in the classroom is still relatively limited. Although teachers consider manipulatives to be beneficial, they usually prefer traditional teacher-centred and paper-and-pencil instruction to manipulative use. Day-to-day challenges for the use of manipulatives in the classroom include manipulative availability (e.g., a lack of acquisition money or limited numbers of manipulatives available), manipulative organisation (e.g., difficulty in borrowing and returning, setting up and packing away, sorting and storing manipulatives, and damaging and losing them), classroom management (e.g., crowded classes, students not listening to instructions, manipulative misuse, noisiness, and messiness), and a lack of time (e.g., to use, organise, set up, and pack manipulatives away).

4.1.2 Existing solution investigation

Apart from the literature review of equation-solving learning and manipulatives in general, I reviewed previous studies on different manipulatives and educational games for solving equations, as well as trialled them by myself. The aim was to investigate their design in general, as well as their unique strengths and limitations. To date, all three types of manipulatives (i.e., PMs, VMs, and TMs) for solving equations are commercialised, self-made by teachers, or research-based prototypes.

Available PMs for solving equations include physical balance scales, algebra tiles, and cups and chips. Most commonly used PMs are physical balance scales, which are concrete forms of the balance model. A scale balances or tilts (as the metaphor of equal or not equal) in response to the placement and removal of distinct objects (one standing for constants and another standing for unknowns in an equation) on each side. The scale's dynamic actions, according to students' equation-solving actions (i.e., addition or subtraction) concretise the mathematical equivalence concept and provide students with immediate feedback. Unfortunately, the benefits of balance scales come with a price. Similar to Figueira-Sampaio et al. (2009), I found that physical balance scales often pose practical challenges regarding their affordability and mechanical requirements. For example, the balance scale mechanics are typically costly (a set of a balance scale with objects/weights costs about €15–30) and easily broken. Size, capacity, and accuracy of scales determine the sizes and weights of objects that can be used to represent constants and unknowns. Preparation before solving each equation (e.g., scale calibration for zero adjustments and setup of correct weight objects to represent the unknown in different equations) and technical attributes of scales (e.g., balance maintainability and sensitivity) typically cause inconvenience in the classroom. Another limitation is that not all values can be represented using physical balance scales. Physical scales can only represent equations containing natural numbers. Although special four-pan balance scales (see e.g., Learning Resources, 2020) can handle negative integers, they are more expensive (Learning Resources' scale costs about €40), more complicated, and more easily damaged than typical scales. Additionally, it is not easy to use commonly available objects to represent fractions on a balance scale. Representing equations whose solution equals zero ($x = 0$) is also technically challenging, because an object representing the unknown has its weight. Moreover, the balance mechanics (i.e., when placing an object on one side of a scale, the scale beam tilts down on that side) support only the representation of zero and positive values.

Apart from commercialised balance scales, teachers' self-made balance scales from available materials (e.g., paper, clothes hangers, cups, and blocks) have also been used for solving equations. Whereas self-made scales are inexpensive compared to commercialised scales, they take time to make and are usually less accurate. Moreover, rather than tilting dynamically, some self-made scales have to be operated by hand and thus do not provide students with real-time feedback.

Algebra tiles are manipulatives used to represent constants and variables in algebraic expressions. Algebra tiles include small squares (constant values of 1 and –1), rectangles (variable/unknown [in equations] values of x and $-x$), and large squares (variable/unknown [in equations] values of x^2 and $-x^2$). Commonly positive unit tiles, x-tiles, and x^2-tiles are yellow, green, and blue, respectively; all negative tiles are red. Algebra tiles can be used to solve equations on an equation mat, which is vertically divided in half (sometimes with an equal sign in the middle), representing both sides of the equation. Equations containing subtraction, such as $x - 2 = 4$, can be represented with one positive x-tile and two negative unit tiles on the left side of the equation mat and four positive unit tiles on the right. To solve this equation, two positive unit tiles can be placed on both sides of the equation mat, and then two *zero pairs* of unit tiles (i.e., one positive unit tile and one negative unit tile cancel each other out) on the left side are removed, thereby leaving one positive x-tile on the left and six positive unit tiles on the right, so $x = 6$. Negative integers can be represented with algebra tiles, but not easily with balance scales. However, when compared to balance scales, algebra tiles are more abstract, less intuitive for students, and do not illustrate mathematical equivalence well (Braukmüller et al., 2019), nor do they provide physical feedback regarding the correctness of students' actions (Reinschlüssel, Alexandrovsky et al., 2018).

Braukmüller et al. (2019) asked mathematics teachers at different school types and levels, who were also textbook authors, to rate the indispensability of balance scales and algebra tiles to their teaching based on each manipulative type's ability/inability to illustrate equivalence and negative integers. The research findings were in favour of balance scales. In the study of Reinschlüssel, Alexandrovsky et al. (2018), mathematics textbook authors, who were also experienced teachers, found the different colours of each tile type slightly confusing. Moreover, some teachers were unsatisfied with possibly drawn relationships between sizes of x-tiles and unit tiles (one x-tile was approximately three or four times the unit tile) because this could mislead students about the value of the unknown. In terms of practicality, algebra tiles are more affordable (a set of algebra tiles costs about €6–12) and more durable (the tiles have neither moving nor mechanical parts) compared to balance scales.

Cups and chips have also been used to solve equations whose unknowns are represented with cups (facing up for positive and facing down for negative), and constants are represented with two-sided chips (one side for positive and another side for negative). Although cups and chips share a similar idea to algebra tiles, they appear to be less abstract and more intuitive for students than algebra tiles. For example, regarding the concept of the unknown, it is more tangible to picture that each cup contains the same number of chips than picturing that each rectangular x-tile equals the particular number of square unit tiles. Additionally, I found that cups and chips were relatively easier to make and less expensive to acquire than algebra tiles.

To summarise, physical balance scales, algebra tiles, and cups and chips concretise the key concepts of equations and equation solving. The balance principle of scales illustrates mathematical equivalence. In the case of algebra tiles, Magruder's (2012) research findings indicate that the representation of constants as squares and unknowns as rectangles enabled students to distinguish between the two terms and eventually realise their differences. Nevertheless, according to my observations, the use of distinct objects on balance scales, as well as cups and chips to represent constants and unknowns of an equation, can differentiate between the two more explicitly than the algebra tiles. Negative integers can be represented with algebra tiles as well as cups and tiles, but not with typical balance scales. Compared to other PMs, balance scales tend to pose more practical challenges due to their mechanics.

Available VMs are typically digital versions of physical balance scales and algebra tiles. Algebra Balance Scales (NLVM, n.d.) is a free web-based Java applet for sixth-to-eighth graders to learn to solve simple linear equations. The applet consists of a balance scale, unit blocks representing one, and x-boxes representing positive unknowns. Algebra Balance Scales - Negatives applet (NLVM, n.d.), which is recommended for the same school grade level, solves the limitation of physical balance scales in their inability to represent negative values using balloons to lift the beam of balance scales. Unit balloons represent negative ones; *x*-balloons represent the negative unknown. Objects (i.e., blocks, boxes, and balloons) are placed on the scale to digitally model equations randomly generated by the applet or created by students. To solve an equation, students select mathematical operations (i.e., addition, subtraction, multiplication, or division of constants and unknowns) to be performed on both sides of the equation. Each equation-solving step made by students is simultaneously represented visually on the scale (the scale beam tilts according to students' manipulations) and symbolically in an equation window. Dynamic visuals of tilting balance scales assist students in developing their understanding of the equal sign and mathematical equivalence (Magruder & Mohr-Schroeder, 2013). Moreover, NLVM balance scales explicitly link pictorial and symbolic representations (Suh & Moyer, 2007; Magruder & Mohr-Schroeder, 2013) and thus are likely to enhance students' representational fluency. Both scales also provide real-time feedback, step-by-step guidance, and self-checking, thereby promoting students' independent learning.

In line with Magruder (2012), I found various practical advantages of both scales. For approximately €20, all NLVM applets, including the balance scales, can be installed on individual computers and servers for offline use without Java; additional features (e.g., customisation and recording completed work for students' reflection and assessment) are also available. In Magruder's (2012) study, practical requirements for working with the applets appeared to cause challenges regarding accessibility of the computer lab and losing learning time to computer logging in/off.

The Algebra Tiles iPad app (Version 4.1.0; Brainingcamp, LLC, 2020) is algebra tiles in the digital environment. This €2 app is designed for children (no target age group mentioned) to explore various algebraic topics, including solving equations. Shapes and colours of virtual tiles are the same as their physical counterparts. There is also an option to simplify the colours of the tiles to only blue (for all positive tiles) and red (for all negative tiles), in case a variety of tile colours causes confusion. Tile labels (1, –1, x, $-x$, x^2, and $-x^2$), which can be toggled on/off, make each type of virtual tile easier to recognise compared to physical ones. Equations are solved on an equation mat (with an equation sign in the middle) with virtual algebra tiles, similar to with physical tiles. A workspace under the equation mat displays a mathematical sentence according to the tiles on the mat at that moment. This feature supports the connection between pictorial and symbolic representations of the equation.

However, similar to physical algebra tiles, the app does not provide any feedback regarding the correctness of students' equation solving. Students need to keep track of their equation-solving processes by remembering or changing an equal sign to another sign that correctly depicts their recent action. Although the app does not provide any guidance on how to solve equations with virtual tiles, it offers scaffolding for zero pairs: when opposite pairs are dropped on each other, they automatically cancel each other.

DragonBox Algebra 5+ (Version 1.3.7; Kahoot DragonBox, 2019a) is a €5 award-winning mobile educational game for children aged 6–8 to get familiar with algebra and basic processes of solving equations. Players learn to solve equations involving addition, subtraction, division, and multiplication through a game environment, discovery, and experimentation. Each level of the game starts with an animation showing how to win that level, for example, by isolating a dragon box, which later changes into the unknown x, on one side of the game board (an equation mat). The game uses familiar concepts to introduce the mathematical concepts required for solving equations. For example, picture cards, day-and-night cards, and a black hole are gradually replaced with numbers and variables, zero pairs, and the additive identity of zero, respectively. The game provides step-by-step guidance at the beginning of each level and instant feedback throughout the game, so it is possible for children to play it without adults' supervision.

DragonBox Algebra 12+ (Version 2.3.1; Kahoot DragonBox, 2019b) is a sequel for DragonBox Algebra 5+. The €9 educational game is designed for 12–17-year-olds to learn advanced topics in mathematics and algebra, such as collection of like terms, factorisation, and substitution. Although both games offer a playful learning environment for algebra and equation solving, research indicates their possible limitations. Some teachers felt that the design of the DragonBox [Algebra 12+] game was possibly too childish for ninth graders (Reinschlüssel, Alexandrovsky et al., 2018). Many fifth graders did not perceive the mathematical attributes involved in playing DragonBox Algebra 12+ (Tucker & Johnson, 2017).

In short, existing VMs for solving equations have advantages and disadvantages similar to VMs in general, as addressed in the literature. They provide various benefits to the digital environment (e.g., step-by-step guidance, real-time feedback, connection between pictorial and symbolic representations, and affordability), which their physical counterparts cannot provide. Nevertheless, they lack the benefits that come with real physical interaction (Reinschlüssel, Alexandrovsky et al., 2018). While available educational games appear to provide fun, discovery, and independent learning, their suitability to the mathematics classroom is somewhat uncertain.

Currently available manipulatives for solving equations are mainly physical and virtual. Recently, the Multimodal Algebra Learning (MAL) tangible system (Reinschlüssel, Alexandrovsky et al., 2018; Reinschlüssel, Döring et al., 2018) was developed. The system consists of smart algebra tiles (width × depth × height = 7 × 7 × 5 cm) and a 2 × 2 interactive tabletop. The two areas on the left of the tabletop stand for the left side of the equation, while the two areas on the right stand for the right side of the equation. The upper areas of the tabletop are the 'addition zones', where all tiles are linked by addition. The lower areas are the 'subtraction zones', where all tiles are deducted from the upper areas. Lights illuminating each tile indicate tile types through shapes (squares for unit tiles and rectangular for x-tiles) and colours (blue for positive tiles and red for negative tiles). Four tile colours typically used for physical algebra tiles were reduced to two colours for clarity and intuitiveness.

The MAL tangible system enhances learning in various ways that physical or virtual algebra tiles alone do not. The system provides multimodal input (i.e., direct haptic interaction with the tiles) and output, for example, the current +1 or −1 value of a tile on its LED display, a tile's light [sound and vibration under testing] feedback/hints about equation-solving operations, and a real-time symbolic representation of an equation-solving step in response to students' interaction with smart tiles. The smart tiles also provide dynamic constraints that are magnetic hints for directing the grouping of tiles (which is an important action for transforming and solving equations) and preventing incorrect combinations of tiles. When two tiles are placed next to each other, based on the current value of each tile and their possible combinations, the magnets inside the tiles attach them together if they are fit for each other (e.g., +1 and −1 can be paired as a zero pair) or repel them if they are unfit for each other (e.g., a unit tile and an x-tile). The system can, to some extent, automatically adapt feedback and hints to suit students' levels and needs.

MAL interactive prototypes were planned to be tested with students to evaluate the system's benefits. The MAL system utilises technological solutions purposely designed for assisting students with different levels and needs in learning to solve equations through multimodal interaction, dynamic constraints, feedback, and adaptivity. Nevertheless, the use of advanced technologies comes with a cost. Based on the published information, the system's current design appears to be impractical for the classroom due to the smart tile size and possible weight, as well as the overall cost of the system.

4.1.3 Initial fieldwork

To better understand the educational problems, real-world educational settings, and classroom utilisation of manipulatives, initial fieldwork was conducted in third-to-sixth-grade classrooms of a middle-size school in spring 2015. Figure 3 illustrates the research design of the initial fieldwork (for more details, see Publication II).

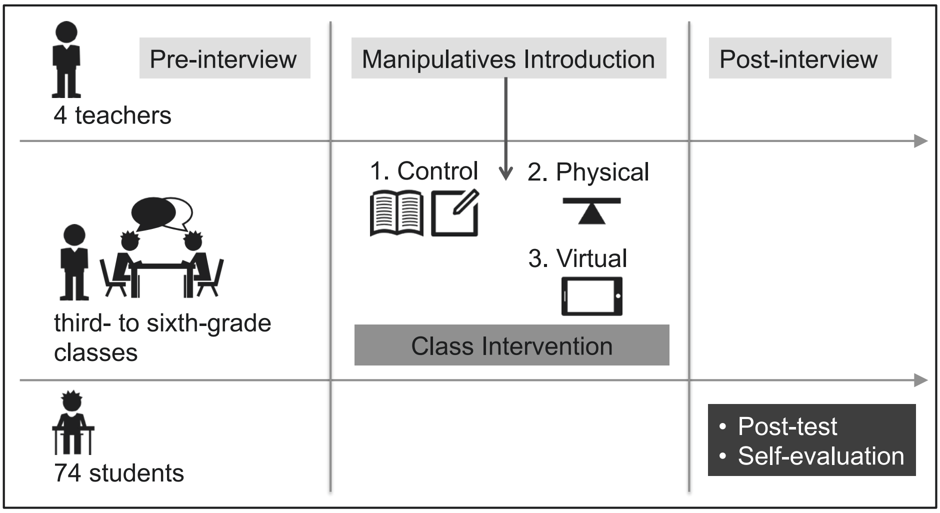


Figure 3. The research design of the initial fieldwork. From 'Using manipulatives for teaching equation concepts in languaging-based classrooms,' by D. Lehtonen and J. Joutsenlahti, in N. Pyyry, L. Tainio, K. Juuti, R. Vasquez & M. Paananen (Eds.), *Changing subjects, changing pedagogies: Diversities in school and education*, p.168, 2017, Finnish Research Association for Subject Didactics. Copyright 2017 by the Finnish Research Association for Subject Didactics. Reprinted with permission.

Four primary school teachers and 74 students (with no/low prior knowledge of equation solving) participated in the study. The teachers were interviewed about their experiences and opinions regarding teaching equation solving and using manipulatives in their classrooms from pedagogical and practical perspectives. The interview questions were informed by the research objectives and the literature. One-lesson class interventions were conducted after the teacher interviews. The teachers divided their students with different mathematics attainments equally into three groups: learning with paper-and-pencil (n = 25), a physical balance scale (n = 25), and a virtual balance scale (n = 24). Balance scales were chosen for the interventions because they are intuitive and illustrate mathematical equivalence, and therefore suitable for helping novice students (the target of this study) learn linear equation solving. Hands-On Equations® balance scale (Borenson & Associates, n.d.) and its mobile app, Hands-On Equations 1 (Version 3.7; Borenson & Associates, 2015), were used for the interventions to investigate the strengths and limitations of physical and virtual forms of the same manipulative.

The same teacher taught their own students in all instructional groups how to solve equations (by substituting values of the unknown in the third- and fourth-grade interventions and by performing the same operation on both sides of the equation in the fifth- and sixth-grade interventions). To ensure the development of students' equation-solving understanding, class activities were designed to promote learning through multiple representations, discovery, social interaction, reasoning, and reflection. The students were asked to work in pairs to translate and solve equations through talking, drawing, and writing mathematical symbols, and in the PM and VM groups, also through using manipulatives.

After the interventions, the students completed the paper-based test (which was informed by existing instruments, the literature, textbooks, and an input from mathematics education experts) with no access to the manipulatives. Informed by the literature, the test contained a variety of tasks. It was used to evaluate how well each instructional condition enhanced the students' equation-solving performance and their representational fluency (i.e., ability to make links between multiple external representations), which indicates their conceptual understanding of equation solving. Then, all students evaluated their learning development after the intervention; the students in the PM and VM groups also evaluated their learning experiences with the manipulatives. The teachers were again interviewed about their experiences and opinions regarding each instructional condition.

Initial research findings were presented at the Fifth Nordic Conference on Subject Education (Lehtonen & Joutsenlahti, 2015). The findings regarding pedagogical aspects are thoroughly reported in Publication II. The overall results indicate that the class activities (encouraging learning through multiple representations, discovery, social interaction, reasoning, and reflection) during the interventions are likely to help students across the three instructional conditions learn to translate equations into different representations and solve them. Among all instructional conditions, PM-based instruction appeared to benefit students' equation-solving learning and performance most.

Students in the PM group outperformed their classmates in the other conditions on the test, and most (21/25) of the PM students believed that the PM helped them learn equation solving. Likewise, all teachers unanimously regarded their lessons with the PM as the most successful instructional condition. In their opinion, concrete interaction with the PM supported students' peer interaction, multimodal expression of mathematical thinking (i.e., languaging), and understanding of mathematical equivalence, different terms in an equation, and equation solving. Some teachers noticed that their low-attaining students in the VM group tended to use the VM to solve equations by scrolling and trying different values for the unknown until they arrived at the correct solutions. Therefore, they felt that these students might not really understand the to-be-learnt content and therefore would perform worse on the test than other students.

All in all, the teachers believed that among all instructional groups, the PM group had the best conceptual understanding of equation solving and would perform best on the test. Moreover, teachers also found that the PM was straightforward and helped their students complete the exercises faster than the students in the other groups. Similar to the teacher interviews, the class intervention observations provide evidence of PM benefits and VM hindrances to students' peer interaction, languaging, and equation-solving learning. During their pair work, PM students usually used the PM to translate and solve equations together. While manipulating the PM, students tended to discuss with each other and say aloud what they were doing or thinking (Figure 4a). In contrast, VM students were less likely to discuss and express their actions or thinking in words. Instead of using the VM together to complete the exercise, students usually worked separately and sometimes even held an iPad for themselves (Figure 4b). Moreover, many students appeared to manipulate the VM in a rote, procedural manner to complete exercises. Regarding ease of use, PM students learnt how to use the PM without difficulty, whereas many VM students had difficulty in using the VM to solve equations and check the solutions.

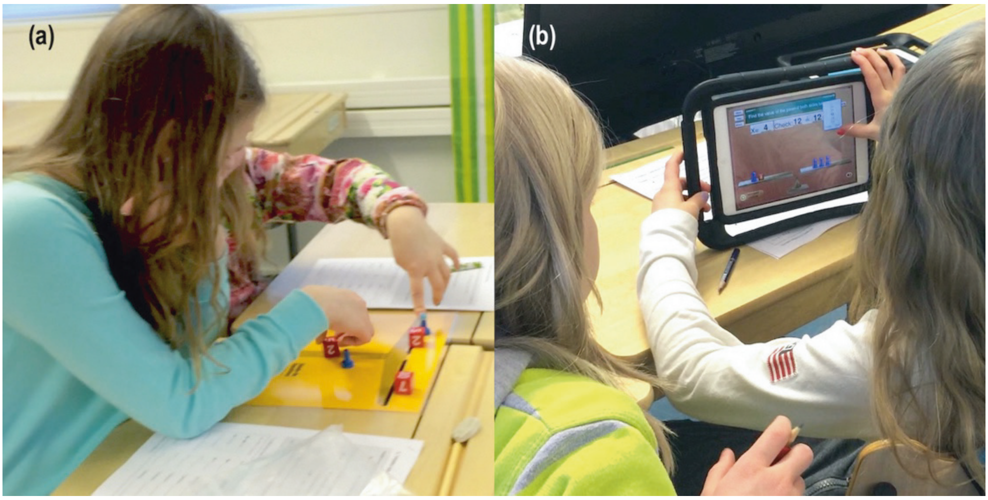


Figure 4. Third graders working in pairs during the class interventions: (a) PM students discussing and helping each other to solve an equation with the PM; (b) a VM student holding an iPad for herself and using it to solve an equation, while her partner silently looks at the iPad

Although the overall findings indicate a number of disadvantages of the VM, potential benefits to students' learning and performance were also found. The VM students performed somewhat better than their classmates in the paper-and-pencil condition on the test. Almost two-thirds (17/24) of them considered that the VM helped them learn equation solving. During the interventions, students in the PM and VM groups worked more independently, with less support from the teachers compared to the paper-and-pencil group. The VM also provided students with a link between pictorial and symbolic representations of equations as well as immediate feedback regarding the correction of equation solutions and solution checking. Additionally, some teachers thought that high-attaining students could independently use the VM to learn equation solving at their own pace.

Unlike most balance scale manipulatives, the scale beam of the PM and VM is fixed (i.e., always in balance regardless of objects on both sides of the scale). Therefore, both manipulatives provide no information regarding the balance/imbalance of the scale. This potentially influenced the research findings. For example, the inability to illustrate equivalence could have some effect on PM and VM students' learning and performance. The PM was found to be easy to use, which could be because the scale does not require any balance calibration and setup. Additionally, no balance scale mechanics makes the PM more affordable (one student set costs about €4) than typical physical balance scales.

The teacher interviews reveal several factors regarding classroom and school practice that potentially prevent the acquisition and adoption of manipulatives. All teachers considered time limitations to be a barrier to the use of manipulatives in the classroom. Acquisition budget was also a concern for most teachers, due to their school's budget constraints. Other factors, including available class and storage space, classroom management, and teacher skills and knowledge, were also mentioned.

4.2 Phase 2: Concept development

Phase 2 started with design opportunity and tentative design principle identification to initially address RQ 2. Based on this identification, four potential design concepts were generated. After that, the generated concepts were evaluated with teachers to preliminarily address RQ 1.2 and select promising concept(s) for further development.

4.2.1 Four manipulative concepts

The contextual knowledge derived from Phase 1 indicates an opportunity for design solutions that embrace the strengths of existing manipulatives, and at the same time, address their limitations regarding successful classroom utilisation and adoption. The tentative design principles informed by the Phase 1 results provided guidance and direction on how to design a solution that helps primary school students understand equation-solving concepts and encourages teachers to adopt it in their classrooms. Publication IV presents four manipulative concepts and describes how their design was guided by tentative design principles.

Regarding pedagogy, a manipulative should assist students in learning through their firsthand experience and provide them with appropriate guidance and scaffolding. It should also enable students to connect multiple representations of equation-solving concepts and express their mathematical thinking through different modes of meaning making. Moreover, a manipulative should encourage students to construct their knowledge together with their peers. In terms of content to be covered, a manipulative should use the balance model to concretise concepts of equation, mathematical equivalence, different terms in an equation, and equation solving. The strengths and limitations of existing manipulatives were also used to guide the manipulative design. During the manipulative concept exploration, most attention was given to the pedagogy and to-be-covered content; later, when evaluating design concepts and further developing the design solution, practicality was also taken into account. After exploring different design alternatives, four manipulative concepts (Figure 5) were generated based on tentative design principles.

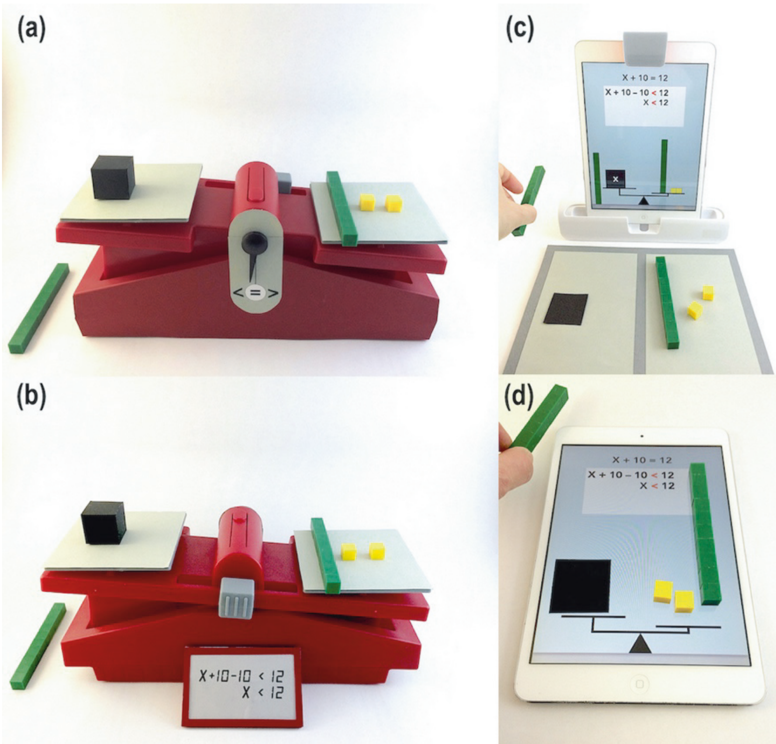


Figure 5. How to solve an equation (x + 10 = 12) using each of the four generated design concepts (a) Concept A; (b) Concept B; (c) Concept C; and (d) Concept D. From 'Constructing a Design Framework and Design Methodology from Educational Design Research on Real-World Educational Technology Development,' by D. Lehtonen, 2021, *Educational Design Research, 5*(2), Article 38, p. 10 (https://doi.org/10.15460/eder.5.2.1680). CC BY 4.0.

All the concepts share similar core ideas and only differ in their utilisation of existing technologies to meet the design objectives. Concept A (Figure 5a) is composed of a tiltable physical balance scale and physical objects: black boxes representing unknowns and base-10 blocks representing constants. Concept B (Figure 5b) is composed of a tiltable physical balance scale, a digital display showing mathematical sentences for the current stage of equation solving, and physical objects. Concept C

(Figure 5c) is composed of a tablet app, a mirror placed in front of a tablet camera for physical object detection, physical objects, and a mat representing both sides of the scale. The tablet screen displays mathematical sentences for the current stage of equation solving and images of the objects on and off the scale. Concept D (Figure 5d) is composed of a tablet app and physical objects. The tablet touchscreen detects physical objects on it and provides outputs (i.e., images and mathematical symbols) accordingly.

4.2.2 Concept evaluation

To determine how well each envisioned concept could promote students' understanding of equation-solving concepts and encourage teachers to adopt it in the classroom (RQ 1.2), the concept evaluation was conducted through teacher questionnaires and interviews. Initial results of the concept evaluation were presented at the Sixth Nordic Conference on Subject Education (Lehtonen & Joutsenlahti, 2017a). The evaluation methods and results are reported in Publication IV.

Nonfunctional mock-ups describing key functional features and the initial visual appearance of each concept were introduced to 12 primary school teachers (four participated in the initial fieldwork). The teachers then rated each concept regarding how well they potentially provide pedagogical benefits and were compatible with classroom and school practice, on a scale of 1 (*not at all*) to 4 (*very well*). The evaluation criteria were informed by the literature, Marshall and Swan's (2008) manipulative survey, Collins et al.'s (2004) independent variables affecting the success of the design in practice, the Finnish NCC 2014 (EDUFI, 2016), and the initial fieldwork results. After rating the concepts, the teachers were required to explain the reasons behind their rating responses.

All four concepts were rated relatively highly (M = 3–3.5) regarding their potential benefits for students' understanding of equation-solving concepts, discovery learning, social interaction, and multimodal expression of mathematical thinking. Nevertheless, only Concept D was highly rated (M = 3.5, SD = 0.78) for its compatibility with classroom and school practices (e.g., acquisition budget, preparation, and class management), whereas the others were rated below 3. At the end, the teachers were required to make an acquisition decision by taking into account all pedagogical and practical factors. The majority of the teachers (9/12) would definitely acquire Concept D for their class, and none said they would not acquire it for their class in any case. According to their explanations, there were no significant differences between the concepts regarding pedagogical advantages; thus, compatibility with classroom and school practice was the deciding factor in their acquisition decisions. Concept D was teachers' favourite because it potentially provides high pedagogical benefits and seems to be straightforward, usable, compact, portable, durable, compatible with existing school tablets, attractive to diverse learners (regarding age and attainment levels), and multifunctional. Based on the teachers' responses, Concept D was selected for further development.

The evaluation results bring to attention that practicality is likely to play an important role in manipulative implementation and adoption in the classroom. The results from the initial fieldwork and the concept evaluation regarding this notion were presented at the 17th Biennial European Association for Research on Learning and Instruction Conference (Lehtonen & Joutsenlahti, 2017b).

4.3 Phase 3: Design development

Phase 3 aimed to refine design principles to address RQ 2. Moreover, it sought to develop the selected concept based on the refined design principles and then evaluate the developed solution in classrooms to address RQ 1.2. The evaluation results were also used to guide the refinement of the solution.

4.3.1 Design solution

At the beginning of Phase 3, another literature review (e.g., technology-enhanced learning and tangible technologies) and investigation of educational products (e.g., textbooks and educational technologies) that were relevant to the design development were conducted. The knowledge underlined by the literature and the investigation of existing solutions and educational products were incorporated with the Phase 2 findings (i.e., initial outcomes knowledge) to refine the design principles. Then, the selected design concept (Concept D) was developed based on the refined design principles: concretising key equation-solving concepts; supporting multimodality and languaging, discovery learning, and social interaction; in agreement with curriculum; suitable for diverse learners; easy to use; and feasible for classroom and school practice. The design principles were presented thoroughly in Publication III. According to Edelson (2006) and Collins et al. (2004), lessons can also be learnt from unsuccessful design. Teachers' feedback about rejected concepts during Phase 2 was also taken into account when developing the selected concept. A TM, instructional materials, and class activities were developed as a solution to the educational problem.

As I have a strong background in design, I took a designer's role in developing the solution and building most parts of its working prototypes myself. Mathematical contents of the design solution were developed under the supervision of my supervisors, who are experts in this subject matter and mathematics education. The TM, particularly its technological parts, was developed and prototyped in collaboration with a team of computer science students and their supervisor. During the TM prototyping, some trade-off decisions were made to balance its pedagogy, practicality, and technological feasibility. The development and prototyping processes of the design solution were described in Publications III–IV.

X-is ('*X* is equal to') is a TM designed based on the selected concept to concretise the key concepts of equation solving. Aiming at deployment in classrooms today, the TM employs off-the-shelf technology to ensure its feasibility and affordability. During the development, it became clear that the original object-tracking idea used in the concept was not technologically feasible, so the current design tracks the objects on the table screen with image recognition via an external USB web camera (Figure 6a vs. Figure 5d). *X*-is consists of a tablet app and two kinds of physical objects: *X*-Boxes specially designed to represent unknowns and existing base-10 blocks used to represent constants. The app has two levels for learning to solve equations by substituting values for the unknown at Level 1 and by doing the same operation on both sides of the equation at Level 2. When students model and solve an equation by manipulating a *X*-Box(es) and a base-10 block(s) on a table screen, the app provides pictorial (i.e., tilting digital balance scale) and symbolic (i.e., mathematical sentences) feedback according to their actions (Figure 6b). The app also gives students multimodal guidance and tips. Implemented architecture and description of the TM were presented in detail in Publication III.

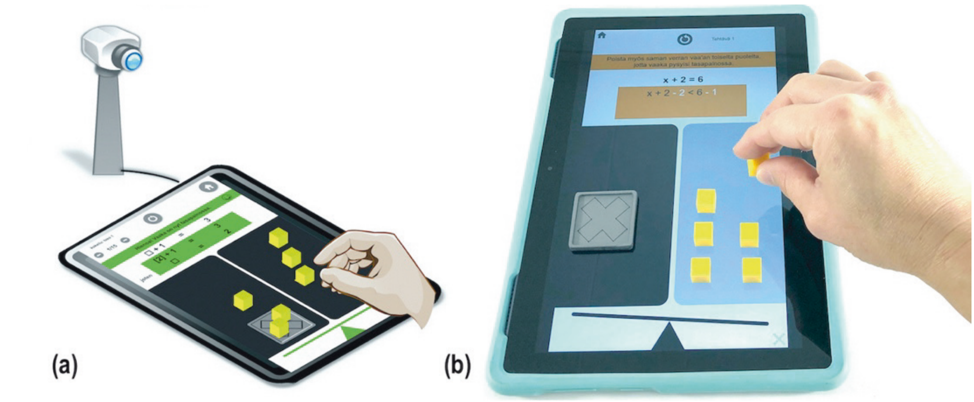


Figure 6. (a) Image recognition via an external USB web camera used for tracking objects in *X*-is; (b) multimodal inputs and outputs of *X*-is. From 'The Potentials of Tangible Technologies for Learning Linear Equations,' by D. Lehtonen et al., 2020, *Multimodal Technologies and Interaction, 4*(4), Article 77, pp. 8, 11 (https://doi.org/10.3390/mti4040077). CC BY 4.0.

The instructional materials consisted of two sets of student worksheets and teacher guides designed to be used together with *X*-is. The first set, together with *X*-is Level 1, is for lower-grade students to learn to solve equations by substituting values for the unknown. The second set, together with *X*-is Level 2, is for upper-grade students to learn to solve equations by doing the same operation on both sides. The worksheets encourage students to learn equation solving through multiple representations, while the teacher guides suggest how to implement equation-solving lessons meaningfully. The instructional materials were designed to have a structure similar to the textbooks and teacher guides typically used in Finland for user-friendly navigation. Publication IV provides more information regarding the instructional materials.

To ensure meaningful lessons, class activities were designed to support students' learning through multimodality and languaging, discovery, and social interaction. During the lessons, teachers should supervise students to use *X*-is in pairs/small groups to model and solve equations provided on the worksheets before writing down their equation-solving processes and solution(s) on their own worksheets.

4.3.2 Design evaluation

Design evaluation was conducted to determine whether and to what extent the developed design solution helped students understand equation-solving concepts and encouraged teachers to adopt it in their classrooms (RQ 1.2). The evaluation results helped to develop the outcomes theory and inform the design refinement. Initial results of the design evaluation were presented at the Annual Symposium of the Finnish Mathematics and Science Education Research Association (Lehtonen et al., 2019). The evaluation methods and results regarding the pedagogical benefits of the design solution were thoroughly reported in Publication III; results regarding its compatibility with classroom and school practice were presented in Publication IV. Figure 7 shows the research design of the design evaluation.

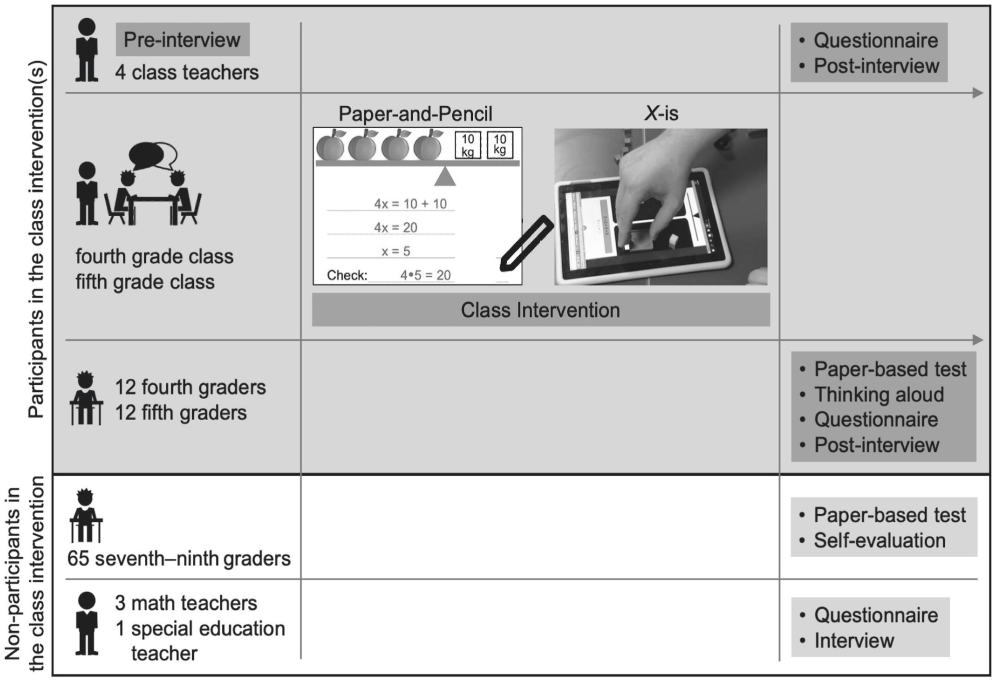


Figure 7. The research design of the design evaluation. Icons were designed by Freepik.

Apart from students' learning, there are other dependent variables to be assessed to determine the success or failure of a design solution (Collins et al., 2004; Plomp, 2013). In this EDR, different aspects relevant to the success of the developed solution, including its pedagogical benefits, usability, and compatibility with classroom and school practice, were assessed. A variety of qualitative and quantitative methods—observations, paper-based tests, questionnaires and interviews, and thinking aloud—were employed to address the evaluation aspects.

One-lesson class interventions were implemented with one fourth- and one fifth-grade teacher, and their students (12 fourth graders and 12 fifth graders), who had no/low prior knowledge of equations solving. Each teacher divided their students with different mathematics attainments equally into two instructional conditions: learning with or without *X*-is. Aside from that, both conditions used the same teacher guides, worksheets, and class activities, which are parts of the design solution. Following the class intervention, the students took a paper-based test. The class intervention and the test were conducted similarly to those of Phase 1. To examine how well the design solution enhanced students' equation-solving performance and representational fluency compared to traditional instruction, a comparison group of mixed-attaining seventh-to-ninth graders (n = 65), who had several equation-solving lessons in their normal school curricula, took the same test without participating in the intervention. After the posttest, all students in the *X*-is condition individually participated in a thinking aloud session in which they solved equations with *X*-is, and at the same time, talked about their actions. It should be noted that the thinking aloud session took place after the posttest to prevent potential impacts of the session on *X*-is students' posttest performance. The session aimed to assess *X*-is usability and students' mathematical understanding by taking into account the content and process that students use to arrive at their solution, as proposed in the literature. The *X*-is students rated and justified the extent to which they regarded *X*-is as easy to use, pleasant to use, and helpful for their equation-solving learning, and would consider using it again. *X*-is was also evaluated by both teachers using a process similar to the Phase 2 concept evaluation. Additionally, a special education teacher and three lower secondary school mathematics teachers also evaluated *X*-is to examine its possible utilisation and adoption in other educational contexts.

Overall, highly favourable results from the design evaluation suggest successful classroom utilisation and adoption of the design solution. The class intervention observations, as well as teacher questionnaires and interviews, reveal that learning with *X*-is better promoted students' languaging, peer interaction, and discovery learning compared to learning without it. *X*-is enabled students to express their mathematical thinking physically and verbally. It also encouraged students to interact with each other and independently experiment and discover to-be-learnt content together. The comparison between the average test scores of the students in the interventions (learning through the proposed class activities with the designed worksheets with/without *X*-is) and the comparison group of students indicates a positive impact of the design solution on students' equation-solving achievement and representational fluency. The test results and thinking aloud session observations indicate the benefits of *X*-is to students' equation-solving performance and understanding.

Although the test performance of the students in both instructional conditions was relatively similar, *X*-is students were more likely to use the strategies that were taught during the class interventions to solve equations correctly than paper-and-pencil students. Moreover, during the thinking aloud sessions, students used *X*-is to model and solve equations correctly and were able to justify their equation-solving processes (see Videos S1–2 in Publication III). The design evaluation results demonstrate the satisfactory usability of *X*-is. Based on students' questionnaire and interview responses and observations from the class interventions and the thinking aloud sessions, *X*-is was easy and pleasant to use. Most students found it beneficial to their equation-solving learning and said they would like to use it again. Teachers' questionnaire and interview responses also indicated the possibility of the adoption of *X*-is in the classroom and its use in other educational settings. Teachers considered *X*-is to be highly compatible with classroom and school practice; all of them indicated they intended to acquire *X*-is for their classrooms. They also believed that *X*-is was likely to enhance pre-primary-to-ninth-grade students' understanding of equation concepts.

Despite highly favourable evaluation results for *X*-is, some refinement is required to increase its technical stability, pedagogical value, usability, and practicality. For example, the formative evaluation during the working prototype development revealed the unreliability of object tracking using a webcam. To overcome this image recognition challenge, different alternatives for *X*-Box redesign were explored (Figure 8).

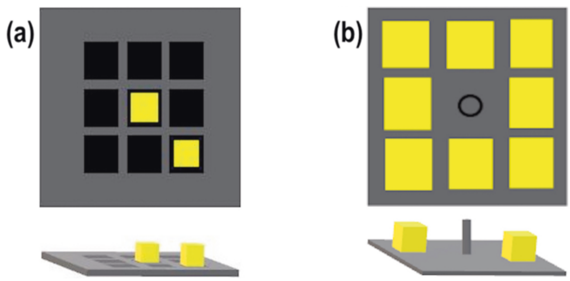


Figure 8. Alternatives for *X*-Box redesign to overcome image recognition challenge: (a) engraved squares for placing base-10 units not too close to each other; (b) added pole in the middle of the *X*-Box to increase *X*-Box's visibility. Illustrated by J. Korkala.

Results from the design evaluation with teachers and students also informed minor refinement of *X*-is. For instance, new features, such as free experiments with the balance scale and students' own equation setup, can be added to future design.

5 EXPERIENCE FROM EDUCATIONAL DESIGN RESEARCH

In the previous chapter, I provided research findings (regarding the needs, challenges, and opportunities of using manipulatives in primary school classrooms and outcomes of implementing the design solution in real classrooms), which can be used to construct domain theories (RQ 1). In this chapter, I reflect on the knowledge that I needed to develop the design solution and my experience of conducting this EDR, which can be used to establish design frameworks (RQ 2) and design methodologies (RQ 3), respectively.

5.1 Developing a real-world tangible manipulative

To develop a design solution that helps primary school students understand equation-solving concepts and encourages teachers to adopt it in their classrooms, I needed to iteratively acquire and synthesise various types of knowledge throughout this EDR. During Phase 1, I established my contextual understanding of the educational problem and the target real-world setting through a literature review (see Chapter 2), investigation of existing manipulatives and educational games (see Chapter 4.1.2), and fieldwork (see Chapter 4.1.3).

The literature provided the theoretical background of learning mathematics with understanding (e.g., what conceptual understanding is and how to develop it), manipulatives (e.g., types of manipulatives, whether and how they support mathematics learning, and challenges to their classroom utilisation and adoption), linear equation solving (e.g., key concepts in equation solving, equation-solving approaches, and how to learn equation solving), and instructional recommendations and equation-solving content recommended in the Finnish NCC (EDUFI, 2016).

The existing solution investigation reveals the strengths and limitations (i.e., what does or does not work) of available manipulatives and educational games, thereby providing insight into potential solutions to yield the desired results. The fieldwork was conducted in real classrooms to empirically investigate the educational problems, target context, target users, and their needs, as well as challenges and opportunities regarding manipulative use and adoption in the classroom. At the end of Phase 1, it became clear that the design solution needed to address the following aspects: equation-solving content, support for understanding of equation-solving concepts, and day-to-day practice related to classroom utilisation and adoption of manipulatives.

During Phase 2, the knowledge gained from Phase 1 was used to guide the design concepts. Teachers' responses during the concept evaluation not only supported Phase 1 knowledge, but also provided insightful information that could be used directly to inform the design. For example, their explanations indicated that the number of manipulative components had an influence on its setup, portability, and organisation, thereby affecting its practicality.

At the beginning of Phase 3, I conducted another literature review and educational product investigation to acquire new knowledge regarding possible technologies for developing the selected design concept, which employs tangible technologies. An understanding of digital, tangible, and object-tracking technologies were added to the knowledge gained from the two previous phases. The design evaluation results indicate that a design solution informed by all the knowledge is likely to be beneficial to students' equation-solving learning, practical, usable, and feasible for production with today's technologies.

5.2 Conducting the educational design research

My research journals, field notes, and communication records with the computer science student team and teachers assisted me in reflecting on my experience in conducting this EDR. The following themes regarding the benefits of EDR for my study and challenges that I encountered were recognised: iterations, data triangulation, various participants, multidisciplinary collaboration, technological innovations, alternative designs, and solitary researcher (for more details, see Publication IV).

The iterative investigation, development, assessment, and refinement helped me construct knowledge of the educational problems, the target educational contexts, and how the design solution could promote equation-solving learning and day-to-day practice. The design solution's feasibility and successful implementation and adoption in the classroom were also ensured through this iterative process. However, the iterations were intensive and required considerable resources. Partly for this reason, the study lasted over 6 years compared to 3- to 14-year completion time of the Finnish EDR doctoral dissertations on mathematics, science, and technology education, which were reviewed in Publication I.

The data triangulation not only assisted me in understanding complex and dynamic real-world educational phenomena, but also enhanced the research reliability and validity. Nevertheless, the collection and analysis of a large and varied dataset was resource intensive. Due to limited resources, a large amount of collected data that was not directly related to the RQs was left unused.

The involvement of teachers and students (i.e., target users) guided the solution development towards the desired outcomes. Target users' involvement also contributed to the advancement of theoretical knowledge; part of the research findings would not have been discovered without their involvement. Teachers and students benefited from their participation in the research; for example, the teachers stated that their participation allowed them to experience the influence of different instructional approaches on students' learning and to understand how to best assist their students. However, involving teachers and students in the research also posed challenges, such as participant recruitment and research permission acquisition.

Multidisciplinary collaboration helped to ensure solution viability regarding subject matter, pedagogy, and technology. Good communication, mutual respect, and shared understanding contributed to the collaboration success. Nevertheless, there were some challenges in collaborating with different disciplines, such as difficulties in establishing the collaboration and securing collaborators' commitment to the study.

While tangible technologies played a part in the satisfactory results of the design evaluation, the development of such educational technology caused several challenges. Originally, a fully working prototype, stable and containing all the features, would be built for the Phase 3 class interventions. However, due to limited time and the demanding technology implementation, only a Wizard of Oz prototype (see Beaudouin-Lafon & Mackay, 2012) was built, and the originally planned research design had to be modified.

Instead of working with only one design idea, I first explored different design concepts before evaluating them and selecting the most promising one for further development. This concept development phase reduced the possibility of discovering later that the developed solution might not be the best, thereby efficiently utilising available resources. Moreover, the evaluation of various concepts appeared to be a good way to collect in-depth information from teachers.

As the only researcher in this EDR, I had multiple roles and was involved in all processes. While this provided me with a profound understanding of the whole process, it was challenging at the same time. The complex nature of EDR and the scope of the research required substantial time, patience, and multidisciplinary expertise. With a background in design and educational sciences, I was able to conduct this EDR mostly alone, although it took many years to complete the study, and the study's quality was affected by limited human resources. For example, there were only 12 students working with *X*-is during the Phase 3 interventions, because I could operate only one Wizard of Oz prototype at a time. Because one cycle of the iterations was devoted to concept development, the developed solution was implemented in real classrooms only during Phase 3. A single implementation of the solution in actual educational settings is not likely to be sufficient to collect evidence indicating the solution's success.

The research objectivity, validity, and reliability were also challenged due to the lack of researcher triangulation and my multiple roles (i.e., researcher, designer, and evaluator). Different attempts were made to prevent a possible conflict between my roles. As a designer, I considered design evaluation a means to collect feedback to improve the design rather than demonstrating its perfection. As recommended by Edelson (2002), I acted as an EDR researcher to develop a novel solution to improve educational practice and use design implementation to establish theoretical knowledge.

6 RESEARCH QUALITY EVALUATION

The research quality of this study was evaluated at two levels, as recommended by Juuti and Lavonen (2006). At the first level, the quality of the whole EDR study is assessed. At the second level, the quality of the mixed methods research design employed as the strategy of enquiry is assessed. Both evaluation levels are presented separately for clarity. Unavoidably, this method of presentation results in repeated information.

6.1 Quality in educational design research

EDR has different objectives and characteristics from traditional empirical research; therefore, its quality should be evaluated differently from that of traditional research (e.g., Edelson, 2002; Kelly, 2004; McKenny & Reeves, 2019; Phillips, 2006). Because EDR aims to produce new theories and solutions that contribute to the improvement of educational practice, the *novelty* and *usefulness* of EDR can be used to assess its quality (Edelson, 2002; Juuti & Lavonen, 2006). This study provided novel and useful theoretical and practical outcomes; thus, it can be argued that it achieved the goals of EDR. Regarding novelty, a novel design solution—a TM, student worksheets, teacher guides, and class activities—was developed to promote students' understanding of equation-solving concepts and classroom practice (see Chapter 4.3.1). Additionally, the study constructed three kinds of knowledge: (1) knowledge of the challenges and opportunities of using manipulatives in mathematics classrooms, the strengths and limitations of existing manipulatives, and how the proposed solution encouraged students' understanding of equation-solving concepts and the classroom adoption; (2) a design framework for real-world educational technologies; and (3) guidelines for conducting EDR (see Chapter 7.1). Regarding usefulness, the design evaluation results indicate that the developed solution not only enhanced students' equation-solving learning and achievement, but it also conformed to classroom practice (see Chapter 4.3.2). All teachers who participated in the design evaluation stated that they would acquire the manipulative for their equation-solving instruction. Therefore, the developed solution has the potential to be useful in mathematics classrooms.

Objectivity, *validity*, and *reliability* can be used to address rigour in EDR, as in traditional empirical research; however, these qualities are established rather differently than in traditional research (DBRC, 2003). In EDR, researchers are typically also designers, implementors, and evaluators of the solution, so these potentially conflicting roles make it challenging for researchers to maintain their objectivity, that is, the absence of bias (e.g., DBRC, 2003; McKenny & Reeves, 2019; Plomp, 2013; Publication I). The designer's presence in the evaluation of solutions can also cause potential bias. When research participants are aware that the researcher has designed the solution, they may respond differently, and the designer may intentionally or unintentionally not be open to critique (McKenny & Reeves, 2019).

To address these threats to the objectivity of this study, triangulation of multiple data sources (i.e., students and teachers), data collection methods (e.g., observations, interviews, and questionnaires), and data types (i.e., quantitative and qualitative) was employed, as recommended in the literature (e.g., McKenny & Reeves, 2019; Plomp, 2013). Teachers were the designated implementors of all class interventions, so there was no influence of the researcher or designer on any interventions. I also saw the concept and design evaluation as an opportunity to gain information for developing the design solution and theoretical knowledge instead of a showcase for demonstrating the solution's perfection (see e.g., DBRC, 2003; McKenny & Reeves, 2019; Ulrich & Eppinger, 2016).

The validity of this study is assessed on three aspects: systemic, internal, and external validity. The study developed knowledge of how to use manipulatives to promote students' understanding of equation-solving concepts and classroom practice and then used the developed knowledge to design a TM and accompanying materials to improve educational practice. Therefore, the study achieved *systematic validity* (i.e., the extent to which practice is informed by theories, which are informed by the study), which is particularly important for EDR (Hoadley, 2004), and exhibited characteristics of good EDR (DBRC, 2003).

According to Collins et al. (2004), methodologically, EDR differs from experimental research; for example, regarding research location (messy real-world settings vs. controlled laboratories), variable quantities (multiple dependent variables vs. a single dependent variable), variable treatment (identifying all variables during the evolving study vs. selecting a few variables in advance and constantly controlling them), and research design (adaptable procedures as the study unfolds vs. fixed procedures according to plan). Given these differences, EDR does not make simple causal claims in complex, real-world educational contexts (Kelly, 2006). Consequently, it is not an easy task to convince others about *internal validity* (i.e., the extent to which valid inferences can be drawn from the research results; e.g., Creswell & Creswell, 2018) of EDR (e.g., Barab, 2006; DBRC, 2003).

In this study, internal validity was enhanced in several ways, as recommended in the literature. While giving up control of research settings, variables, and research design; the multiple iterations of investigation, design and construction, and evaluation and reflection established justification for the research results, theoretical knowledge, and design solution (DBRC, 2003; Edelson, 2002, 2006; McKenny & Reeves, 2019). Moreover, in line with DBRC's (2003) and Ørngreen's (2015) recommendation, the involvement of teachers and students—potential users of the design solution—in the empirical study largely contributed to the rigour of the research findings. Triangulation can enhance the internal validity of EDR (DBRC, 2003; McKenny & Reeves, 2019; Plomp 2013). In this study, in addition to the triangulation of data sources, data types, and research methods, the triangulation of theoretical frameworks (e.g., constructivism, representational fluency, and multimodality) was also employed. Reporting EDR to a research community can also enhance the rigour of the research findings (Juuti & Lavonen, 2006; McKenny & Reeves, 2019). Apart from my supervisors, I discussed the study with my research group and other fellows in the faculty to obtain their feedback for developing the research design and analysing the data. I also presented the initial findings of each research phase at four conferences to communicate with other scholars during the ongoing process. Feedback received from the conferences assisted me in finalising data analysis and interpretation, as well as writing Publications II–IV and this dissertation. All four publications underwent a peer-review process in which reviewers (i.e., external auditors) assessed the validity of the research findings (Creswell & Creswell, 2018). Additionally, the publications make the study open to an audit by scientific community. Each publication transparently provides sufficient information (e.g., to-be-solved educational problem, target educational setting, and rationale for design choices, besides traditionally reported research methods and results) to allow readers to evaluate for themselves the rigour of the findings.

External validity is the extent to which the research results can be transferable/generalisable beyond the study context to other populations, settings, and times (e.g., Creswell & Creswell, 2018). The findings of EDR, similar to those of case and experimental studies, cannot be statistically generalised from sample to population, as in traditional empirical studies, such as survey research (e.g., Juuti &

Lavonen, 2006; Kelly, 2006; Plomp 2013). Being contextualised in nature, EDR seeks to generalise knowledge gained from its context to a broader theory that is applicable to other educational contexts (Hoadley, 2004; Juuti & Lavonen, 2006; Plomp 2013). It is noteworthy that each educational setting is unique and full of multiple variables; therefore, generalised knowledge does not provide certainties, but rather, guidance and direction (Hoadley, 2004; Plomp, 2013).

Generalisability of this study was promoted in different manners as recommended by EDR scholars. The empirical studies involved both teachers and students and were conducted in their classrooms to account for the reality of how teaching and learning occur in a complex social context (e.g., Brown, 1992; McKenny & Reeves, 2019). Various research methods have been employed to enhance the holistic study of educational phenomena in messy real-world settings (e.g., Anderson & Shattuck, 2012; McKenny & Reeves, 2019). The development of knowledge and the design solution was also driven by the literature, previous research, and empirical studies of this EDR (Edelson, 2002, 2006). Throughout the study, the design process was thoroughly and systematically documented to promote retrospective analysis, which is important to EDR (Edelson, 2002, 2006). Publication IV also provides a description of problem analysis, target educational context, design process, and solution design and construction so that others will be able to relate the knowledge gained from this study to similar educational contexts (McKenny & Reeves, 2019). As recommended by Edelson (2002, 2006) and Plomp (2013), the generalised knowledge (i.e., domain theory, design framework, and design methodology; see Chapter 7.1) was constructed and refined during the multiple iterations. For example, in line with Juuti and Lavonen's (2006) and Plomp's (2013) recommendations, the design principles were inductively generated and refined to inform successful design solutions during the iterations. Altogether, the research results of this study should be applicable to other educational settings, at least in the context of Finnish schools, which, according to OECD (2017, 2020), are relatively similar to this study context regarding the education system and teacher quality, as well as students' mathematical performance.

Reliability refers to the extent to which a study produces consistent findings when repeated (e.g., Creswell & Creswell, 2018). It is typically challenging to achieve reliability in EDR. Unlike controlled laboratory studies, EDR is difficult to replicate precisely (DBRC, 2003). This is because EDR is situated in complex real-world settings full of various variables, and its research design is often altered as the study evolves (e.g., Collins et al., 2004; Hoadley, 2004; Phillips, 2006), as was the case with this EDR. An example of an uncontrolled environment is that one student was upset during the class intervention because of an argument with her classmate right before the intervention, and thus was possibly not at her best during the intervention.

The research design was also modified during the study; for example, the worksheet and paper-based test used during Phase 3 differed slightly from those used during Phase 1. The Phase 3 worksheet and test were modified based on lessons learnt from Phase 1 and the emphasis of the Phase 3 investigation. Reliability of this study was promoted through triangulation, as recommended by EDR scholars (e.g., DBRC, 2003; Juuti & Lavonen, 2006; McKenny & Reeves, 2019). Based on the recommendations of DBRC (2003) and McKenny and Reeves (2019), reliability was also improved through multiple iterations of research activities. Moreover, the research methods and procedures, as well as the justification for the research design alteration, were documented so that other researchers could replicate them.

6.2 Quality in mixed methods research

The quality of a mixed methods study depends on the quality of its quantitative and qualitative parts, so the quality of its individual parts should be also evaluated (Fàbregues & Molina-Azorín, 2016; Ihantola & Kihn, 2011). Following this recommendation, the quality of the quantitative and qualitative parts of this study were first evaluated separately using the evaluation criteria of each. The quality of the quantitative part was assessed using common criteria for quantitative research: internal validity, external validity, and reliability (for the definition of each criterion, see Chapter 6.1). The quality of the qualitative part was assessed using Ryan et al.'s (2002) alternative criteria for case studies: contextual validity, transferability, and procedural reliability. Then, the quality of the overall mixed methods research was evaluated.

The internal validity of the quantitative part of this study was ensured through various actions. Relevant literature and previous studies from different continents were reviewed to construct a theoretical framework and assist the research result interpretation. The students participating in the class interventions were equally distributed among the instructional groups in terms of numbers and attainment levels. The class interventions during Phases 1 and 3 were well controlled. For example, the interventions of each instructional group used identical research methods, procedures, and instruments; the same teacher taught all instructional groups of their own classrooms the same content for the same period of time. Cross contamination of different instructional groups was minimised by administering the paper-based test for all student groups of each classroom at the same time, right after the last intervention. Students who failed to take part in all research activities were excluded from the research.

Nevertheless, there were some threats to the internal validity of the quantitative part. All class interventions were conducted in real-world educational settings (instead of a controlled laboratory setting); thus, the differences between study contexts might influence the research findings. Despite the homogenous quality of Finnish teachers, the various backgrounds and experiences of the teachers might also affect the results. Additionally, it was not possible to conduct all interventions in each classroom at the same time or to keep the students participating in different instructional groups separately during the study. This resulted in the possibility that students in different groups communicated with each other, causing potential cross contamination of groups.

Ryan et al. (2002) defined *contextual validity* as the credibility of evidence and the conclusions drawn from a case study (pp. 155–156; cf. Lincoln and Guba's [1985] *credibility* and Creswell and Creswell's [2018] *qualitative validity*). In this study, the contextual validity of the qualitative part was enhanced in a number of ways. Conducting the research in authentic educational settings provided a better understanding of the actual contexts. Cross-sectional field studies of classrooms at different grade levels (i.e., primary and lower secondary) also broadened the contexts under study. As previously mentioned, different data sources, research methods, and theories were triangulated throughout the study. Clear and standard interview instructions and questions were used to ensure mutual understanding of all participants, and being in a familiar environment (their own classrooms as opposed to laboratory settings) was likely to help the participants act more naturally. The validity of the results was evaluated by external auditors through conference and journal reviewing processes.

In spite of this, threats to the contextual validity of the qualitative part were recognised. While conducting this study alone guaranteed conformity between all research activities, the lack of researcher triangulation may have risked the contextual validity of the study. On the one hand, the cross-sectional field studies promoted a better understanding of broader educational contexts. On the other hand, it required significant resources, thereby limiting the class intervention duration to only 45 minutes per instructional group. The short class interventions possibly resulted in a limited understanding of each instructional group. Furthermore, all the research activities were conducted in Finnish, which is the participants' mother tongue, to facilitate their communication. Finnish is not my native language, even though I am fluent, so there could still be some language barriers that might threaten the accuracy of the communication between the participants and myself, as well as the data transcription and interpretation.

The external validity of the quantitative part of this study was promoted through different means. The uniform quality of Finnish schools in terms of teachers' qualifications and students' mathematics performance allowed for convenience sampling and increased population and environmental validity. Thus, inferences drawn from the primary and lower secondary classroom settings under study are likely to be generalisable to other Finnish educational settings and populations of the same grade levels. Nevertheless, there are a few threats to the external validity of the research. The sample size of students participating in each instructional group was relatively small for statistically generalising the results to individuals in other contexts, and rapid technological development makes it difficult to generalise findings regarding educational technologies to future situations.

Ryan et al. (2002) used the term *transferability* to refer to two types of theoretical generalisations: (1) refinement and generalisation of theory to a broader context and (2) applicability of research results to other settings (pp. 149–150; cf. Lincoln and Guba's [1985] *transferability* and Creswell and Creswell's [2018] *qualitative generalisation*). Transferability of the qualitative part is rather similar to the external validity of EDR in Chapter 6.1, and was strengthened through different actions; for example, by conducting research in real classroom environments, investigating both pedagogical and practical aspects of the classroom, collecting data from both teachers and students, undertaking cross-sectional field studies (i.e., several cases), triangulating data sources and research methods, comparing the findings to the literature and previous studies, constructing and refining theoretical knowledge through iterations, and documenting research contexts, procedures, and findings well. However, a potential threat to the transferability of the qualitative part was the relatively short duration of each class intervention.

The reliability of the quantitative part of this study was enhanced in a number of ways. Research instruments, including instructional materials (i.e., student worksheets and teacher guides), paper-based tests, and questionnaires were informed by existing instruments, the literature, previous studies, school textbooks, and the input of mathematics education experts. Clarity and conformability of research activity instructions, class intervention materials, and measuring instruments were ensured to avoid participants' misinterpretation and reduce errors of measurement. Sufficient indicators (e.g., different types of test items and questionnaire items) were used to measure students' mathematical concept understanding and performance, as well as teachers' opinions and perceptions. All classroom interventions were controlled in terms of classroom environment, numbers and characteristics of participants, instructional tools, procedures, and duration.

However, there were several threats to the quantitative research reliability. The empirical research did not take place in a laboratory environment, but rather in real classroom settings. This made it difficult to rule out the influence of confounding variables on the research results. The teachers were allowed to adjust their class intervention lessons according to classroom dynamics, as long as the adjustment did not conflict with the intervention instructions. While the adjustment reflected real-world phenomena and thus increased the contextual validity of the qualitative part, it threatened the reliability of the quantitative part.

The reliability of some research instruments was somewhat questionable for several reasons. For example, interrater reliability was not possible in this single-researcher study. The research instruments did not undergo a pilot test due to time constraints and limited resources, and some parts of the measuring instruments were particularly designed for this study because appropriate instruments did not exist. Nevertheless, the developed instruments were reviewed by mathematics education experts and then revised accordingly.

The Cronbach's alpha of the paper-based test, all student questionnaire scales, and almost all teacher questionnaire scales (Cronbach's alpha of one scale was 0.53, possibly due to a small number of scale items and participants) were above the generally acceptable level (see Hinton et al., 2014). This indicates the internal consistency of the measuring instruments. The last threat to reliability was my inexperience in quantitative research. This threat was minimised through consulting experts and the literature about quantitative research design, analysis, and interpretation, as well as developing my quantitative research skills.

Ryan et al. (2002) proposed *procedural reliability* (i.e., the use of suitable and trustworthy research methods and procedures; p. 155; cf. Lincoln and Guba's [1985] *dependability* and Creswell and Creswell's [2018] *qualitative reliability*) as analogous to traditional notions of reliability for quantitative research. In this study, the procedural reliability of the qualitative part was ensured in different ways. Research design, methods, and procedures, as well as data transcriptions, analysis, and interpretation were systematically and descriptively documented and reported. The interview questions were informed by the RQs, the literature, and previous studies. Furthermore, the interview form assisted in systematic interviewing, whereas the face-to-face interviews allowed all interview questions to be posed to the participants unambiguously. All interviews and class intervention observations were audio- or video-recorded; notes were taken to capture as much data as possible; the data sources and data collection methods were triangulated. However, a potential threat to procedural reliability was recognised in that working alone in this study did not allow for researcher triangulation.

Over the past two decades, scholars have proposed various quality criteria for mixed methods research, ranging from lengthy and comprehensive frameworks (e.g., Onwuegbuzie & Johnson, 2006; Tashakkori & Teddlie, 2008) to minimal sets of core criteria (e.g., Bryman, 2014; Creswell & Plano Clark, 2017). The latter type of quality frameworks with minimum key criteria are easier to use, more flexible to accommodate diverse research contexts, and easier for others to understand (Fàbregues & Molina-Azorín, 2016). Thus, the quality of this mixed methods study is assessed using one of the latter type of frameworks, which was proposed by Creswell and Plano Clark (2017). A good mixed methods study should have the following four core characteristics (Creswell & Plano Clark, 2017, pp. 282–284):

1. *Rigorous quantitative and qualitative data collection and analysis in response to research questions/hypotheses*. In this study, quantitative and qualitative research methods were selected and implemented based on their ability to answer the RQs (see Table 1). The quantitative strand addressed 'whether' and 'to what extent' questions, whereas the qualitative strand addressed 'what' and 'how' questions. For example, during Phase 1, RQ 1.1 was answered using both research strands. The teacher interviews and classroom intervention observations were used to address the challenges and opportunities of using manipulatives in the target settings, as well as the strengths and limitations of existing manipulatives. The paper-based tests and student self-evaluations were used to address how well each existing manipulative enhanced students' learning and achievement compared to those of students learning without manipulatives. Throughout the study, the two research methods were implemented following the practices of each tradition. The quality of each strand and threats to its quality were addressed at the beginning of this subchapter.
2. *Intentional integration of both forms of data and their findings*. In this study, the convergent design of mixed methods research (see Chapter 3.4) was employed as a strategy of enquiry for confronting the complex research problem. Throughout the study, both qualitative and quantitative data were simultaneously collected and analysed, then concurrently triangulated and combined to holistically address RQs 1–2. For example, during Phase 2, the teacher questionnaire and interview were conducted

at the same time. The questionnaire was used to assess teachers' perceptions of how well each design concept was likely to promote students' learning and comply with real-world practice. After each questionnaire item, the teachers were interviewed to provide an explanation for their questionnaire responses. Moreover, the interpretation of the empirical findings in Publications II–IV and Chapter 4 explicitly brings together both forms of data.

3. *Logical research design, in which the key elements of mixed methods research fit well together*. To address the research purpose and RQs, the research design of this study was planned and implemented, so that the strengths of the quantitative and qualitative strands compensated for the weakness of the other. The triangulation of quantitative and qualitative findings about the same phenomenon was used to enhance the confidence of the findings, thereby promoting research validity. For example, during Phase 3, the paper-based test was conducted to assess the learning achievement of the students participating in the class interventions compared to the comparison group. Nevertheless, the sample size of the students in the interventions was rather small, and thus, the interpretation of their test score analysis was somewhat questionable. To deal with this limitation, the quantitative findings were compared with the qualitative findings from the observations, interviews, and thinking aloud sessions, which not only confirmed but also explained the quantitative findings.

4. *Research procedures were informed by theoretical/philosophical considerations.* The research procedures of this study were framed within an EDR paradigm. During Phase 1, different qualitative data were obtained to develop an in-depth understanding of the research problems, complex real-world educational contexts, and existing manipulatives (RQ 1.1). At the same time, quantitative data were collected mainly to investigate the impact of different manipulatives on students' learning achievement. After Phase 1, both research strands played a relatively equal role in studying the same variables. During Phase 2, the quantitative results assisted in selecting the promising concept for further development; the qualitative results helped to understand the reasons behind teachers' concept evaluation, which provided useful information for developing the design principles and design solution. During Phase 3, the quantitative and qualitative data were used together to evaluate the developed design solution (RQ 1.2).

Despite meeting all four criteria for good mixed methods research, two key potential threats to the quality of this convergent design study were acknowledged. First, in most parts of Phase 1, different variables were used to collect quantitative and qualitative data, thereby making it difficult to merge the findings. It should be noted that the use of different variables on both strands was intended to connect findings regarding different quantitative and qualitative variables to establish an overview contextual understanding. Second, unequal sample sizes of the quantitative and qualitative data in some parts of the study (e.g., out of six teachers who took part in Phase 3 questionnaires and interviews, only two teachers participated in the class interventions) were likely to provide an unequal picture of both research parts. Because of limited resources, this was unavoidable.

Apart from that, most parts of the study achieved equality in both sample sizes. For example, during Phase 1, the same students participated in both the qualitative (class intervention) and quantitative (paper-based test and self-evaluation) parts. Students who failed to participate in both parts were excluded from the study. The qualitative sample was increased in an attempt to achieve equality. However, this limited the amount of qualitative data collected from each individual, as well as the quantitative sample.

7 DISCUSSION

This EDR seeks to contribute to both practice and theory in mathematics education. The research-based design solution—the TM, student worksheets, teacher guides, and class activities—is the practical outcome of this research to directly improve educational practice (i.e., promoting students' understanding of equation-solving concepts and classroom adoption). Additionally, the research contributes to the three types of theories (domain theory, design framework, and design methodology) proposed by Edelson (2002). In this chapter, I summarise the research's theoretical outcomes and then reflect on the implications, limitations, and future research.

7.1 Theoretical outcomes

7.1.1 Domain theory

Domain theories are twofold descriptive knowledge about a real-world educational problem to be solved (context theory) and outcomes of implementing a design solution to solve that particular problem (outcomes theory). This chapter first outlines the context theory addressing RQ 1.1 and then the outcomes theory addressing RQ 1.2. The context theory was reported in Publications II and IV; the outcomes theory was presented in Publications III–IV.

RQ 1.1: What are the needs, challenges, and opportunities of using manipulatives in primary school classrooms? What are the strengths and limitations of existing manipulatives? (context theory)

Previous research has found mixed results regarding the effectiveness of manipulative use on students' mathematics learning and achievement (e.g., Manches & O'Malley, 2012; Uribe-Flórez & Wilkins, 2017). In line with the literature (e.g., Carbonneau et al., 2013; McNeil & Jarvin, 2007), the initial research results demonstrate that the use of manipulatives can help students construct their understanding of abstract mathematical (in this case, equation-solving) concepts.

When learning equation solving with physical or virtual forms of the same manipulative or without manipulatives through multiple representations, discovery, peer interaction, reasoning, and reflection, the PM most benefited primary school students' learning and performance. The findings on the superiority of learning with PM over other instructional conditions contradict previous research findings. Suh and Moyer's (2007) study reported that learning with both PM and VM were effective equation-solving instructional conditions, whereas Magruder's (2012) study reported that learning without manipulatives was the most effective. These contradictory findings may be because, in these two previous studies, students used manipulatives to learn equation solving mostly though discovery and multiple representations, while in this study, students were encouraged to learn through peer interaction, reasoning, and reflection as well. Moreover, the intervention duration difference may also have had an impact on the findings.

The initial research also discovered strengths and limitations of each manipulative, similar to those found in Magruder's (2012) and Suh and Moyer's (2007) studies. For example, the tactile features of PMs enhanced learning and made manipulatives easy to use. VMs explicitly link pictorial and symbolic representations of equations as well as provide immediate feedback and guidance, but can lead to rote learning. Additionally, the initial research reveals evidence of PM benefits and VM hindrances to students' peer interaction and verbalisation of their mathematical thinking. These two classroom activities appeared to play a significant role in how each manipulative benefited students' equation-solving learning and performance.

The initial research results regarding day-to-day practice are in agreement with the literature (e.g., Bedir & Özbek, 2016; Hatfield, 1994; Marshall & Swan, 2008). Practical issues, including time constraints, manipulative availability, and manipulative organisation, can potentially prevent classroom utilisation and adoption of manipulatives. The Phase 2 design evaluation results also confirm that day-to-day practice has an important role to play in teachers' decisions to acquire manipulatives.

RQ 1.2: How does the developed design solution help students understand equation-solving concepts and encourage teachers to adopt it in their classrooms? (outcomes theory)

Highly desirable outcomes from the Phase 3 design evaluation suggest successful classroom utilisation and adoption of the developed design solution. All in all, *X*-is, the developed TM, is likely to be adopted in the classroom due to its pedagogical benefits and compatibility with school and classroom practice.

In terms of pedagogy, *X*-is better supported students' understanding of equation-solving concepts through discovery learning, peer interaction, and multimodal expression of mathematical thinking compared to learning without it. The evaluation results are consistent with those of Starcic et al. (2013) and Zamorano Urrutia et al. (2019), who found that TMs assisted students in developing their understanding of mathematical concepts and encouraged peer interaction during mathematics learning. The present study also supports evidence from previous observations by Zamorano Urrutia et al. (2019) that TMs stimulated students to autonomously experiment and discover to-be-learnt mathematics content. In accordance with the literature (e.g., Manches & O'Malley, 2012; Price, 2013), the present findings demonstrate that the combined strengths of physical and digital parts of the TM appear to contribute to the desirable outcomes regarding pedagogy.

The design evaluation results also demonstrate that *X*-is is likely to be usable and compatible with school and classroom practice. Consistent with the literature (Manches & O'Malley, 2012), this research found that *X*-is was easy to use because students were not required to learn how to manipulate the physical objects. Similar to previous findings found by Salvador et al. (2012) and Sapounidis and Demetriadis (2013), all students considered *X*-is pleasant to use and intended to use it again. The teachers also regarded *X*-is as highly compatible with classroom and school practice. Their responses indicated the possibility of the adoption of *X*-is in primary school classrooms and its use in pre-primary and lower secondary school classrooms.

7.1.2 Design framework

Design frameworks describe the key characteristics of successful design solutions to a particular educational problem in a particular setting (Edelson, 2002). This chapter reports a *real-world educational technology design framework* that addresses RQ 2. The design framework was presented in Publication IV.

RQ 2: What key aspects should be taken into account when developing a manipulative to ensure its successful classroom utilisation and adoption?

Different types of knowledge are required for the successful development of solutions to improve real-world educational practice (Brown, 1992). As described in Chapter 5.1, I needed to acquire and synthesise different knowledge throughout this EDR to develop the solution, which appeared to promote students' understanding of equation-solving concepts and its classroom adoption. Based on this experience, I have developed a design framework with usable and generalisable knowledge of what should be considered when developing educational technologies for real-world educational contexts. The framework takes into account four essential aspects—content, pedagogy, practice, and technology—that play a part in educational benefits, feasibility, and classroom utilisation and adoption of educational technologies (Figure 9).

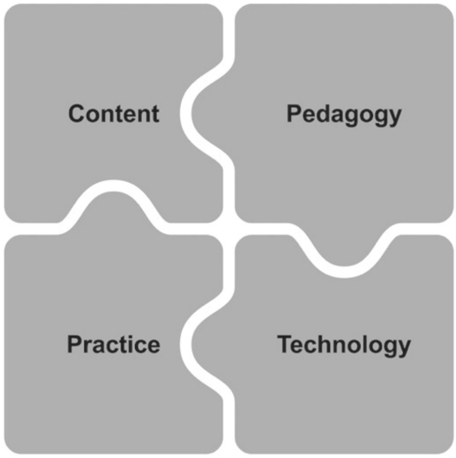


Figure 9. The real-world educational technology design framework takes into account content, pedagogy, practice, and technology. From 'Constructing a Design Framework and Design Methodology from Educational Design Research on Real-World Educational Technology Development,' by D. Lehtonen, 2021, *Educational Design Research, 5*(2), Article 38, p. 19 (https://doi.org/10.15460/eder.5.2.1680). CC BY 4.0.

An understanding of content to be learnt is essential for the development of any educational solutions. The development of *X*-is and the instructional materials required knowledge of equation solving, important concepts required for understanding equation solving, different models for teaching equations, and equation-solving content stated in the current Finnish NCC (EDUFI, 2016) and used in textbooks.

Knowledge of pedagogy is needed to ensure the meaningful use of educational solutions. Learning through discovery and social interaction, multimodal expression of mathematical thinking, and teaching and learning approaches recommended by the NCC (EDUFI, 2016) were used to inform the design of *X*-is, instructional materials, and class activities to enhance students' understanding of equation-solving concepts.

An understanding of practice in the target educational context contributes to the success of educational solutions in the real world. Considering practical issues such as class management when developing *X*-is proved to play an important part in teachers' possible acquisition.

It is essential to know about technological possibilities to design feasible educational technologies. To develop *X*-is with computer science students, I required knowledge of digital technologies, tangible technologies, and object-tracking alternatives. When developing educational solutions, technology should be used because of its appropriateness and contribution, instead of for its own sake, to avoid the *technology before pedagogy* effect (see D. M. Watson, 2001; cf. technology-driven products; Ulrich & Eppinger, 2016). In this study, the design concepts were not driven by technologies but instead by what technologies could offer to help students learn equation-solving concepts.

It is noteworthy that the importance of each aspect in the framework is typically unequal and depends on the nature of the educational problem, target context, and possible technologies. Consistent with that of McKenney and Reeves (2019), design decision making of this EDR often involved simultaneous consideration of various aspects. Similarly to what Ulrich and Eppinger (2016) described, *trade-off decisions* were often made during the design development to fine-tune the design that would best meet the research aims, for example, the change of object tracking as described in Publications III–IV. The understanding of theoretical background, classroom practice, and technologies, derived from the literature, fieldwork, and multidisciplinary collaboration, assisted in making the trade-off decisions.

7.1.3 Design methodology

Design methodology provides guidelines for conducting EDR to achieve the research objectives (Edelson, 2002). Built on the literature and my reflection on my own experience from undertaking this EDR, this chapter addresses RQ 3. The guidelines were reported in Publications I and IV.

RQ 3: What guidelines for conducting successful EDR can be drawn from the lessons learnt from undertaking this study?

To date, various EDR models (e.g., Easterday et al., 2017; McKenney & Reeves, 2019) have been adopted in different EDR projects. For example, the overall process of this EDR was based on McKenney and Reeves' (2019) model (see Chapter 3.3). Usually, some adaptation of these models is required to meet the uniqueness of each EDR project. Thus, rather than proposing another model, I present general guidelines (Figure 10) for conducting EDR to help other researchers embrace opportunities and overcome the challenges that may emerge from their EDR.

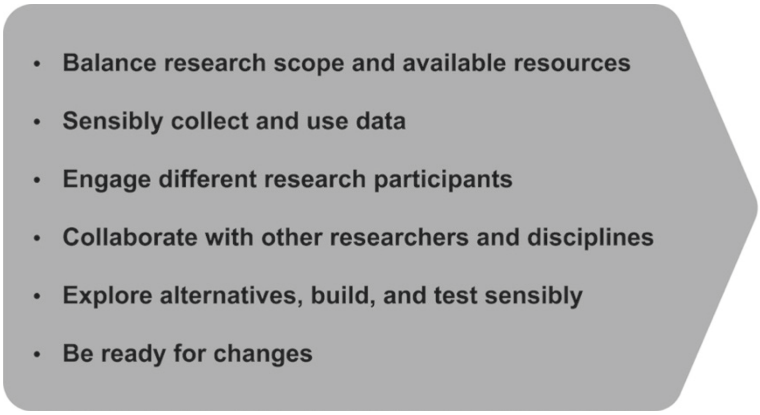


Figure 10. General guidelines for conducting EDR. From 'Constructing a Design Framework and Design Methodology from Educational Design Research on Real-world Educational Technology Development,' by D. Lehtonen, 2021, *Educational Design Research, 5*(2), Article 38, p. 22 (https://doi.org/10.15460/eder.5.2.1680). CC BY 4.0.

This doctoral study and those of others (see Publication I) have demonstrated that EDR is resource intensive, and thus it is important to assess the available resources and adjust the scope of a project correspondingly. This helps to ensure that the project is accomplishable and that real-world iterative enquiry is preserved. In line with Kennedy-Clark's (2013) recommendation, the research team size should be appropriate to the intensiveness and scale of the project. Additionally, disciplines of project members (i.e., monodisciplinary vs. multidisciplinary) and other resources (e.g., time and budget) should be considered when planning EDR. A multifaceted and long-term project can be divided into achievable parts, for example, doctoral and postdoctoral research (Goff & Getenet, 2017) or small-scale studies for doctoral students to undertake individually (Anderson & Shattuck, 2012). Findings from a single iteration can be used to inform further research (Di Biase, 2020).

As in the EDR literature (e.g., McKenney & Reeves, 2019; Plomp, 2013) and this study, data triangulation can promote an understanding of complex real-world phenomena and EDR trustworthiness. Data triangulation often requires intensive resources to collect substantial amounts of data (Collins et al., 2004), most of which are left unanalysed due to limited resources (Kelly, 2006). This is often the case with many doctoral students, such as Goff (2016, as cited in Goff & Getenet, 2017) and me. Thus, data collection and analysis should be well planned and implemented by balancing resource use and data triangulation. Additionally, sharing collected data (i.e., open data) can progress the research community (Collins et al., 2004).

It has been shown in my study and those of others (e.g., Cowling & Birt, 2018) that stakeholders' engagement in EDR is essential for the success of EDR. My experience supports the literature (e.g., McKenney & Reeves, 2019; Ørngreen, 2015) that direct users and other relevant stakeholders should be involved in different phases of EDR. My experience from this study supports Herrington et al.'s (2007) suggestion that stakeholder participation should be intended to profit both the EDR and themselves. Stakeholder participation should be well planned and implemented by taking into account practical and ethical issues (e.g., research permission, interruptions to their normal activities, and suitable timing for all involved).

It is feasible for a researcher like Di Biase (2020) and me (in most parts of my study) to conduct EDR alone. Nevertheless, EDR usually requires multidisciplinary expertise in both research and design (Collins et al., 2004; Edelson, 2006). In line with the literature (e.g., Kennedy-Clark, 2013; McKenney & Reeves, 2019), the contribution of my supervisors and the computer science students enhanced the feasibility and trustworthiness of this study. Fruitful collaboration can be ensured through good teamwork and communication (McKenney & Reeves, 2019).

While in the design field, it is important to explore alternatives before selecting potential designs for further development (e.g., Ulrich & Eppinger, 2016), only a few EDR researchers have clearly discussed working with alternative solutions (Publication I; Ørngreen, 2015). As demonstrated in my study, alternative solution exploration is likely to ensure that the design developed with substantial resources is the best solution to the educational problem (McKenney & Reeves, 2019; Ørngreen, 2015). Furthermore, my study supports Easterday et al. (2017) that to use resources efficiently, design construction (mock-ups vs. working prototypes) and evaluation methods (e.g., quick surveys vs. class interventions) should be employed according to the stage of design (concept exploration vs. design development) and researchers' theoretical knowledge level.

EDR is carried out in complex, real-world settings full of variables, so the research design of EDR usually evolves during the iterations (e.g., Herrington et al., 2007; Plomp, 2013), as was the case with my study. Unpredictable situations (e.g., in my case, a teacher's withdrawal from the study on short notice and the unreliably functioning prototype) can also occur. Therefore, I agree with Kennedy-Clark's (2013) recommendation that EDR researchers should be prepared for adjustments and changes.

7.2 Research contributions

This EDR contributes to theory and practice in mathematics education. Regarding the theoretical contribution, this research advances three types of usable knowledge. First, the domain theory adds knowledge about the needs, challenges, and opportunities of using manipulatives in primary school classrooms and how multimodal interaction with manipulatives, particularly TMs, in the classroom social context can enhance students' understanding of mathematical concepts. Second, the real-world educational technology design framework informs educational designers about taking important aspects into consideration when developing educational technologies. The framework helps ensure that the development of educational technologies will benefit teaching and learning, be realisable, and succeed in the real world. Third, the guidelines for conducting EDR can assist other researchers in embracing opportunities and overcoming the challenges that may emerge from their EDR.

Regarding practical contributions, the research-based design solution is the direct outcome of this research to directly assist students in understanding equation-solving concepts and enhance day-to-day practice. The research also has practical implications. First, it encourages teacher educators to prepare pre- and in-service teachers for the successful incorporation of manipulatives in their mathematics classrooms. Second, it guides practitioners in how to support their students to benefit from manipulatives. Third, it urges schools to support the acquisition and utilisation of manipulatives. Finally, it calls for school curricula to encourage the use of manipulatives in the mathematics classroom to support students' conceptual understanding.

7.3 Limitations and future research

As mentioned in Publications II–IV, the limitations of the current study have been acknowledged, and possibilities for future research have also been recognised. The research quality might have been affected by limitations regarding research methodology: small convenience sample size, short interventions, only posttests employed, questionable reliability of some instruments, lack of researcher triangulation, and the researcher's inexperience of quantitative research. Future research that has larger random samples, prolongs interventions, uses pre-post-test design, ensures reliability of all instruments, has more than one researcher to achieve researcher triangulation, and develops/ensures sufficient research skills of researcher(s) would yield favourable quality. Moreover, research employing control (traditional classroom practice) and experimental groups (with the developed design solution: *X*-is, instructional materials, and class activities) from the same grade level would help to better determine the impacts of the design solution on students' equation-solving achievement compared to the traditional mathematics classroom instruction. As reported in Publication III, it was not possible to transcribe the students' dialogues during the Phase 3 pair work due to the classroom noisy background. To better understand the contributions of the developed TM (*X*-is) to students' peer interaction, lavalier (clip-on) microphones could be used to record pairs' communications in the future.

The quality of this EDR suffered from limited human resources. As only one researcher worked in this doctoral study, only one concept evaluation and one classroom implementation of the design solution was conducted. A single implementation of the solution in a real-world educational context is unlikely to be enough to collect evidence that indicates the success of the proposed solution. Postdoctoral research, in which the solution will be refined based on the doctoral research results and working prototypes of the refined solution will be constructed and tested in the real world, would contribute to the research validity. A multidisciplinary team working with full commitment would enhance EDR quality.

Regarding domain knowledge of manipulatives for understanding of mathematical concepts, this research concentrated only on the use of manipulatives, particularly TMs, for promoting mixed-attaining primary school students' understanding of equation-solving concepts and their classroom adoption. Further research could continue to investigate these particular aspects in more depth. To better understand the potential of TMs in mathematics classrooms, future research could investigate the benefits of TMs for students with a specific degree of attainment, with visual impairment, or at different educational levels, as well as other mathematics topics and other forms of education (e.g., distance learning). Future research focusing on the influence of schools and policymakers on classroom implementation and adoption of manipulatives would contribute to manipulatives' success in the real world. Regarding design framework and methodological knowledge of EDR in technology-enhanced mathematics learning, the proposed framework and guidelines were only informed by a single researcher's experience from one EDR project. More similar EDR projects by other researchers would help validate the results of this study and advance the design framework and methodological knowledge of EDR.

Regarding technology, the results of this EDR might have been influenced by the limitations of the Wizard of Oz prototype used during the Phase 3 class interventions. A fully working reliable prototype with complete functions and features would overcome these limitations and expanded the assessed attributes of *X*-is. Future development of tangible technology may allow *X*-is to detect objects without external devices and connections, and thus increase its usability, practicality, and reliability. Added features, such as free experiments with the balance scale and students' own equation setup, would enhance the user experience.

The current design of *X*-is is unable to represent equations containing negative integers and subtraction. With new technologies, future design could overcome this limitation and thus broaden the potential benefits of *X*-is in equation-solving learning. On the other hand, mathematics education scholars (e.g., Goldin, 2002; A. Watson, 2009) recommend that concrete models should be used to help novice students construct their equation-solving foundations. From this perspective, future design should assist students in moving away from nonsymbolic representations (working with the manipulative) to mathematical symbolics, which can deal with all types of numbers and operations. Future research could explore and investigate these two approaches. It is noteworthy that the results of this EDR are only relevant to TMs employing available off-the-shelf technology. Therefore, future research on TMs using made-to-order technologies might discover different findings.

REFERENCES

Alfaro, H., & Joutsenlahti, J. (2020). Promoting learning with understanding: Introducing languaging exercises in calculus course for engineering students at the university level. *International Journal on Math, Science and Technology Education, 8*(1), 229–251. https://doi.org/10.31129/LUMAT.8.1.1412

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences, 21*(2), 247–286. https://doi.org/10.1080/10508406.2011.611446

All European Academies. (2017). *European code of conduct for research integrity* (Revised ed.). https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

Alshabeb, A. (2020). *The potential of using mobile social media applications for language learning: A case study in Saudi higher education* [Doctoral dissertation, University of Wolverhampton]. WIRE: Wolverhampton Intellectual Repository and E-Theses.

Alshwaikh, J. & Morgan, C. (2013). Analysing the Palestinian school mathematics textbooks: A multimodal (multisemiotic) perspective. In C. Smith (Ed.), *Proceedings of the British Society for Research Into Learning Mathematics, 33*(2), 70–75.

Andamon, J. C., & Tan, D. A. (2018). Conceptual understanding, attitude and performance in mathematics of grade 7 students. *International Journal of Scientific & Technology Research, 7*(8), 96–105.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Australian Curriculum, Assessment and Reporting Authority. (2016). *Australian F–10 curriculum: Mathematics.* https://www.australiancurriculum.edu.au/umbraco/Surface/Download/Pdf?subject=Mathematics&type=F10

Ball, D. L. (1992). Magical hopes: Manipulatives and the reform of math education. *American Educator, 16*(1), 14–18, 46–47.

Barab, S. (2006). Design-based research. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 153–169). Cambridge University Press.

Barmby, P., Harries, T., & Higgins, S. (2010). Teaching for understanding/understanding for teaching. In I. Thompson (Ed.), *Issues in teaching numeracy in primary schools* (2nd ed., pp. 45–57). Open University Press.

Barmby, P., Harries, T., Higgins, S., & Suggate, J. (2007). How can we assess mathematical understanding? In J. H. Woo, H. C. Lew, K. S. Park, & D. Y. Seo (Eds.), *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education: Vol. 2 Research reports* (pp. 41–48). Korea Society of Educational Studies in Mathematics.

Baroody, A. J. (1989). Manipulatives don't come with guarantees. *Arithmetic Teacher, 37(*2), 4–5. https://doi.org/10.5951/AT.37.2.0004

Bauersfeld, H. (1995). Language games in the mathematics classroom: Their function and their effects. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 271–294). Lawrence Erlbaum.
Beaudouin-Lafon, M., & Mackay, W.E. (2012). Prototyping tools and techniques. In J. A. Jacko (Ed.), *Human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (3rd ed., pp. 1081–1104). CRC Press.
Bedir, G., & Özbek, Ö. Y. (2016). Primary school teachers' views on the preparation and usage of authentic material. *Higher Education Studies, 6*(3), 138–146. http://dx.doi.org/10.5539/hes.v6n3p138
Borenson & Associates. (n.d.). *Hands-On Equations* [Equipment]. https://www.borenson.com/Products/Teacher-Materials-Hands-On-Equations
Borenson & Associates. (2015). *Hands-On Equations 1* (Version 3.7) [Mobile app]. App Store. https://apps.apple.com/fi/app/hands-on-equations-1/id505948222?l=en
Bogomolny, M. (2007). Raising students' understanding: Linear algebra. In J. H. Woo, H. C. Lew, K. S. Park, & D. Y. Seo (Eds.), *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education: Vol. 2 Research reports* (pp. 65–72). Korea Society of Educational Studies in Mathematics.
Booth, J. L., & Koedinger, K. R. (2008). Key misconceptions in algebraic problem solving. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 571–576). Cognitive Science Society.
Brainingcamp, LLC. (2020). *Algebra tiles* (Version 4.1.0) [Mobile app]. App Store. https://apps.apple.com/fi/app/algebra-tiles/id568896224
Braukmüller, M., Bikner-Ahsbahs, A., & Wenderoth, D. F. (2019). Towards linking teaching, technology and textbooks. In U. T. Jankvist, M. van den Heuvel-Panhuizen & M. Veldhuis (Eds.), *Proceedings of the 11th Congress of the European Society for Research in Mathematics Education* (pp. 2671–2678). Freudenthal Group & Freudenthal Institute, Utrecht University, and ERME.
Bray, A., & Tangney, B. (2017). Technology usage in mathematics education research—A systematic review of recent trends. *Computers and Education, 114,* 255–273. https://doi.org/10.1016/j.compedu.2017.07.004
Brenner, M. E., Herman, S., Ho, H.-Z., & Zimmer, J. M. (1999). Cross-national comparison of representational competence. *Journal for Research in Mathematics Education, 30*(5), 541–557. https://doi.org/10.2307/749773
Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2
Bruner, J. S. (1961). The act of discovery. *Harvard Education Review, 31*, 21–32.
Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.
Bryman, A. (2014). June 1989 and beyond: Julia Brannen's contribution to mixed methods research. *International Journal of Social Research Methodology, 17*(2), 121–131. https://doi.org/10.1080/13645579.2014.892653
Buchbinder, O., Chazan, D., & Fleming, E. (2015). Insights into the school mathematics tradition from solving linear equations. *For the Learning of Mathematics, 35*(2), 2–8.
Capraro, M. M., & Joffrion, H. (2006). Algebraic equations: Can middle-school students meaningfully translate from words to mathematical symbols? *Reading Psychology, 27*(2–3), 147–164. https://doi.org/10.1080/02702710600642467

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400. https://doi.org/10.1037/a0031084

Ceuppens, S., Deprez, J., Dehaene, W., & De Cock, M. (2018). Design and validation of a test for representational fluency of 9th grade students in physics and mathematics: The case of linear functions. *Physical Review Physics Education Research, 14*(2), 020105:1–020105:19. https://doi.org/10.1103/PhysRevPhysEducRes.14.020105

Clements, D. H. (1999). 'Concrete' manipulatives, concrete ideas. *Contemporary Issues in Early Childhood, 1*(1), 45–60. https://doi.org/10.2304/ciec.2000.1.1.7

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13. https://doi.org/10.3102/0013189X032001009

Collins, A. (1992). Towards a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Springer.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Science, 13*(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2

Cowling, M., & Birt, J. (2018). Pedagogy before technology: A design-based research approach to enhancing skills development in paramedic science using mixed reality. *Information, 9*(2), Article 29. https://doi.org/10.3390/info9020029

Creamer, G. E. (2018). *An introduction to fully integrated mixed methods research.* SAGE Publications. https://www.doi.org/10.4135/9781071802823

Creswell, J. W., & Creswell J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

de Lima, R. N., & Tall, D. (2008). Procedural embodiment and magic in linear equations. *Educational Studies in Mathematics, 67*(1), 3–18. https://doi.org/10.1007/s10649-007-9086-0

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational Inquiry. *Educational Researcher, 32*(1), 5–8. https://doi.org/10.3102/0013189X032001005

Di Biase, R. (2020). Using design-based research to explore the influence of context in promoting pedagogical reform. *Educational Design Research, 4*(2), Article 21. https://doi.org/10.15460/eder.4.2.1427

Domino, J. (2010). *The effects of physical manipulatives on achievement in mathematics in Grades K–6: A meta-analysis* [Doctoral dissertation, State University of New York at Buffalo]. ProQuest Dissertations Publishing.

Driver, M. K., & Powell, S. R. (2015). Symbolic and nonsymbolic equivalence tasks: The influence of symbols on students with mathematics difficulty. *Learning Disabilities Research and Practice, 30*(3), 127–134. https://doi.org/10.1111/ldrp.12059

Duval, R. (2006). A cognitive analysis of problems of comprehension in the learning of mathematics. *Educational Studies in Mathematics, 61*, 103–131. https://doi.org/10.1007/s10649-006-0400-z

Easterday, M. W., Lewis, D. G. R., & Gerber, E. M. (2017). The logic of design research. *Learning: Research and Practice, 4*(2), 131–160. https://doi.org/10.1080/23735082.2017.1286367

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences, 11*(1), 105–121. https://doi.org/10.1207/S15327809JLS1101_4
Edelson, D. C. (2006). Balancing innovation and risk: Assessing design research proposals. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 100–106). Routledge.
Figueira-Sampaio, A. S., dos Santos, E. E. F., & Carrijo, G. A. (2009). A constructivist computational tool to assist in learning primary school mathematical equations. *Computers & Education*, *53*(2), 484–492. https://doi.org/10.1016/j.compedu.2009.03.012
Finlex. (2021, March 8). *Decree on Qualification Requirements for Teaching Staff 986/1998*. Finland's Ministry of Justice. https://finlex.fi/en/laki/kaannokset/1998/en19980986
Finnish National Agency for Education. (2016). *National core curriculum for basic education 2014* (Publication 2016:5). https://www.ellibslibrary.com/book/9789521362590
Finnish National Board on Research Integrity. (2009). *Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review*. https://tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf
Finnish National Board on Research Integrity. (2012). *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*. https://tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf
Flyvbjerg, B. (2011). Case study. In N. K. Denzin & Y. S. Lincoln (Eds.), *Sage handbook of qualitative research* (4th ed., pp. 301–316). SAGE Publications.
Foster, D. (2007). Making meaning in algebra: Examining students' understandings and misconceptions. In A. H. Schoenfeld (Ed.), *Assessing mathematical proficiency* (Vol. 53, pp. 163–176). Cambridge University Press. https://doi.org/10.1017/CBO9780511755378
Fox, A. (2018). *Becoming an instructional leader for elementary mathematics: Transforming principal learning through a research-practice partnership* [Doctoral dissertation, University of Washington]. ProQuest Dissertations Publishing.
Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of 'concreteness fading' for children's mathematics understanding. *Learning and Instruction, 35*, 104–120. https://doi.org/10.1016/j.learninstruc.2014.10.004
Fàbregues, S., & Molina-Azorín, J. F. (2016). Addressing quality in mixed methods research: A review and recommendations for a future agenda. *Quality & Quantity, 51*(6), 2847–2863. https://doi.org/10.1007/s11135-016-0449-4
Goff, W. M., & Getenet, S. (2017). Design based research in doctoral studies: Adding a new dimension to doctoral research. *International Journal of Doctoral Studies, 12*, 107–121. https://doi.org/10.28945/3761
Goldin, G. A. (2002). Representation in mathematical learning and problem solving. In L. English (Ed.), *Handbook of international research in mathematics education* (pp.197–218). Erlbaum. https://doi.org/10.4324/9781410602541
Goldin, G. A. (2014). Mathematical representations. In S. Lerman (Ed.), *Encyclopedia of mathematics education*. Springer. https://doi.org/10.1007/978-94-007-4978-8_103
Groves, S. (2012). Developing mathematical proficiency. *Journal of Science and Mathematics Education in Southeast Asia, 35*(2), 119–145.

Hatfield, M. (1994). Use of manipulative devices: Elementary school cooperating teachers self-report. *School Science and Mathematics, 94*(6), 303–309. https://doi.org/10.1111/j.1949-8594.1994.tb15680.x

Herrington, J., McKenney, S., Reeves, T., & Oliver, R. (2007). Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. In C. Montgomerie & J. Seale (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007* (pp. 4089–4097). AACE.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 65–97). Information Age Publishing.

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

Hinton, P. R., McMurray, I., & Brownlow, C. (2014). *SPSS explained* (2nd ed.). Routledge.

Hoadley, C. M. (2004). Methodological alignment in design-based research. *Educational Psychologist, 39*(4), 203–212. https://doi.org/10.1207/s15326985ep3904_2

Ihantola, E.-M., & Kihn, L.-A. (2011). Threats to validity and reliability in mixed methods accounting research. *Qualitative Research in Accounting and Management, 8*(1), 39–58. https://doi.org/10.1108/11766091111124694

Ishii, H., & Ullmer, B. (2012). Tangible user interfaces. In J. A. Jacko (Ed.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (3rd ed., pp. 465–490). Taylor & Francis.

Jewitt, C., Bezemer, J. J., & O'Halloran, K. L. (2016). *Introducing multimodality*. Routledge.

Joutsenlahti, J., & Kulju, P. (2017). Multimodal languaging as a pedagogical model: A case study of the concept of division in school mathematics. *Education Sciences, 7*(1), Article 9. https://doi.org/10.3390/educsci7010009

Joutsenlahti, J., & Rättyä, K. (2015). Kielentämisen käsite ainedidaktisissa tutkimuksissa [The concept of languaging in subject-didactic studies]. In M. Kauppinen, M. Rautiainen & M. Tarnanen (Eds.), *Rajaton tulevaisuus: Kohti kokonaisvaltaista oppimista* [Unlimited Future: Towards Holistic Learning] (pp. 45–61). Finnish Research Association for Subject Didactics.

Joutsenlahti, J., & Vainionpää, J. (2010). Oppimateriaali matematiikan opetuksessa ja osaamisessa [Learning materials in mathematics teaching and competencies]. In E. K. Nieminen & J. Metsämuuronen (Eds.), *Miten matematiikan taidot kehittyvät? Matematiikan oppimistulokset peruskoulun viidennen vuosiluokan jälkeen vuonna 2008* [How do mathematical skills develop? Mathematics learning outcomes after the fifth grade of primary school in year 2008] (Education follow-up reports 2010:2, pp. 137–148). Finnish National Agency for Education.

Juuti, K., & Lavonen, J. (2006). Design-based research in science education: One step towards methodology. *Nordic Studies in Science Education, 4*, 54–68.

Kahoot DragonBox. (2019a). *DragonBox Algebra 5+* (Version 1.3.7) [Mobile app]. App Store. https://apps.apple.com/fi/app/dragonbox-algebra-5/id522069155

Kahoot DragonBox. (2019b). *DragonBox Algebra 12+* (Version 2.3.1) [Mobile app]. App Store. https://apps.apple.com/fi/app/dragonbox-algebra-12/id634444186

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives, 3(*3), 151–155. https://doi.org/10.1111/j.1750-8606.2009.00096.x

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *The Journal of the Learning Sciences, 13*(1), 115–128. https://doi.org/10.1207/s15327809jls1301_6

Kelly, A. E. (2006). Quality criteria for design research: Evidence and commitments. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 107–118). Routledge.

Kelly, A. E. (2013). When is design research appropriate? In T. Plomp & N. Nieveen (Eds.), *Educational design research, Part A: An introduction* (pp. 135–150). Institute for Curriculum Development. https://www.slo.nl/publish/pages/4474/educational-design-research-part-a.pdf

Kennedy-Clark, S. (2013). Research by design: Design-based research and the higher degree research student. *Journal of Learning Design, 6*(2), 26–32. https://doi.org/10.5204/jld.v6i2.128

Kerslake, L. S. (2019). *The playground of ideas: A design-based research investigation into dialogic thinking with six- and seven-year-old children in England* [Doctoral dissertation, University of Cambridge]. Apollo.

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

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312.

Laine, A., Ahtee, M., Näveri, L., Pehkonen, E., & Hannula, M. (2018). Teachers' influence on the quality of pupils' written explanations – Third-graders solving a simplified arithmagon task during a mathematics lesson. *International Journal on Math, Science and Technology Education, 6*(1), 87–104. https://doi.org/10.31129/LUMAT.6.1.255

Learning Resources. (2020). *Algebra Balance* [Equipment]. https://www.learningresources.com/algebra-balance

Lehtonen, D., & Joutsenlahti, J. (2015, May 27–28). *Manipulatives in primary school mathematics classroom: Needs, benefits, and limitations on equations teaching and learning* [Conference presentation]. The Fifth Nordic Conference on Subject Education (NoFa5), Helsinki, Finland.

Lehtonen, D., & Joutsenlahti, J. (2017a, May 29–31). *Promoting the use of manipulatives for mathematical concepts understanding in comprehensive school classrooms: Taking pedagogy and practicality into account* [Conference presentation]. The Sixth Nordic Conference on Subject Education (NoFa6), Odense, Denmark.

Lehtonen, D., & Joutsenlahti, J. (2017b, August 27–September 2). *The use of manipulatives in mathematics classrooms from the pedagogical and practical perspective* [Conference presentation]. The 17th Biennial European Association for Research on Learning and Instruction Conference (EARLI 2017), Tampere, Finland.

Lehtonen, D., Joutsenlahti, J., & Perkkilä, P. (2019, October 24–25). *The potentials of tangible technologies in primary school mathematics classrooms* [Conference presentation]. Annual Symposium of the Finnish Mathematics and Science Education Research Association, Helsinki, Finland.

Lepik, M., Grevholm, B., & Viholainen, A. (2015). Using textbooks in the mathematics classroom – the teachers' view. *Nordic Studies in Mathematics Education, 20*(3–4), 129–156.

Lesh, R., Post, T., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of representations in the teaching and learning of mathematics* (pp. 33–40). Lawrence Erlbaum.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Magruder, R. (2012). *Solving linear equations: A comparison of concrete and virtual manipulatives in middle school mathematics*. [Doctoral dissertation, University of Kentucky]. UKnowledge.

Magruder, R., & Mohr-Schroeder, M. (2013). Solving equations is all about balance: Using virtual manipulatives in the middle school classroom. In D. Polly (Ed.), *Common core mathematics standards and implementing digital technologies* (pp.201–214). IGI Global. http://doi:10.4018/978-1-4666-4086-3.ch013

Manches, A., & O'Malley, C. (2012). Tangibles for learning: A representational analysis of physical manipulation. *Personal and Ubiquitous Computing, 16*(4), 405–419. https://doi.org/10.1007/s00779-011-0406-0

Manches, A., O'Malley, C., & Benford, S. (2010). The role of physical representations in solving number problems: A comparison of young children's use of physical and virtual materials. *Computers and Education, 54*(3), 622–640. https://doi.org/10.1016/j.compedu.2009.09.023

Marichal, S., Rosales, A., Perilli, F. G., Pires, A. C., Bakala, E., Sansone, G., & Blat, J. (2017). CETA: Designing mixed-reality tangible interaction to enhance mathematical learning. In M. Jones, M. Tscheligi, Y. Rogers & R. Murray-Smith (Eds.), *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Article 29). Association for Computing Machinery. https://doi.org/10.1145/3098279.3098536

Marley, S. C., & Carbonneau, K. J. (2014). Theoretical perspectives and empirical evidence relevant to classroom instruction with manipulatives. *Educational Psychology Review, 26*(1), 1–7. https://doi.org/10.1007/s10648-014-9257-3

Marshall, L., & Swan, P. (2008). Exploring the use of mathematics manipulative materials: Is it what we think it is? In J. Renner, J. Cross & L. McCormack (Eds.), *Proceedings of EDU-COM 2008 International Conference—Sustainability in higher education: Directions for change* (pp. 338–350). Edith Cowan University.

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29*(4), 587–625. https://doi.org/10.1207/s15516709cog0000_15

Martin, T. (2008). Physically distributed learning with virtual manipulatives for elementary mathematics. In D. H. Robinson & G. Schraw (Eds.), *Recent innovations in educational technology that facilitate student learning* (pp. 253–275). Information Age Publishing.

McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research* (2nd ed.). Routledge.

McNeil, N. M., Hornburg, C. B., Devlin, B. L., Carrazza, C., & McKeever, M. O. (2019). Consequences of individual differences in children's formal understanding of mathematical equivalence. *Child Development, 90*(3), 940–956. https://doi.org/10.1111/cdev.12948

McNeil, N. M., & Jarvin, L. (2007). When theories don't add up: Disentangling the manipulatives debate. *Theory into Practice, 46*(4), 309–316.

McNeil, N. M., & Uttal, D. H. (2009). Rethinking the use of concrete materials in learning: Perspectives from development and education. *Child Development Perspectives, 3*(3), 137–139. https://doi.org/10.1111/j.1750-8606.2009.00093.x

Morgan, C., Planas, N., & Schütte, M. (2021). Developing a perspective on multiplicity in the study of language in mathematics classrooms. In N. Planas, C. Morgan & M. Schütte (Eds.), *Classroom research on mathematics and language: Seeing learners and teachers differently* (1st ed., pp. 3-21). Routledge. https://doi-org.libproxy.tuni.fi/10.4324/9780429260889

Moschkovich, J. N. (2015). Academic literacy in mathematics for English learners. *Journal of Mathematical Behavior, 40*, 43–62. https://doi.org/10.1016/j.jmathb.2015.01.005

Moyer-Packenham, P. S., & Bolyard, J. J. (2016). Revisiting the definition of a virtual manipulative. In P. S. Moyer-Packenham (Ed.), *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 3–23). Springer.

Moyer-Packenham, P. S., Roxburgh, A. L., Litster, K., & Kozlowski, J. S. (2021). Relationships between semiotic representational transformations and performance outcomes in digital math games. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-021-09506-5

Moyer-Packenham, P. S., Salkind, G. M., Bolyard, J. J., & Suh, J. M. (2013). Effective choices and practices: Knowledgeable and experienced teachers' uses of manipulatives to teach mathematics. *Online Journal of Education Research, 2*(2), 18–33.

Moyer-Packenham, P. S., & Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *International Journal of Virtual and Personal Learning Environments*, *4*(3), 35–50. https://doi.org/10.4018/jvple.2013070103

Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics, 47*(2), 175–197. https://doi.org/10.1023/A:1014596316942

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all.*

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf

National Library of Virtual Manipulatives. (n.d.). *Algebra Balance Scales* [Applet]. http://nlvm.usu.edu/en/nav/frames_asid_201_g_3_t_2.html?open=instructions&from=topic_t_2.html

Neber, H. (2012). Discovery learning. In N. Seel (Ed.), *Encyclopedia of the sciences of learning* (1st ed., pp. 1009–1012). Springer.

Organisation for Economic Cooperation and Development. (2017). *Education at a glance 2017: OECD indicators*. OECD Publishing. http://dx.doi.org/10.1787/eag-2017-en

Organisation for Economic Cooperation and Development. (2020). *Education policy outlook: Finland*. www.oecd.org/education/policy-outlook/country-profile-Finland-2020.pdf

O'Halloran, K. L. (2004). *Mathematical discourse: Language, symbolism and visual images.* Continuum.

O'Halloran, K. L. (2015a). The language of learning mathematics: A multimodal perspective. *Journal of Mathematical Behaviour, 40*, 63–74. https://doi.org/10.1016/j.jmathb.2014.09.002

O'Halloran, K. L. (2015b). Mathematics as multimodal semiosis. In E. Davis & P. J. Davis (Eds.), *Mathematics, substance and surmise* (1st ed., pp. 287–303). Springer. https://doi.org/ 10.1007/978-3-319-21473-3_14

Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools, 13*(1), 48–63.

Otten, M., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2019). The balance model for teaching linear equations: A systematic literature review. *International Journal of STEM Education, 6*, Article 30. https://doi.org/10.1186/s40594-019-0183-2

Pape, S. J., & Tchoshanov, M. A. (2001). The role of representation(s) in developing mathematical understanding. *Theory Into Practice, 40*(2), 118–127. https://doi.org/10.1207/s15430421tip4002_6

Phillips, D. C. (2006). Assessing the quality of design research proposals: Some philosophical perspectives. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 93–99). Routledge.

Piaget, J. (1965). *The child's conception of number*. W. W. Norton & Company.

Piaget, J. (1970). *Science of education and the psychology of the child*. Orion Press.

Pires, A. C., González Perilli, F., Bakala, E., Fleisher, B., Sansone, G., & Marichal, S. (2019). Building blocks of mathematical learning: Virtual and tangible manipulatives lead to different strategies in number composition. *Frontiers in Education, 4*, Article 81. https://doi.org/10.3389/feduc.2019.00081

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research, Part A: An introduction* (pp. 11–50). Institute for Curriculum Development. https://www.slo.nl/publish/pages/4474/educational-design-research-part-a.pdf

Pontual Falcão, T., Meira, L., & Gomes, A. S. (2007). Designing tangible interfaces for mathematics learning in elementary school. In *Proceedings of the Third Latin American Conference on Human-Computer Interaction*. Association for Computing Machinery.

Poon, K., & Leung, C. (2010). Pilot study on algebra learning among junior secondary students. *International Journal of Mathematical Education in Science and Technology, 41*(1), 49–62. https://doi.org/10.1080/00207390903236434

Pouw, W. T. J. L., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review, 26*(1), 51–72. https://doi.org/10.1007/s10648-014-9255-5

Price, S. (2013). Tangibles: Technologies and interaction for learning. In S. Price, C. Jewitt & B. Brown (Eds.), *The Sage handbook of digital technology research* (pp. 307–325). SAGE Publications. http://dx.doi.org/10.4135/9781446282229

Reeves, T. C. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 52–66). Routledge.

Reinschlüssel, A. V., Alexandrovsky, D., Döring, T., Kraft, A., Braukmüller, M., Janßen, T., Reid, D., Vallejo, E., Bikner-Ahsbahs, A., & Malaka, R. (2018). Multimodal algebra learning: From math manipulatives to tangible user interfaces. *I-Com, 17*(3), 201–209. https://doi.org/10.1515/icom-2018-0027

Reinschlüssel, A. V., Döring, T., Alexandrovsky, D., Thieme, D., & Malaka, R. (2018). Beyond math manipulatives: Smart tangible objects for algebra learning. In F. Müller, D. Schnelle-Walka, S. Günther & M. Funk (Eds.), *Proceedings of the 6th Workshop on Interacting with Smart Objects* (Vol. 2082, pp. 9–16). CEUR-WS.

Resnick, M., Martin, F., Berg, R., Borovoy, R., Colella, V., Kramer, K., & Silverman, B. (1998). Digital manipulatives: New toys to think with. In C.-M. Karat, A. Lund, J. Coutaz & J. Karat (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 281-287). Association for Computing Machinery. https://doi.org/10.1145/274644.274684

Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. L. (2011). Assessing knowledge of mathematical equivalence. *Journal of Educational Psychology, 103*(1), 85–104. https://doi.org/10.1037/a0021334.

Ryan, B., Scapens, R. W., & Theobald, M. (2002). *Research method and methodology in finance and accounting* (2nd ed.). Thomson.

Salvador, G., Pérez, D., Ortega, M., Soto, E., Alcañiz, M., Contero, M. (2012). Evaluation of an augmented reality enhanced tabletop system as a collaborative learning tool: A case study on mathematics at the primary school. In G. Gallo & B. S. Santos (Eds.), *Proceedings of the 33rd Annual Conference of the European Association for Computer Graphics* (pp. 9–16). Eurographics Association. http://dx.doi.org/10.2312/conf/EG2012/education/009-016

Sandoval, W. A., & Bell, P. (2004). Design-based research methods for studying learning in context: Introduction. *Educational Psychologist, 39*(4), 199–201. https://doi.org/10.1207/s15326985ep3904_1

Sapounidis, T., & Demetriadis, S. (2013). Tangible versus graphical user interfaces for robot programming: Exploring cross-age children's preferences. *Personal and Ubiquitous Computing, 17*(8), 1775–1786. https://doi.org/10.1007/s00779-013-0641-7

Schleppegrell, M. J. (2010). Language in mathematics teaching and learning: A research review. In J. N. Moschkovich (Ed.), *Language and mathematics education: Multiple perspectives and directions for research* (pp. 73–112). Information Age Publishing.

Schoenfeld, A. H. (2007). What is mathematical proficiency and how can it be assessed? In A. H. Schoenfeld (Ed.), *Assessing mathematical proficiency* (Vol. 53, pp. 59–73). Cambridge University Press. https://doi.org/10.1017/CBO9780511755378

Sherman, J., & Bisanz, J. (2009). Equivalence in symbolic and nonsymbolic contexts: Benefits of solving problems with manipulatives. *Journal of Educational Psychology, 101*(1), 88–100. https://doi.org/10.1037/a0013156

Sierpinska, A. (1994). *Understanding in mathematics*. Routledge.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*(1), 20–26.

Slavin, R. E. (2010). Cooperative learning. In P. Peterson, E. Baker & B. McGaw (Eds.), *International encyclopaedia of education* (3rd ed., pp. 177–183). Elsevier.

Sowell., E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20*(5), 498–505. https://doi.org/10.2307/749423

Starcic, A. I., Cotic, M., & Zajc, M. (2013). Design-based research on the use of a tangible user interface for geometry teaching in an inclusive classroom. *British Journal of Educational Technology, 44*(5), 729–744. https://doi.org/10.1111/j.1467-8535.2012.01341.x

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.

Suh, J., & Moyer, P. S. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching, 6*(2), 155–173.

Swain, M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp.95–108). Continuum.

Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social & behavioral research*. SAGE Publications.

Tashakkori, A., & Teddlie, C. (2008). Quality of inferences in mixed methods research: Calling for an integrative framework. In M. M. Bergman (Ed.), *Advances in mixed methods research* (pp. 101–119). SAGE Publications.

Teddlie, C., & Tashakkori, A. (2012). Common 'core' characteristics of mixed methods research: A review of critical issues and call for greater convergence. *American Behavioral Scientist, 56*(6), 774–788. https://doi.org/10.1177/0002764211433795

Toptaş, V., Çelik, S., E., & Karaca, T. (2012). Pedagogical materials use of primary grade teachers in mathematics education. *Elementary Education Online, 11*(4), 1121–1130. http://ilkogretim-online.org.tr/index.php/io/article/view/1470/1326

Tran, C., Smith, B., & Buschkuehl, M. (2017). Support of mathematical thinking through embodied cognition: Nondigital and digital approaches. *Cognitive Research: Principles and Implications, 2*(1), Article 16. https://doi.org/10.1186/s41235-017-0053-8

Tripathi, P. N. (2008). Developing mathematical understanding through multiple representations. *Mathematics Teaching in the Middle School, 13*(8), 438–445. http://www.jstor.org/stable/41182592

Tucker, S. I., & Johnson, T. N. (2017). I thought this was a study on math games: Attribute modification in children's interactions with mathematics apps. *Education Sciences, 7*(2), Article 50. https://doi.org/10.3390/educsci7020050

Tuohilampi, L., & Hannula, M. S. (2013). Matematiikkaan liittyvien asenteiden kehitys sekä asenteiden ja osaamisen välinen vuorovaikutus 3., 6. ja 9. luokalla [The development of attitudes towards mathematics and the interplay between attitudes and competence during third, sixth, and ninth grades]. In J. Metsämuuronen (Ed.), *Perusopetuksen matematiikan oppimistulosten pitkittäisarviointi vuosina 2005−2012* [Longitudinal assessment of mathematics learning outcomes in basic education during 2005−2012] (Education follow-up reports 2013:4, pp. 231–253). Finnish National Agency for Education.

Ulrich, K. T., & Eppinger, S. D. (2016). *Product design and development* (6th ed.). McGraw-Hill Education.

Uribe-Flórez, L. J., & Wilkins, J. L. M. (2010). Elementary school teachers' manipulative use. *School Science and Mathematics, 110*(7), 363–371. https://doi.org/10.1111/j.1949-8594.2010.00046.x

Uribe-Flórez, L. J., & Wilkins, J. L. M. (2017). Manipulative use and elementary school students' mathematics learning. *International Journal of Science and Mathematics Education, 15*(8), 1541–1557. https://doi.org/10.1007/s10763-016-9757-3

Uttal, D. H., Amaya, M., del Rosario Maita, M., Hand, L. L., Cohen, C. A., O'Doherty, K., & DeLoache, J. S. (2013). It works both ways: Transfer difficulties between manipulatives and written subtraction solutions. *Child Development Research, 2013*, Article 216367. https://doi.org/10.1155/2013/216367

Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. J., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives, 3*(3), 156–159. https://doi.org/10.1111/j.1750-8606.2009.00097.x

Vessonen, T., Väisänen, E., Laine, A., & Aunio, P. (2020). Virtuaalisten ja konkreettisten apuvälineiden käytön vaikutus viidesluokkalaisten murtolukujen oppimiseen [The impact of the use of virtual and concrete manipulatives on fifth graders' fraction learning]. *NMI-bulletin, 30*(4), 72–82.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Warren, E., & Cooper, T. J. (2005). Young children's ability to use the balance strategy to solve for unknowns. *Mathematics Education Research Journal, 17*(1), 58–72. https://doi.org/10.1007/BF03217409
Warren, E., & Cooper, T. J. (2009). Developing mathematics understanding and abstraction: The case of equivalence in the elementary years. *Mathematics Education Research Journal, 21*(2), 76–95. https://doi.org/10.1007/BF03217546
Warren, E., Trigueros, M., & Ursini, S. (2016). Research on the learning and teaching of algebra. In Á. Gutiérrez, G. C. Leder & P. Boero (Eds.), *Second handbook of research on the psychology of mathematics education: The journey continues* (pp. 73–108). Sense Publishers.
Watson, A. (2009). Algebraic reasoning. In T. Nunes, P. Bryant & A. Watson (Eds.), *Key understandings in mathematics learning* (Paper 6). Nuffield Foundation. https://www.nuffieldfoundation.org/project/key-understandings-in-mathematics-learning
Watson, D. M. (2001). Pedagogy before technology: Re-thinking the relationship between ICT and teaching. *Education and Information Technologies, 6*(4), 251–266. https://doi.org/10.1023/A:1012976702296
Zamorano Urrutia, F., Cortés Loyola, C., & Herrera Marín, M. (2019). A tangible user interface to facilitate learning of trigonometry. *International Journal of Emerging Technologies in Learning, 14*(23), 152–164. http://dx.doi.org/10.3991/ijet.v14i23.11433
Zbiek, R. M., Heid, M. K., Blume, G. W., & Dick, T. P. (2007). Research on technology in mathematics education. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1169–1207). Information Age Publishing.
Zhou, Y., Wang, M., 2015. Tangible user interfaces in learning and education. In: J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., Vol. 24., pp. 20–25). Elsevier.
Zito, L., Cross, J. L., Brewer, B., Speer, S., Tasota, M., Hamner, E., Johnson, M., Lauwers, T., & Nourbakhsh, I. (2021). Leveraging tangible interfaces in primary school math: Pilot testing of the Owlet math program. *International Journal of Child-Computer Interaction, 27*, Article 100222. https://doi.org/10.1016/j.ijcci.2020.100222
Ørngreen, R. (2015). Reflections on design-based research: In online educational and competence development projects. In J. Abdelnour-Nocera, B. R. Baricelli, A. Lopes, P. Campos & T. Clemmensen (Eds.), *Human work interaction design: Analysis and interaction design methods for pervasive and smart workplaces* (pp. 20–38). Springer.

PUBLICATIONS

PUBLICATION
I

A Systematic Review of Educational Design Research in Finnish Doctoral Dissertations on Mathematics, Science, and Technology Education

Daranee Lehtonen, Anne Jyrkiäinen, & Jorma Joutsenlahti

International Journal on Math, Science and Technology Education, 7(3), 140–165
https://doi.org/10.31129/LUMAT.7.3.399

A systematic review of educational design research in Finnish doctoral dissertations on mathematics, science, and technology education

Daranee Lehtonen, Anne Jyrkiäinen and Jorma Joutsenlahti

Faculty of Education and Culture, Tampere University, Finland

Since educational design research (EDR) was introduced to educational research at the beginning of the 1990s, it has gained recognition as a promising research approach that bridges the gap between research and practice in education. This paper aims to investigate how EDR has been utilised and developed and which challenges it has faced by systematically reviewing 21 Finnish EDR doctoral dissertations on mathematics, science, and technology education published between January 2000 and October 2018. The findings indicate that all dissertations yielded practical and theoretical contributions. Moreover, common EDR characteristics, including the use of educational problems in practice as a point of departure, research in real-world settings, evolution through an iterative process, development of practical interventions, and refinement of theoretical knowledge, were found in all dissertations. Most of the doctoral researchers were confronted with challenges, such as high demand for EDR with limited resources and difficulties associated with multidisciplinary teamwork. However, the dissertations were diverse in terms of research contexts, practical educational problems, research outcomes, research methodologies, scale, and collaboration. This systematic review not only enhances the understanding of the utilisation, development, and challenges of EDR but also provides implications for future EDR.

Keywords
educational design research, doctoral dissertation, mathematics education, science education, technology education

Correspondence
daranee.lehtonen@tuni.fi

DOI
https://doi.org/10.31129/LUMAT.7.3.399

1 Introduction

Since educational design research (EDR) was introduced to educational research at the beginning of the 1990s, it has gained recognition as a promising research approach that bridges the gap between theoretical research and practice in education. Globally, EDR is still developing (Easterday, Lewis, & Gerber, 2017), as it is relatively young compared to other research approaches in education (Bell, 2004; Ørngreen, 2015). Over the past three decades, researchers have conducted EDR from a variety of theoretical perspectives and traditions for various purposes and contexts using different research methods (Bell, 2004; Prediger, Gravemeijer, & Confrey, 2015). While they have provided evidence supporting the usefulness of EDR, some have critiqued its limitations and challenges.

To better understand how EDR has been utilised and developed and which challenges it has faced, we systematically reviewed EDR studies conducted in the context of mathematics, science, and technology education at all levels. This lens was chosen for two reasons. First, it is likely that EDR is conducted differently in different educational fields, and therefore examining its application in specific fields may help refine the understanding of how to carry out EDR (McKenney & Reeves, 2012). Second, EDR has been adopted in a growing body of research on mathematics, science, and technology in education (Anderson & Shattuck, 2012; Prediger et al., 2015; Zheng, 2015).

2 Educational design research (EDR)

2.1 Overview of EDR

In this paper, we use the term *educational design research* to describe a research approach that is also known as *design experiments*, *design research*, *design-based research*, and *development (al) research*. EDR uses educational problems in practice as a point of departure and seeks to develop practical solutions to improve educational practices and advance usable knowledge through iterative processes in real-world settings (McKenney & Reeves, 2019; Plomp, 2013).

The manifold studies on EDR differ in terms of goals, forms, processes, outcomes, and other aspects (e.g., Bell, 2004; Plomp, 2013; Prediger et al., 2015). In addition, scholars have defined EDR in a variety of ways. Table 1 provides examples of EDR characteristics proposed by Anderson and Shattuck (2012); Cobb, Confrey, diSessa, Lehrer, and Schauble (2003); Juuti and Lavonen (2006); McKenney and Reeves (2019); and Wang and Hannafin (2005). Nevertheless, there are some commonalities among the definitions: intervention in real-world settings to improve practices, evolution through iterative cycles, development of practical solutions (i.e., interventions), and refinement of theoretical knowledge.

Table 1. Variants of educational design research (EDR) characteristics proposed by different scholars.

Title	**Characteristics of EDR**	**Reference**
Design-based research	1. Situated in real educational contexts 2. Focusing on the design and testing of interventions 3. Utilising mixed methods 4. Involving multiple iterations 5. Entailing partnership between researchers and practitioners 6. Providing design principles 7. Different from action research 8. Having a practical impact on practice	Anderson and Shattuck, 2012, pp. 16–18
Crosscutting features of design experiments	1. Developing theories about the learning process and ways to facilitate that learning 2. Interventionist: bringing about educational innovation 3. Prospective and reflective 4. Iterative cycles of intervention and revision 5. Practice orientated	Cobb, Confrey, diSessa, Lehrer, and Schauble, 2003, pp. 9–11
Features of the design-based research	1. Iterative process 2. Developing usable artefacts 3. Rendering novel educational knowledge	Juuti and Lavonen, 2006, pp. 59–63
Features of the design research process	1. Theoretically oriented 2. Interventionist: developing solutions informed by existing knowledge, testing, and participants 3. Collaborative: working in collaboration with others 4. Responsively grounded process 5. An Iterative process of investigation, development, testing, and refinement	McKenney and Reeves, 2019, pp. 12–16
Characteristics of design-based research	1. Pragmatic: refining theory and practice 2. Grounded in relevant research, theory, and practice 3. Interactive: working together with participants; an iterative cycle of analysis, design, implementation, and redesign; and flexible when necessary 4. Integrative: using mixed research methods 5. Contextual research results and generated design principles	Wang and Hannafin, 2005, p. 8

Descriptions of phases of EDR differ between scholars (cf. Cobb et al., 2003; Easterday et al., 2017). According to Plomp (2013), there are three main phases: (1) preliminary research (i.e., literature research, needs and context analysis, and theoretical framework development), (2) the development phase (i.e., the iterative design phase), and (3) the assessment phase (i.e., the summative evaluation of the intervention and recommendations for improvement; cf. McKenney & Reeves, 2019, who described the initial phase, design phase, and evaluation). McKenney and Reeves (2019) divided EDR into cycles of different sizes: single subcycle, multiple subcycles, and overall design research project. A single subcycle is the completion of one of the three main phases (i.e., preliminary research, development, or assessment). Multiple subcycles consist of several subcycles, but not as many as the whole EDR project. An overall design research project can range from one multiple subcycle that consist of three subcycles of each phase to several multiple subcycles.

EDR contributes to both practice and theory. In terms of its practical contribution, EDR uses an iterative process of design, assessment, and redesign in authentic contexts to develop an intervention to solve an educational problem (McKenney & Reeves, 2019). Additionally, according to Edelson (2002), EDR can help to develop three types of theory: *domain theories*, *design frameworks*, and *design methodologies* (cf. Plomp, 2013). Domain theories describe real-world phenomena and the outcomes of design implementation; design frameworks describe the characteristics of successful solutions to the problem in the studied context; design methodologies provide guidelines for successfully achieving the research aims.

2.2 EDR challenges and recommendations

Scholars have addressed several challenges of EDR and provided recommendations for how to overcome them. First, the triangulation of data sources, data collection methods, data types, theories, and evaluators is recommended to better understand complex real-world phenomena and enhance the reliability and validity of EDR (e.g., Design-Based Research Collective [DBRC], 2003; McKenney & Reeves, 2019). Nevertheless, triangulation and the iterative nature of EDR usually lead to *over methodologisation*—that is, the collection and analysis of excessive amounts of data—which sometimes many not lead to adequate results (e.g., Brown, 1992; Dede, 2004). Second, EDR researchers often take on multiple roles (e.g., researcher, designer, implementor, and evaluator of the intervention), which may lead to conflicts of interest (e.g., Plomp, 2013). Triangulation of researchers can enhance the objectivity of EDR (Plomp, 2013). Third, several EDR studies tend to be *under conceptualised*, as they lack a profound theoretical foundation and do not seek to provide theoretical contributions (e.g., Dede, 2004). Therefore, EDR should not only provide solutions to problems but also yield a variety of theories, particularly theories related to the design process (McKenney & Reeves, 2019). Fourth, a multidisciplinary collaboration among various experts from relevant fields is recommended for ensuring the feasible and successful development of solutions to complex educational problems (e.g., Wang & Hannafin, 2005). However, multidisciplinary teamwork requires, for example, a shared understanding among team members, strong group cohesion, and respect for others, and thus teamwork can be tiresome and contentious (McKenney & Reeves 2019). Fifth, the involvement of various participant groups that are relevant to the implementation of the intervention (e.g., teachers, students, and organisations) is advised to better understand complex authentic contexts and enhance respondent triangulation (McKenney & Reeves, 2019; Ørngreen, 2015). Sixth, rather than refining only one design idea, working with alternative designs and exploring solutions is recommended to ensure that the proposed intervention is the best solution to the problem (McKenney & Reeves, 2019; Ørngreen, 2015). Finally, Kelly (2013) proposed that, as EDR requires the investment of considerable resources, EDR should be employed only when truly needed, such as when facing a challenging educational problem with no satisfactory solution.

2.3 Previous reviews of EDR

Previous studies have investigated the utilisation and progress of EDR and other relevant issues with various focuses and review processes.

Anderson and Shattuck (2012) reviewed and defined the characteristics of EDR, including interventions in real educational contexts, a focus on the design and testing of a significant intervention, the use of mixed methods, multiple iterations, a collaborative partnership between the researcher and practitioners, the provision of design principles, differences from action research, and practical impact. The authors also conducted a review of the 47 most cited EDR articles from 2002 to 2011. Quantitative and qualitative content analyses were conducted to investigate geographic, disciplinary, and curricular focuses and the interventions, iterations, and outcomes of the articles. They found that design research was increasingly employed in educational contexts and that the majority of studies were conducted in North America. The most commonly studied subject was science; the main context was K–

12; and most interventions involved technology. Thirty-one articles were empirical studies that were part of a multi-iterative research project. All of the empirical studies involved were either technological and instructional design interventions or instructional methods, models, and strategies. Typically, mixed methods were employed. Most focused on furthering theoretical knowledge and developing applications to improve learners' learning outcomes or attitudes. Although the results of their review affirmed the great promise of EDR due to its integration of educational theory and practice, Anderson and Shattuck (2012) argued that work still needs to be done regarding educational innovations. Moreover, they recommended that future reviews perform a more detailed investigation of the full text of articles and investigate a broader set of articles. Their characterisation of EDR has been cited numerous times.

According to McKenney and Reeves (2013), most of the EDR characteristics defined by Anderson and Shattuck (2012) are similar to those reported by other authors. However, McKenney and Reeves identified that departure from a problem is an important characteristic of EDR that is missing from Anderson and Shattuck's (2012) list. Moreover, they criticized Anderson and Shattuck's (2012) systematic review for its limited search terms (*design-based research* and *education*), narrow dataset (i.e., only the most cited articles), and the use of only abstracts for a number of analyses (McKenney & Reeves, 2013). They called for the use of diverse search terms, an adequate dataset, and in-depth analyses of full texts to assess EDR progress in future studies (McKenney & Reeves, 2013).

Kennedy-Clark (2013) provided an overview of EDR as well as emphasised Plomp's (2007) three phases of EDR (i.e., initial, prototyping, and assessment phases) and the contribution of iterative cycles to the development of design principles and the refinement of theories. Furthermore, she investigated how EDR characteristics were used in doctoral dissertations by critically reviewing six education dissertations utilising EDR that were published by different institutions in Australia, Europe, Africa, and North America from January 2000 to January 2013. Her search terms included *design research*, *design-based research*, *education*, *phases*, *cycles*, and *iteration*. The research contexts (i.e., teaching subjects and education levels), focuses, and duration of data collection cycles varied among the dissertations, but they all utilised mixed methods for data collection. Conducting iterative data collection phases, engaging with several expert groups, testing designs with different participation groups, and being flexible and adaptive appeared to assist the researchers in reflecting on their research, understanding the educational problem, and avoiding overstated claims and conclusions. Finally, Kennedy-Clark's review demonstrated that the use of iterative design and development cycles or micro phases could increase the reliability and trustworthiness of research.

As researchers interested in EDR, we appreciate Kennedy-Clark's in-depth review of the potential benefits of EDR for education dissertations. However, the method was not sufficiently elaborated, and no overview of the information in the dissertations was provided. Revisiting the original article (Kennedy-Clark, 2013), Kennedy-Clark (2015) highlighted that researchers tend to concentrate on publishing their research findings and neglect to report their research methodologies. Therefore, there is a need for further investigation of how researchers employ EDR in their studies (Kennedy-Clark, 2015).

Zheng (2015) noted that applications of EDR do not appear to live up to expectations. She investigated empirical studies that adopted EDR through a systematic review of 162 journal articles published between 2004 and 2013 and quantitative content analysis of the selected EDR studies in terms of demographics, research methods, intervention characteristics, and research outcomes. The findings show that higher education was the most common sample group, and natural science was the most commonly studied learning domain. Qualitative methods were most often adopted, mixed methods were the second most popular, and solely quantitative methods were not used in any studies. Nearly all studies collected miscellaneous data, including interviews, questionnaires, and notes; and most performed technological interventions. More than half of the studies designed, developed, and redesigned educational interventions in only one iteration cycle. Although the majority revised their interventions, only approximately half of the studies reported how they did so. Moreover, most studies relied heavily on measurements of learners' cognitive outcomes. Based on her findings, Zheng (2015) proposed that there is a need for EDR studies to apply multiple iterations and new approaches that pay more attention to the design process.

We value her work for its thorough review of a large number of EDR studies and because it improves the understanding of the EDR landscape over the past decade. Nevertheless, a more detailed qualitative analysis would have complimented her quantitative analysis and contributed to an even deeper understanding of the selected studies. Zheng (2015) recognised the shortcomings of her research and recommended more deliberate investigation and analysis of design activities and their functions.

3 Methodology

3.1 Dissertation search and selection

To investigate how EDR has been employed in research on mathematics, science, and technology education and which challenges have confronted EDR researchers, we conducted a systematic review based on the recommendations of Anderson and Shattuck (2012), Kennedy-Clark (2015), McKenney and Reeves (2013), and Zheng (2015; see Section 2.3). Our data was collected from Finnish doctoral dissertations on mathematics, science, and technology education published between January 2000 and October 2018. We chose dissertations as our dataset because they report all iterative phases of the completed research, unlike articles, which often report only specific phases of research. We focused on Finnish dissertations because, as researchers in Finland, we expected our familiarity with the Finnish education system and practices to assist our review. It was not feasible to review all related dissertations completed at all Finnish universities because each university's repository uses a different database system, and there is no shared database containing all Finnish dissertations. Therefore, we decided to retrieve our data from the institutional repositories of the five Finnish universities that awarded the most qualifications and degrees in 2014: the University of Helsinki, University of Jyväskylä, University of Oulu, University of Tampere, and University of Turku (Official Statistics of Finland, 2015). The repository of the University of Eastern Finland, which provided the fourth most qualifications and degrees in 2014 and where a number of EDR dissertations have been completed, did not support the use of search terms for data retrieval. We also tried to retrieve dissertations of the University of Eastern Finland from Finna, a collection of search services providing access to material from Finnish university libraries. However, the Finna portal did not support a full-text search, which we used in our systematic review. Thus, we excluded the University of Eastern Finland and included the University of Tampere instead. Although our list of dissertations is not comprehensive, we believe that it provides an overview of the various dissertations published in Finland.

Our search terms included different terminologies that have been used to describe EDR in both English (*design research*, *design-based research/design based research*, *development research/developmental research,* and *design experiments*) and Finnish (*design-tutkimu*/suunnittelututkimu**, *design-perustai*/design-perustei*/suunnitteluperustai*/suunnitteluperustei**, *kehittämistutkimu*,* and *design-eksperiment**). The initial search resulted in 625 dissertations. One of the authors and a research assistant screened these results using the following inclusion criteria: (1) at least one of the search terms is visible in the English or Finnish title, abstract, or keywords and (2) the full text is openly available digitally. After applying these criteria, 55 dissertations remained. Each of the authors independently read one-third of this list according to our own interests and expertise. Thereafter, we jointly decided to exclude dissertations that did not utilise EDR as a strategy of inquiry, leaving 49 dissertations. At the beginning of this research, we decided not to use search terms similar to *mathematics* OR *science* OR *technology* AND *teach** OR *learn** OR *class** to locate all dissertations on mathematics, science, or technology education because doing so would not be possible. Instead, we carefully read the remaining EDR dissertations, identified which dissertations concerned mathematics, science, and technology education, and jointly excluded dissertations in fields other than mathematics, science, and technology education, such as other taught subjects (e.g., language, design, and nursing), skill and competence development, teaching and learning support, and learning environments in general.

3.2 Dataset

After the final screening process, the full texts of 21 EDR dissertations (10 in English and 11 in Finnish; 18 monographs and 3 article-based dissertations) on mathematics, science, and technology education from three universities (the University of Helsinki, University of Jyväskylä, and University of Oulu; *n* = 14, 6, and 1, respectively) remained for statistical and content analysis. Table 2 presents the number of EDR dissertations on mathematics, science, and technology education and on other educational domains by the university during the periods of 2000–2009 and 2010–2018.

Table 2. Frequency of EDR dissertations by the university and educational domain.

Educational Domains	University of Helsinki		University of Jyväskylä		University of Oulu		University of Tampere		University of Turku	
Year	2000 – 2009	2010 – 2018	2000 – 2009	2010 – 2018	2000 – 2009	2010 – 2018	2000 – 2009	2010 – 2018	2000 – 2009	2010 – 2018
Mathematics, science, and technology	4	10	3	3	0	1	0	0	0	0
Other	2	8	2	5	2	3	2	3	0	1
Total	6	18	5	8	2	4	2	3	0	1

Among the EDR dissertations on mathematics, science, and technology education, those of Aksela (2005) and Juuti (2005) were the first two published at the University of Helsinki, that of Leppäaho (2007) was the first at the University of Jyväskylä, and that of Oikarinen (2016) was the only one published at the University of Oulu. Altogether, there were 19 supervisors for the 21 dissertations. Aksela, who completed her EDR dissertation in 2005, supervised nine dissertations (43%), while Lavonen supervised six dissertations (29%).

3.3 Data analysis

After the final screening, each author coded one-third of the dissertations using a jointly constructed coding table. The coding categories were initially based on the previous literature, but we regularly discussed and modified existing categories and added relevant categories during the coding to best answer our research questions.

We coded the dissertations according to the following categories: (1) use of EDR terms and theoretical frameworks, (2) research contexts (i.e., educational sectors, settings, and domains), (3) educational problems in practice and research outcomes, (4) research methodology (i.e., research methods, data collection methods, and data sources), (5) scale, collaboration, and researcher's roles, (6) EDR process (i.e., phases of EDR, iterations, alternative design interventions, and issues during development of the intervention), and (7) EDR challenges. After the coding, we analysed the coded data quantitatively and qualitatively. Our findings are presented according to these seven categories in tables, figures, and descriptive analyses in the following section.

During the study, we strived to enhance the validity and reliability of our study by performing a precise research process, making joint decisions, crosschecking our data and analysis, consulting the literature for interpretations of the data, and comparing our research results to previous studies.

4 Results

4.1 Use of EDR terms and theoretical frameworks

EDR is referred to by a variety of names, and different scholars define it as having different goals, characteristics, and processes. Thus, investigating how EDR terms and theoretical frameworks have been used in dissertations published during the last two decades improves the understanding of how EDR is utilised and developed.

Of the four terms in each language used for our dissertation search, only three — design research, design-based/design based research, and development/developmental research in English and design-tutkimu*/suunnittelututkimu*, design-perustai*/design-perustei*/suunnitteluperustai*/suunnitteluperustei*, and kehittämistutkimu* in Finnish — appeared in the titles, abstracts, or keywords of the 21 dissertations. The dissertations did not apply a uniform format: while all of the dissertations included English versions of the title and abstract, only 18 included Finnish versions. We counted the appearance of each term only once per dissertation. Vartiainen (2016) used two terms in her English abstract, and Hassinen (2006) used two terms in her Finnish abstract. Thus, we also included them in our data (English: n = 22; Finnish: n = 19).

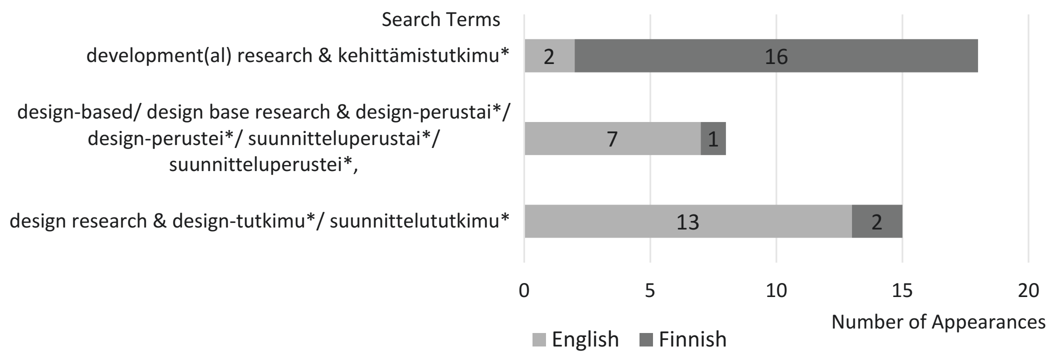


Figure 1. Frequency of search terms appearing in dissertation title, abstract, or keywords (English: n = 22; Finnish: n = 19).
Note: Only 18 dissertations had Finnish abstracts. One researcher used two terms in the English abstract, and another used two terms in the Finnish abstract.

Figure 1 shows the frequency of each search term in the English or Finnish titles, abstracts, or keywords of the dissertations. The most commonly used English term was ‘design research’, which appeared in 13 dissertations (69%), followed by ‘design-based/design based research’ in 7 dissertations (32%). The most commonly used Finnish term was ‘kehittämistutkimu*’ (development/developmental research), which appeared in 16 dissertations (84%). Interestingly, for 13 of the 18 dissertations (72%) that provided the title and abstract in both languages, the English and Finnish terms were not consistent. These dissertations used ‘kehittämistutkimu*’ (development/developmental research) in their Finnish titles or abstracts but either ‘design research’ or ‘design-based/design based research’ in their English titles or abstracts.

Our search terms appeared in the titles of 12 dissertations (57% of the 21 dissertations). Of these, six (50%) included the search terms in their primary titles, such as “**Design-Based Research** of a Meaningful Nonformal Chemistry Learning Environment in Cooperation with Specialists in the Industry” (Ikävalko, 2017) and “A **Design Research**: Problem and Inquiry Based Higher Education of Chemistry” (Rautiainen, 2012).

The comprehensiveness with which EDR theoretical frameworks were presented in the methodology sections of the dissertations varied from relatively superficial to exceedingly thorough. To investigate the use of these theoretical frameworks, we focused on the main EDR literature cited in the dissertations’ methodology sections, such as those regarding the principles, key characteristics, and processes of EDR. We found that early EDR works (e.g., Brown, 1992; Edelson, 2002; DBRC, 2003) and recent works (e.g., Anderson & Shattuck, 2012; McKenney & Reeves, 2019) were used as the main theoretical frameworks. The most cited article was that of Edelson (2002), which described the three types of theories (i.e., domain theories, design frameworks, and design methodologies) that can guide EDR. This article was cited in 18 dissertations (86%). The next most cited article was that of the DBRC (2003), which identified five characteristics of good design-based research and provided recommendations on how to increase the reliability and validity of EDR. This article was cited in 10 dissertations (48%). Of the Finnish EDR literature, Juuti and Lavonen’s (2006) article concerning the three pragmatic features of EDR was cited by nine dissertations (43%).

4.2 Research contexts

To obtain an overview of the authentic educational contexts in which the EDR dissertations were conducted, we examined their research contexts, including the educational sector (i.e., educational levels based on the Finnish educational system), setting (i.e., formal education vs. nonformal education), and domain (i.e., teaching and learning subjects).

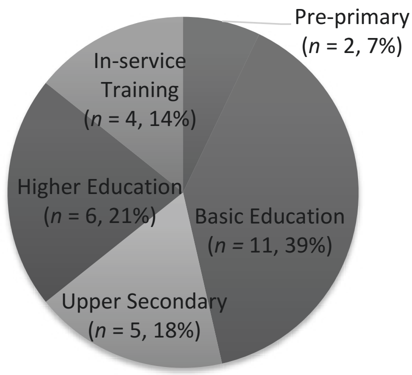


Figure 2. Frequency of educational sectors examined by the dissertations (*n* = 28)

Note: Five dissertations were carried out in more than one educational sector.

All of the dissertations were conducted in real-world educational contexts, and five were carried out in more than one educational sector. We included all of these sectors in our data (*n* = 28). Figure 2 shows a pie chart of the various educational sectors examined by the dissertations. Basic education (Grades 1–9; *n* = 11, 39%) was the most studied educational sector in the dissertations, while pre-primary school (*n* = 2; 7%) was the least.

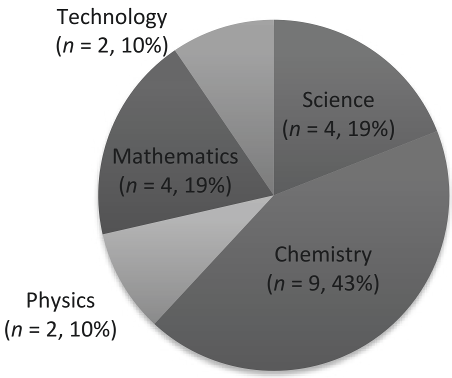


Figure 3. Frequency of the educational domains on which the dissertations focused (n = 21)

Note: The education domains were categorised based on the vocabulary used in the dissertations.

The majority of the 21 dissertations (n = 14, 67%) were conducted in a formal educational setting leading to formal qualifications, while the others were conducted in either a nonformal setting (n = 3, 14%) or in both types of settings (n = 4, 19%). The research interventions were conducted in various educational domains. Some researchers described these domains in a general way (e.g., science, mathematics, or technology in education), while others referred to specific subjects (e.g., chemistry and physics). We categorised our data accordingly. Moreover, we included upper secondary school statistics for mathematics, which is in line with the Finnish national core curriculum. Figure 3 illustrates that the most common domain was chemistry (n = 9, 43%), followed by science in general (n = 4, 19%) and mathematics (n = 4, 19%).

Table 3. Three dissertations serving as examples of variations in the research contexts of the dissertations

Research contexts	**Rukajärvi-Saarela (**2015**)**	**Ekonoja (**2014**)**	**Vartiainen (**2016**)**
Educational sector	Pre-service teacher education and in-service teacher training	Lower and upper secondary education	Pre-primary education (ages 3–6)
Educational setting	Formal and nonformal education	Formal education	Nonformal education
Educational domain	Primary school chemistry teaching	Teaching information and communication technology	Science club for small children

In sum, the dissertations were conducted in various research contexts (i.e., educational sectors, settings, and domains). Table 3 illustrates the differences in the research contexts using three dissertations as examples.

4.3 Educational problems in practice and research outcomes

All dissertations took at least one of four types of practical educational problems as a point of departure. Two dissertations took two types of problem as a point of departure; thus, we also included them in our data (*n* = 23). Figure 4 shows that the most common problem (*n* = 11, 48%) was students' lack of motivation and interest (e.g., Vartiainen, 2016), low performance (Hassinen, 2006), or deficient understanding (e.g., Oikarinen, 2016). The second most common problem (*n* = 7, 30%) was a lack of teaching and learning materials (e.g., Hongisto, 2012) or challenges in adapting to a new teaching and learning environment (e.g., Nieminen, 2008). The third type of problem (*n* = 3, 13%) was a teachers' deficient understanding and pedagogical skills (e.g., Juntunen, 2015). The last type (*n* = 2, 9%) concerned changes in a new curriculum (e.g., Kallunki, 2009).

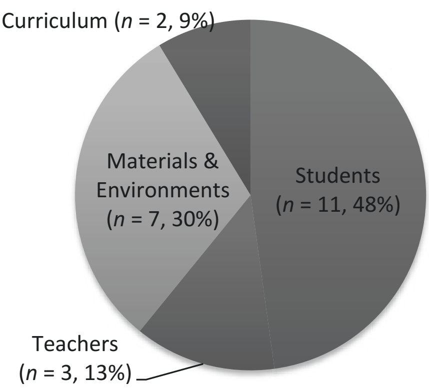


Figure 4. Types of practical educational problems that the dissertations took as points of departure (*n* = 23)

Note: Two dissertations took two problems as points of departure.

With regard to the practical contributions of the dissertations, various educational interventions were developed to respond to educational challenges in practice. Kallunki (2009) developed a teaching model and a learning environment, and we included both in our data (*n* = 22). The most common type of intervention involved teaching and learning environments (*n* = 10, 45%), such as a virtual science club

(Vartiainen, 2006) or a chemistry information and communication technology (ICT)-based learning environment (Pernaa, 2011). Another major type concerned teaching and learning concepts or models (n = 9, 41%), such as new chemistry teaching concepts for sustainability education (Juntunen, 2015) or a teaching model for algebra (Hassinen, 2006). Teaching and learning materials (n = 3, 14%), such as textbooks and electronic learning materials for teaching ICT (Ekonoja, 2014), were also developed.

We also investigated the theoretical contributions of the dissertations. Figure 5 shows that the majority of the dissertations (n = 15, 71%) developed all three types of theory (i.e., domain theories, design frameworks, and design methodologies) described by Edelson (2002). Nonetheless, only 11 of 15 developed all these theories thoroughly (e.g., Vartiainen, 2016). The remainder (n = 6, 29%) only developed domain theories and design frameworks (e.g., Tomperi, 2015) or domain theories and design methodologies (Leppäaho, 2007).

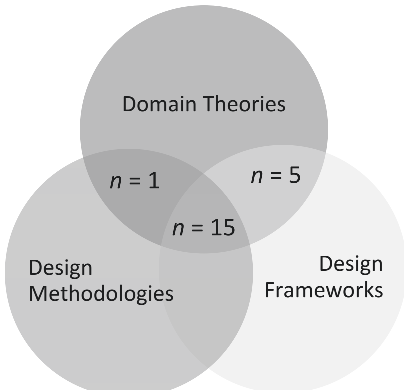


Figure 5. Venn diagram illustrating the theoretical contributions of the dissertations (n = 21)

4.4 Research methodology

The way in which EDR projects are conducted plays an important role in the success and reliability of those projects. Research triangulation is highly recommended to ensure the quality of EDR. Therefore, we examined how the triangulation of research methods, data collection methods, and data sources was implemented in the dissertations.

We coded the research methods as qualitative, quantitative, and mixed methods (see e.g., Creswell & Creswell, 2018). Fourteen dissertations (67%) gathered and analysed data with mixed methods (i.e., both qualitative and quantitative methods), while the remainder (*n* = 7, 33%) used only qualitative methods. None were conducted with only quantitative methods. Nevertheless, some of those dissertations that adopted mixed methods did not utilise qualitative and quantitative methods equally. For example, Ratinen's (2016) dissertation consisted of three substudies, only the first of which adopted mixed methods (i.e., a qualitative and quantitative questionnaire).

The dissertations used various methods to collect empirical data. The most common data collection methods were observation and questionnaires (each of which was used by 15 dissertations), followed by written documents, such as essays, diaries, and reports (which were used by 14 dissertations), and then interviews and group interviews (used by 13 dissertations). Some dissertations used tests and exams (e.g., Nieminen, 2008), tasks and exercises (e.g., Juntunen, 2015), and design intervention analysis (e.g., Pernaa, 2011). With regard to data sources, approximately half (11 of 21) of the dissertations collected data from both students and teachers, while the other half (*n* = 10) collected data from only students or only teachers. Additionally, several dissertations collected data from sources other than students and teachers; for example, Ikävalko (2017) collected data from company specialists, and Vartiainen (2016) collected data from parents.

In addition to investigating the dissertations' data collection methods and sources, we investigated how they collected data with multiple methods and from multiple sources to enhance their research triangulation. The number of data collection methods used in each dissertation ranged from one (Hongisto, 2012) to seven (Juuti, 2005), and the majority used three (*n* = 7, 33%) or four (*n* = 5, 24%). The number of data sources used in each dissertation varied from one (e.g., Rukajärvi-Saarela, 2015) to five (Tuomisto, 2018). Most of the researchers collected their data from one (*n* = 7, 33%) or two sources (*n* = 10, 48%).

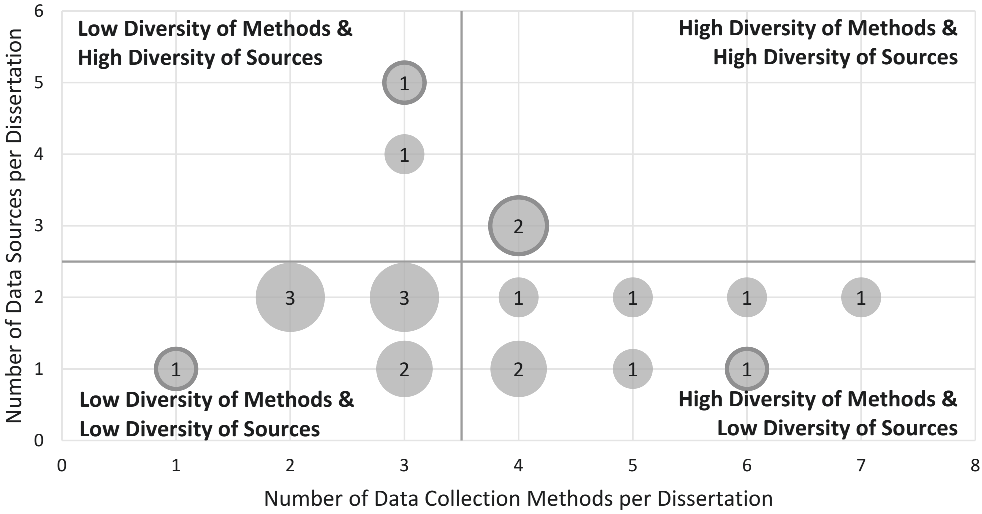


Figure 6. The positioning of dissertations in a research triangulation matrix with two dimensions: data collection methods (x-axis) and data sources (y-axis)

Note: Bubble size is based on the number of dissertations with the same coordinates. Dissertations from the far corner of each quadrant were highlighted.

We further analysed the research triangulation by using a matrix with two dimensions: the number of data collection methods used in each dissertation on the x-axis and the number of data sources used in each dissertation on the y-axis. As Figure 6 shows, the matrix is composed of four quadrants: (1) low diversity of methods and low diversity of sources (lower left quadrant), (2) high diversity of methods and low diversity of sources (lower right quadrant), (3) low diversity of methods and high diversity of sources (upper left quadrant), and (4) high diversity of methods and high diversity of sources (upper right quadrant). The majority of dissertations are located in the lower quadrants; nine dissertations (43%) had low diversity of methods and low diversity of sources, and eight (38%) had high diversity of methods and low diversity of sources. Only two dissertations (Loukomies, 2013; Vartiainen, 2016) had high diversity of methods and high diversity of sources. Table 4 provides four examples of dissertations from the far corner of each quadrant.

Table 4. Four dissertations that illustrate the variation in research methodologies in the dissertations

	Hongisto (2012)	**Ratinen** (2016)	**Tuomisto** (2018)	**Loukomies** (2013)
Collection Methods	Essay	Questionnaire, observations, group interviews, drawing, essays, and lesson plans	Questionnaire, observations, and diaries	Questionnaire, observations, interviews, and meeting memoranda
Data Sources	Students	Pre-service teachers	Teacher educators, pre-service and in-service teachers, students, and peers	Students, teachers, and experts

4.5 Scale, collaboration, and researcher's roles

The scale of the dissertations varied widely in terms of the size of the research team (from an individual researcher to a large multidisciplinary team), the number of research participants (from 15 to over 1000 participants), and the time taken to complete the dissertation (from 3 to 14 years). Eight researchers (38%) conducted their dissertations alone, while the remaining 13 (62%) collaborated with other researchers or disciplines. For example, Ratinen (2016) conducted his dissertation in collaboration with another researcher, and Nousiainen (2008) worked in a multidisciplinary team comprised of members from various fields, including educational sciences, natural sciences, mathematical information technology, game design (e.g., multimedia and graphic design), and stakeholders (e.g., industry representatives, biology and geography teachers, and students from several school levels).

Twenty researchers had one additional role besides that of a researcher. The majority (n = 13, 62%) of the researchers (e.g., Leppäaho, 2007) had three roles: a researcher who plans the research, collects data, and analyses data; a developer who designs and develops a design intervention; and a teacher who teaches in the research intervention. Seven researchers (33%), including Ekonoja (2014), had two roles: a researcher and a developer. Juntunen (2015) was the only one who had a single role: a researcher.

4.6 EDR process

To investigate the EDR processes used by the dissertations, we analysed the phases of EDR, iterations, alternative design interventions, and issues that were considered during the intervention development.

To analyse the EDR phases of the dissertations, we coded the progress of EDR according to three main phases: (1) preliminary research, (2) development phase, and (3) assessment phase (see Plomp, 2013). Although the EDR processes of the dissertations were presented in various ways using various terms (e.g., cases, cycles, phases, stages, and substudies), we found that all dissertations progressed through three main phases. However, the first phase (i.e., investigation of problems, needs, and context) was not fully conducted in several dissertations. For example, Hassinen (2006) did not empirically investigate needs or context and only reviewed the literature on school algebra, curricula, related theories, and textbooks; and Ekonoja's (2014) first phase was conducted as part of his master's thesis. Additionally, while the primary research and assessment phase was reported thoroughly in all dissertations, the development phase was rather brief in some examples (e.g., Oikarinen, 2016) and comprehensive in others (e.g., Juuti, 2005).

As an important characteristic of EDR is its iterative process of design, assessment, and redesign, we investigated the dissertations' iterations by examining revisions of the interventions and the number of multiple subcycles implemented throughout each dissertation (see McKenney & Reeves, 2019). Almost all researchers (n = 20, 95%) revised their interventions during their dissertations. Seven also refined their interventions after their final field trials. With regard to the number of multiple subcycles, 19 researchers (90%) revised their intervention through multiple subcycles. Thirteen (62%) employed two multiple subcycles, four (19%) employed three, one (5%) employed four, and one (5%) employed seven. In addition to performing seven multiple subcycles, Rukajärvi-Saarela (2015) refined her pre- and in-service teacher course after the final field trial. In contrast, two dissertations (10%) performed only one multiple subcycle. After the multiple subcycle, Hassinen (2006) did not revise her Idea-based Algebra teaching model, while Leppäaho (2007) developed his problem-solving materials further in a textbook.

To ensure that their interventions contributed to real-world settings, we also investigated whether any dissertations worked with alternative designs or considered issues besides pedagogy when developing the interventions. No one worked with alternative designs except Nousiainen (2008), whose first project included alternative user interfaces with layouts and different interaction styles and whose second project generated initial ideas and then integrated and developed them in greater detail.

With regard to the issues considered during intervention development, we found that besides pedagogical issues, most of the dissertations considered the needs of policymakers, particularly the National Core Curriculum, when developing interventions. Only a few dissertations considered other issues, such as practicality, usability, administration, and organisation. For example, when developing her ICT learning environment, Aksela (2005) considered pedagogy, the needs of policymakers, practicality (e.g., time, ease of use, resource availability, and classroom space), usability, and technical issues.

4.7 EDR challenges

Finally, we investigated which EDR challenges were encountered during the dissertations. The challenges in the dissertations can be classified into five categories, which are described below.

First, it was difficult to generalise the results due to the small number of research participants, the short length of interventions, the small number of iterative cycles, the insufficiency of relying only on qualitative data, or context-bound research results (e.g., Ekonoja, 2014; Kallunki, 2009). Second, the nature of EDR made it challenging to perform the research for the dissertations. For example, in Nousiainen's (2008) dissertation, it was difficult to compare the research results from different phases, and it was difficult for some participants to recall what happened at the beginning of a long intervention. In the case of Ekonoja (2014), the EDR interventions were typically innovative in nature, and thus there were no previous studies related to his research. Moreover, his intervention relied greatly on technology. Third, the researchers had limited resources in relation to the complexity of EDR, which requires a huge amount of work due to the need to gather and analyse a large dataset (Vartiainen, 2016) and explicitly document the whole process (Pernaa, 2011). Fourth, EDR was often conducted with multidisciplinary collaboration, which required mutual understandings and good teamwork (e.g., Ikävalko, 2017). Fifth, when they took on multiple roles, it was sometimes difficult for the researchers to maintain objectivity (e.g., Oikarinen, 2016).

5 Discussion and conclusions

Our study improves the understanding of how EDR has been utilised and developed and which challenges it has faced over the last two decades by systematically reviewing 21 Finnish doctoral dissertations on mathematics, science, and technology education. The findings indicate that all dissertations made practical and theoretical educational contributions. In line with the literature (e.g., DBRC, 2003; McKenney & Reeves, 2019; Plomp, 2013), all of the dissertations exhibited the characteristics of EDR, including the use of educational problems in practice as a point of departure, research in real-world settings, evolution through an iterative process (i.e., preliminary research, development, and assessment), development of practical interventions, and refining of theoretical knowledge. Moreover, the challenges faced by the researchers (e.g., high demand for conducting EDR with limited resources and the difficulties of multidisciplinary teamwork) are generally similar to those stated by other scholars (e.g., Brown, 1992; McKenney & Reeves, 2019). However, the dissertations were distinctly diverse in terms of the research context (i.e., educational sectors, settings, and domains), educational problems in practice, research outcomes, research methodology (i.e., research methods, data collection methods, and data sources), scale, and collaboration. Like the EDR reviews of Anderson and Shattuck (2012) and Zheng (2015), the findings support the plurality of EDR (see Bell, 2004). Our results indicate that it is feasible to conduct EDR dissertations in different educational sectors, in different settings and domains, at various scales, and with different research designs.

Based on our observations, we agree with other researchers (e.g., Easterday et al., 2017; Ørngreen, 2015; Zheng, 2015) that EDR still needs much more work. Thus, we propose several suggestions for future EDR. First, we encourage agreement between the terms used to describe EDR in different languages to promote consistency and avoid confusion. Second, as EDR is an emergent research approach (Easterday et al., 2017), recent literature should be consulted so that researchers can stay up to date. Third, in agreement with the DBRC (2003) and McKenney and Reeves (2019), we believe that the triangulation of research methods, data collection methods, and data sources is needed to better understand complex authentic phenomena and ensure the trustworthiness of EDR. Fourth, we support Kennedy-Clark (2013) and McKenney and Reeves (2019) and highly encourage multidisciplinary collaboration so that EDR researchers benefit from the expertise of others and increase the feasibility and robustness of their research. Fifth, in line with McKenney and Reeves (2019) and

Ørngreen (2015), when developing the intervention, working with alternative designs and considering various issues faced by all people in real-world contexts can enhance the success of EDR and ensure that the intervention continues to be utilised in real-world settings. Sixth, we agree with McKenney and Reeves (2019), Kennedy-Clark (2015), and Zheng (2015) that design activities and processes should be further emphasised so that others can benefit from them. Finally, due to the appearance of EDR terms in the primary titles of six dissertations, which implies that there is an overemphasis on EDR at the expense of the subject of the research, and the fact that EDR requires substantial resources (Kelly, 2013), we recommend that EDR should be undertaken because of its appropriateness and utility rather than for its own sake.

Our research has several limitations. First, our systematic review included only 21 Finnish dissertations on mathematics, science, and technology education from five universities. A broader dataset in terms of both the number of universities, dissertations, and educational fields would greatly improve the understanding of the utilisation and development of EDR. Second, the large dataset (a total of 4187 pages), the lack of a shared writing structure, and the implicit reporting of information that was necessary for this review made it difficult to perform data coding and analysis. More resources for coding and analysis would increase the precision of the research results and decrease the workload of researchers conducting the review. Last, to gain an overview of the utilisation, development, and challenges of EDR, we adopted a broad perspective when systematically reviewing the use of EDR terms and theoretical frameworks, research contexts, educational problems in practice and research outcomes, research methodologies, the dissertation's scale and collaboration, the researcher's roles, EDR processes, and EDR challenges. While our review indeed provides an overview, a review focusing on specific issues would yield profound insights into EDR.

References

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25. doi:10.3102/0013189X11428813

Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational Psychologist, 39*(4), 243–253.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141–178.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: Sage.

Dede, C. (2004). If design-based research is the answer, what is the question? A commentary on Collins, Joseph, and Bielaczyc; diSessa and Cobb; and Fishman, Marx, Blumenthal, Krajcik, and Soloway in the JLS special issue on design-based research. *Journal of the Learning Sciences, 13*(1), 105–114.

Design-Based Research Collective (DBRC). (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Easterday, M. W., Lewis, D. G. R., & Gerber, E. M. (2017). The logic of design research. *Learning: Research and Practice, 4*(2), 131–160. doi:10.1080/23735082.2017.1286367

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences, 11*(1), 105–121.

Juuti, K., & Lavonen, J. (2006). Design-based research in science education: One step towards methodology. *NorDiNa, 4*, 54–68.

Kelly, A. E. (2013). When is design research appropriate? In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (pp. 135–150). Enschede, Netherlands: Netherlands Institute for Curriculum Development.

Kennedy-Clark, S. (2013). Research by design: Design-based research and the higher degree research student. *Journal of Learning Design, 6*(2), 26–32.

Kennedy-Clark, S. (2015). Reflection: Research by design: Design-based research and the higher degree research student. *Journal of Learning Design, 8*(3), 106–107.

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research. London*: Routledge.

McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research*. New York: Routledge.

McKenney, S., & Reeves, T. C. (2013). Systematic review of design-based research progress: Is a little knowledge a dangerous thing? *Educational Researcher, 42*(2), 97–100. doi:10.3102/0013189X12463781

Official Statistics of Finland. (2015). University education. Appendix table 3. Students in universities and completed university degrees by universities in 2014. Retrieved from http://www.stat.fi/til/yop/2014/yop_2014_2015-05-06_tau_003_en.html

Plomp, T. (2007). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (pp. 9–33). Enschede, Netherlands: Netherlands Institute for Curriculum Development.

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research, Part A: An introduction* (pp. 11–50). Enschede, Netherlands: Netherlands Institute for Curriculum Development.

Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes: An overview on achievements and challenges. *ZDM Mathematics Education, 47*(6), 877–891. doi: 10.1007/s11858-015-0722-3

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53(*4), pp. 5–23.

Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. *Journal of Computers in Education, 2*(4), 399–420. doi:10.1007/s40692-015-0036-z

Ørngreen, R. (2015). Reflections on design-based research: In online educational and competence development projects. In J. Abdelnour Nocera, B. R. Barricelli, A. Lopes, P. Campos, & T. Clemmensen (Eds.), *Human work interaction design: Analysis and interaction design methods for pervasive and smart workplaces* (pp. 20–38). London: Springer.

List of the 21 selected disserations

Aksela, M. (2005). *Supporting meaningful chemistry learning and higher-order thinking through computer-assisted inquiry: A design research approach* (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Ekonoja, A. (2014*). Oppimateriaalien kehittäminen, hyödyntäminen ja rooli tieto- ja viestintätekniikan opetuksessa* [The development, invocation and role of learning materials in teaching information and communication technology] (Doctoral dissertation, University of Jyväskylä). Retrieved from https://jyx.jyu.fi

Hassinen, S. (2006*). Idealähtöistä koulualgebraa: IDEAA-opetusmallin kehittäminen algebran opetukseen peruskoulun 7. luokalla* [Idea-based school algebra: IDEAA teaching model development for 7th-grade algebra teaching] (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Hongisto, L. (2012). *Luonnontieteen historian opetus tieteenhistorian sisältöjen avulla: Kehittämistutkimus lukion historianopetuksen mahdollisuuksista* [History of natural sciences teaching with the help of history of science content: Design-based research on the possibilities of upper secondary school history teaching] (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Ikävalko, V.-M. (2017). *Mielekkään kemian non-formaalin oppimisympäristön kehittämistutkimus yhteistyössä työelämän kanssa* [Design-based research of a meaningful non-formal chemistry learning environment in cooperation with specialists in the industry] (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Juntunen, M. (2015). *Holistic and inquiry-based education for sustainable development in chemistry* (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Juuti, K. (2005). *Towards primary school physics teaching and learning: Design research approach* (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Kallunki, V. (2009). A historical approach to children's physics education: Modelling of DC-circuit phenomena in a small group (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Leppäaho, H. (2007). *Matemaattisen ongelmanratkaisutaidon opettaminen peruskoulussa: Ongelmanratkaisukurssin kehittäminen ja arviointi* [Teaching mathematical problem-solving skills in Finnish comprehensive school: Design and assessment of a problem-solving course] (Doctoral dissertation, University of Jyväskylä). Retrieved from https://jyx.jyu.fi

Loukomies, A. (2013). *Enhancing students' motivation towards school science with an inquiry-based site visit teaching sequence: A design-based research approach* (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Nieminen, M. (2008). *Ilmavoimien kadetit verkossa: Kokemuksia verkkopohjaisen oppimisympäristön käytöstä matematiikan perusopetuksessa* [Finnish air force cadets in network – experience in use of an online learning environment in basic studies of mathematics] (Doctoral dissertation, University of Jyväskylä). Retrieved from https://jyx.jyu.fi

Nousiainen, T. (2008). *Children's involvement in the design of game-based learning environments* (Doctoral dissertation, University of Jyväskylä). Retrieved from https://jyx.jyu.fi

Oikarinen, J. (2016). *Technology-enhanced statistics learning experiment: A case study at upper secondary level* (Doctoral dissertation, University of Oulu). Retrieved from http://jultika.oulu.fi

Pernaa, J. (2011). *Kehittämistutkimus: Tieto- ja viestintätekniikkaa kemian opetukseen* [Design research: Information and communication technology-based chemistry education] (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Ratinen, I. (2016). *Primary student teachers' climate change conceptualization and implementation on inquiry-based and communicative science teaching: A design* research (Doctoral dissertation, University of Jyväskylä). Retrieved from https://jyx.jyu.fi

Rautiainen, J. (2012). *Kehittämistutkimus: Ongelmalähtöinen kokeellinen korkeakouluopetus* [A design research: Problem- and inquiry-based higher education in chemistry] (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Rukajärvi-Saarela, M. (2015*). Tutkimuksellisuudesta innostusta alakoulujen kemian opetukseen: Kehittämistutkimus osallistavan luokanopettajan perus- ja täydennyskoulutuksen kehittämisestä* [Inquiry-based approach to inspire chemistry education in primary schools. A design research study to develop participatory pre- and in-service teacher education] (Doctoral dissertation, University of Jyväskylä). Retrieved from https://jyx.jyu.fi

Tomperi, P. (2015). *Kehittämistutkimus: Opettajan ammatillisen kehittymisen tutkimusperustainen tukeminen käyttäen SOLO-taksonomiaa - esimerkkinä tutkimuksellinen kokeellinen kemian opetus* [Design research: Research-based support utilising the SOLO taxonomy for teachers' professional development: Inquiry-based chemistry instruction as an example] (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Tuomisto, M. (2018). *Design-based research: Educational chemistry card and board games* (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Vartiainen, J. (2016). *Kehittämistutkimus: Pienten lasten tutkimuksellisen luonnontieteiden opiskelun edistäminen tiedekerho-oppimisympäristössä* [Design research: Promoting natural sciences inquiry-based learning of young children in a science club learning environment] (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

Vesterinen, V.-M. (2012). *Nature of science for chemistry education: Design of chemistry teacher education course* (Doctoral dissertation, University of Helsinki). Retrieved from https://helda.helsinki.fi

PUBLICATION
II

Using Manipulatives for Teaching Equation Concepts in Languaging-Based Classrooms

Daranee Lehtonen, & Jorma Joutsenlahti

In N. Pyyry, L. Tainio, K. Juuti, R. Vasquez & M. Paananen (Eds.), Changing subjects, changing pedagogies: Diversities in school and education (pp.164–185)
urn:isbn: 978-952-5993-23-3

CHAPTER NINE

USING MANIPULATIVES FOR TEACHING EQUATION CONCEPTS IN LANGUAGING-BASED CLASSROOMS

Daranee Lehtonen and Jorma Joutsenlahti

Introduction

Conceptual understanding is one of the most important proficiencies in mathematics. It enhances students' learning fluency and retention, facilitates their learning of new concepts, helps them to avoid errors, and promotes self-discovery (NRC 2001, 116–20). Moreover, it has been acknowledged that students' low performance in mathematics can result from an inadequate understanding of mathematical concepts (NRC 2001, 17–18; Ojose and Sexton 2009, 4). Nevertheless, school mathematics has typically emphasised algorithmic skills (Attorps 2006, 1; NRC 2001, 4). Recently, several countries have reformed their mathematics curricula in favour of conceptual understanding, instead of relying entirely on algorithms (e.g. Australian Curriculum, Assessment and Reporting Authority 2015; Common Core State Standards Initiative 2010; Finnish National Board of Education 2015).

Based on the work of Piaget, Bruner, and Montessori, educators and researchers have advocated the use of manipulative materials as hands-on learning tools for mathematical concepts understanding (McNeil and Jarvin 2007, 310; Uttal et al. 2013, 2). Previous studies have demonstrated that manipulatives assist children in developing their understanding of abstract mathematical concepts through multimodality and experiential learning (Puchner et al. 2010, 314; Uttal et al. 2013, 2). On the other hand, there is also a considerable amount of research that has demonstrated that manipulatives have no benefits to learners, and can sometimes even obstruct their learning (Martin, Svihla, and Smith 2012, 1–2; McNeil and Jarvin 2007, 312; Uttal et al. 2013, 2). The fact that the benefits of manipulatives are debatable has therefore caused uncertainties when it comes to applying them in practice.

To establish whether it is worth utilising manipulatives in mathematics classrooms, we compared classes using manipulatives to classes that did not. One-variable linear equations in third- to sixth-grade classes were used as a case study for our investigation because this important concept in algebra has usually been taught merely in terms of rules and procedures, rather than focusing on the concepts contributing to those rules (Magruder 2012, 13; NRC 2001, 259). This chapter attempts to use the studied context to resolve the disagreement over the use of manipulatives in practice. First, it reviews some of the proposed reasons that manipulatives may not be beneficial, and could even be damaging to children's learning and achievement. It also reviews current suggestions for benefiting from manipulatives and relevant models. Second, it presents the context, methods, and results of our study. Finally, it discusses the observed benefits of manipulatives in the studied context, and then proposes evidence-based implications for research, practice, and policy.

Literature review

Research into the effectiveness of manipulatives has yielded varying results, suggesting that their use alone may not automatically facilitate learning within mathematics classes. While there have been many explanations as to why earlier research concluded that the use of manipulatives is ineffective, some of these explanations have actually reached the opposite conclusion. However, several of the proposed explanations do signal the same conclusion: that is, there are potential advantages of using manipulatives, but that they do have to be used appropriately and effectively. Two recommendations regarding how to benefit from manipulatives can be drawn from previous studies. First, manipulatives should be used for fostering children's conceptual understanding rather than for acquiring procedural fluency. Second, while interacting with manipulatives, children need to make a connection between different representations constructed through the manipulatives and mathematical symbols of the same concept. (e.g. McNeil and Jarvin 2007; NRC 2001; Uttal et al. 2013.)

According to the recommendations from previous studies, using manipulatives to facilitate linking various representations of mathematical concepts together can contribute to students' conceptual understanding. To date, various translation models of multiple representations in learning mathematical concepts have been recommended (e.g. Goldin and Shteingold 2001; Joutsenlahti and Kulju 2010; Lesh, Landau, and Hamilton 1983). Besides proposing different representations of mathematical concepts, they have emphasised that "representational fluency"—which has been defined as (a) the ability to represent to-be-learnt mathematical concepts in various forms, and (b) the ability to bridge these representations—plays an important role in facilitating children's understanding of mathematical concepts. Several other studies have supported this understanding (e.g. NRC 2001; Suh and Moyer 2007; Teck 2013).

Our research employed "languaging mathematics", one of the translation models proposed by Joutsenlahti and Rättyä (2015), to enhance students' representational fluency while interacting with manipulatives. In this chapter, we refer to languaging mathematics as "languaging". The term languaging was previously introduced to didactic mathematics and second language learning in relation to verbal communication (see Bauersfeld 1995; Swain 2006). However, Joutsenlahti and Rättyä's (2015) concept of languaging goes further. They have defined languaging as an approach where a student expresses their own mathematical thinking by using one or more of the following four types of language: natural (verbal and written), pictorial, mathematical symbolic, or tactile language. Tactile language has been added to the current model so as to take account of mathematical thinking occurring when interacting with hands-on materials (i.e. manipulatives). Languaging-based instruction has been studied at different educational levels (Joutsenlahti and Rättyä 2015, 51–53). Based on these studies, and those of other researchers (e.g. Bauersfeld 1995; Suh and Moyer 2007; Teck 2013), it has been demonstrated that languaging plays a crucial role in mathematics classrooms in three aspects: the development of students' conceptual understanding, co-operative learning, and the assessment of students' mathematical thinking and learning.

Recently, the new Finnish National Core Curriculum for Basic Education (first to ninth grades) has emphasised mathematical concepts understanding as one of the most important mathematical proficiencies the curriculum aims to develop among students. Concrete and experiential teaching and learning have been underlined as a key instructional method. Additionally, languaging-based class activities have been included in the curriculum. Students are encouraged to develop their mathematical thinking and present it to their classmates and teachers through concrete tools, spoken and written language, and drawings (FNBE 2015, 128, 234–35, 374).

Context and methods

To be able to decide whether the use of manipulatives should be adopted into practice, we used one-variable linear equations in third- to sixth-grade classes as the lens through which the benefits of manipulatives were investigated. We conducted cross-sectional case studies utilising a concurrent triangulation approach of mixed methods as a strategy of inquiry. Qualitative and quantitative data were collected from teachers and students and then integrated for data analysis in order to holistically combine the research findings.

Participants

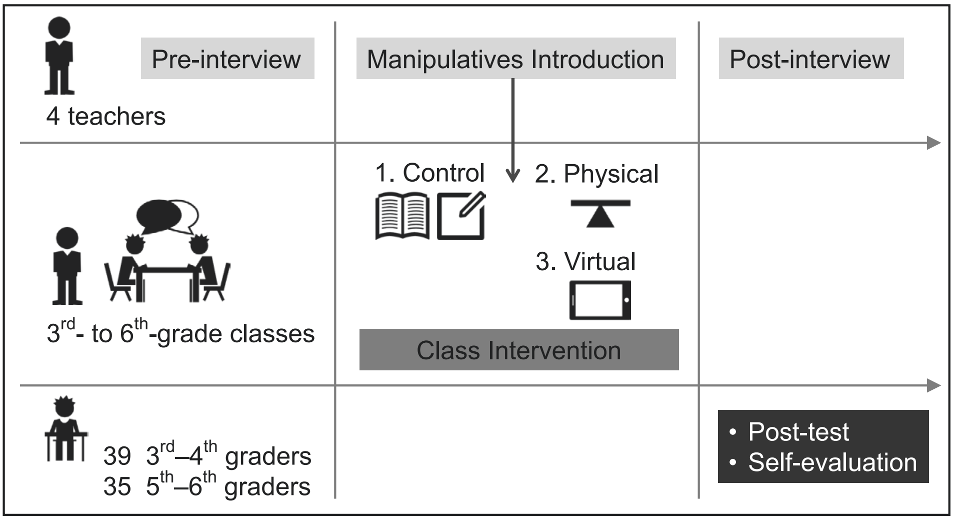
The study was conducted in one third-grade, one fourth-grade, one fifth-grade, and one sixth-grade class in a typical middle-size lower comprehensive school in southern Finland. This particular setting was selected as a case study for two reasons. First, schools and classes in Finland are homogeneous in terms of the students' socioeconomic background and performance (FNBE 2012, 2, 14; OECD 2013, 5–6). Moreover, all permanently employed class teachers in Finnish schools are required to have the same qualifications, including a master's degree and continuing professional development (OECD 2013, 10–11). Consequently, the homogeneity of Finnish comprehensive schools and class teachers made it possible for us to conduct the research in any Finnish school. Second, with limited resources and time, we expected to achieve the most fruitful results by studying third- to sixth-grade classrooms, in which the use of manipulatives has usually declined (Marshall and Swan 2008, 344).

Four class teachers (teaching experience 6–21 years) and 74 students (ages 9–12, N_{3rd}=23, N_{4th}=16, N_{5th}=14, N_{6th}=21) from the school participated in the study. Teachers' pre-interviews revealed that none of them had ever used the manipulatives intended for this study. Due to the mathematics contents included in the previous and new National Core Curriculum (FNBE 2015, 236, 375), all teachers had limited experience in teaching equations. Moreover, the students had low prior experience and knowledge of the mathematical content used in this study. Third- and fourth-grade students had not received any formal instruction in equations, while fifth- and sixth-grade students had received some instruction in solving one-variable linear equations with trial-and-error substitution of values and reasoning for the unknown. It could therefore be claimed that the homogeneity of the participants helped ensure validity and credibility of conclusions to be drawn from the research results.

Procedures

Four separate studies utilising identical research methods and procedures were conducted in the participants' classrooms during regular school hours. The studies were grouped into two grade bands (third- to fourth-grade and fifth- to sixth-grade) according to their similarity of instructional and post-test materials. Each study consisted of the following: 1) teachers' pre- and post-interviews; 2) class intervention, including one control group (a languaging-based classroom without manipulatives) and two treatment groups (a languaging-based classroom with a physical or virtual manipulative); and 3) students' post-test and self-evaluation (Figure 9-1).

Figure 9-1. Mixed-method research design



Teachers' interviews. The teachers participated in face-to-face semi-structured interviews before the treatment groups' lesson to illuminate any prior conceptions and experiences in teaching equations and utilising manipulatives which might affect the study. After all the class interventions, another interview was held to learn about their experiences, perceptions, and opinions about teaching and learning during the interventions.

Class interventions. Based on the students' prior mathematics performance during the academic year, each class teacher categorised them, within-class, into low, medium, and high attaining. They then assigned students from each category randomly, to either a control group or one of the two treatment groups. This was to ensure similarities between the instructional groups; that is, an equal number of students from each attaining level in all groups ($N_{control}$=25, $N_{physical}$=25, $N_{virtual}$=24). The same teacher taught the control and the treatment groups the same content for one 45–minute lesson. Before the study, each class teacher received instructional materials—a teacher guide and a student worksheet that were specially designed for the study to ensure conformity between the four studies. To avoid the influence of manipulatives on the control group's session, the manipulatives intended for the treatment groups' lesson were introduced to each teacher after the control group's lesson.

The lessons in all groups were almost identical. They each: 1) learnt the concepts of equivalence and unknown; 2) developed representational fluency through languaging—that is, translating and making connections between various representations (verbal and written, pictorial, and mathematical symbolic) of equations; 3) solved equations; and 4) checked the solutions. The only difference was that the treatment groups utilised the provided manipulatives (tactile language) to accompany their lesson. According to our literature review, one drawback of previous research on physical and virtual manipulatives for equations is that the manipulatives used in the research differed from each other in several ways (e.g. Magruder 2012; Suh and Moyer 2007). Consequently, the various dissimilarities of the manipulatives make it difficult to compare the research results in terms of representational difference. Thus, this study utilised a physical and a virtual manipulative that shared the same concept and a similar operation for the treatment groups in order to minimise the effect of their other differing attributes on the research results. During the lesson, one of the treatment groups utilised Hands-On Equations® consisting of a balance scale, number cubes representing constants, and pawns representing variables, while another group utilised a virtual version of physical Hands-On Equations® for the iPad, Hands-On Equations 1 applet (Figure 9-2).

Figure 9-2. Hands-On Equations® and Hands-On Equations 1 applet

The instructional materials were divided into two sets, one for the third- and fourth-grade studies and another one for the fifth- and sixth-grade studies. There were only two differences between the two sets. First, the third- and fourth-grade lessons addressed equations with a pictorial unknown and solving equations by trial-and-error substituting values and reasoning for the unknown, whereas the fifth- and sixth-grade lessons addressed equations with a letter as an unknown and equations solving by performing the same operation on both sides of the equation. Second, the number values used in the fifth- and sixth-grade lessons required more arithmetic skills than the ones used in the third- and fourth-grade lessons.

Students' post-tests and self-evaluations. After the class interventions, all students completed the same 45–minute post-test with no access to the manipulatives. The test was administered to determine the relative difference in students' learning achievement across instructional conditions. Two post-tests (one for all third and fourth graders and another one for all fifth and sixth graders) were designed in a similar way to the class intervention worksheets. Each post-test contained six open-response test items requiring students to: 1) translate six equations presented through different representations (written, pictorial, or mathematical symbolic) into two other representations; 2) solve the value of unknowns; and 3) algebraically check their solutions (Figure 9-3). Furthermore, the third and fourth graders had to explain the strategies they used to find the unknown's value, while the fifth and sixth graders had to write down their steps of solving equations. After completing the post-test, students evaluated their learning experiences and achievement.

Figure 9-3. Three types of fifth- and sixth-grade post-test items

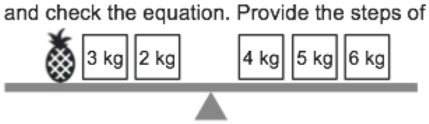
1. Visualize the equation below on the balance scale. Explain the equation in your own words. Solve and check the equation. Provide the steps of your solution.

   x + 7 = 4 • 8

   The equation means that
   Solution:

   x =
   Check:

2. Explain the picture below in your own words. Represent the picture with mathematical equation. Solve and check the equation. Provide the steps of your solution.



   Equation:
   Solution:

   One pineapple weighs kg
   Check:

3. Visualize the word problem below on the balance scale. Represent the word problem with mathematical equation. Solve and check the equation. Provide the steps of your solution.

   When a mother is weighing ingredients for her cake, she notices that three eggs weigh as much as 20 g of butter and 25 g of flour together.

   Equation:
   Solution:

   One egg weighs g
   Check:

Results and discussion

Quantitative data from the post-tests and self-evaluations of students in both grade bands were used to statistically determine whether languaging-based learning with physical or virtual manipulatives enhanced the students' understanding of equation concepts compared with the control groups. Additionally, qualitative data from the teachers' pre- and post- interviews, along with the classroom intervention observations, were concurrently utilised to develop empirical understanding of the research results. Subsequently, all the data was integrated and then interpreted to cross-validate the findings. To address the question of whether manipulatives should be adopted into practice, we next provide and discuss our findings according to our research methods (i.e. students' post-tests and self-evaluations, teachers' pre- and post-interviews, and class intervention observations) before finally turning to our convergent research results.

Students' post-tests

To determine the impact of each instructional condition on students' learning, we examined students' post-tests across three instructional conditions in both grade bands ($N_{\textbf{3rd-4th}}$=39, and $N_{\textbf{5th-6th}}$=35). Overall, both manipulative groups of both grade bands out-performed the control groups on the post-test (Figures 9–4 and 9–5). The third- and fourth-grade physical manipulative groups had the highest post-test average scores (Mean=17.7 out of 24, SD=4.0), followed by the virtual manipulative (Mean=15.9, SD=5.5) and the control groups (Mean=13.6, SD=5.0). Similarly, the fifth- and sixth-grade physical manipulative groups performed better on the post-test (Mean=17.5 out of 30, SD=9.9) than the virtual manipulative (Mean=16.0, SD=9.7) and the control groups (Mean=15.5, SD=8.5).

To test the null hypothesis for the difference of post-test scores across instructional conditions, we examined 95% confidence intervals for the means. We found overlaps of confidence intervals (Mean±1.96 SE) across instructional conditions of both grade bands (see error bars in Figures 9–4 and 9–5). Therefore, we further investigated the test statistic for the difference between two means. We found a significant difference at the 5% level between post-test average scores only in third- and fourth-grade physical manipulative and control groups, that is, the 95% confidence interval for the difference between the means of these two groups (0.6, 7.6) did not contain zero.

The findings from students' post-test average total scores indicate that students in all instructional conditions of both grade bands learned to represent and translate equations into different representations, solve one-variable linear equations, and check the solutions. Nevertheless, the third- and fourth-grade physical manipulative group significantly outperformed the control group on the post-test. When comparing post-test performance by grade band, fifth- and sixth-grade performance was lower than third- and fourth-grade performance. A possible explanation for this might be that the fifth- and sixth-grade content was more challenging than the third- and fourth-grade content. In fact, according to the Finnish National Core Curriculum for Basic Education 2014, the content taught in the fifth- and sixth-grade studies is taught in seventh to ninth grades (FNBE 2015, 236, 375). Based on fifth and sixth graders' post-test response, there is evidence of their equation concepts understanding. A fair number of them showed that they used mathematical operations taught during the intervention for solving equations and were able to arrive at the correct solutions. However, they did not receive full scores because of their incomplete steps of solving equations or arithmetic mistakes.

Figure 9-4. Third- and fourth-grade post-test average total scores (out of 24) by instructional condition (error bars = ± 1.96 SE)

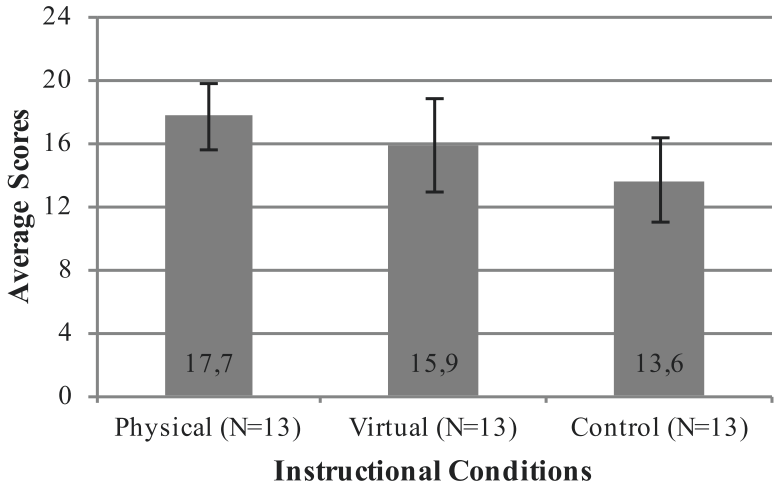
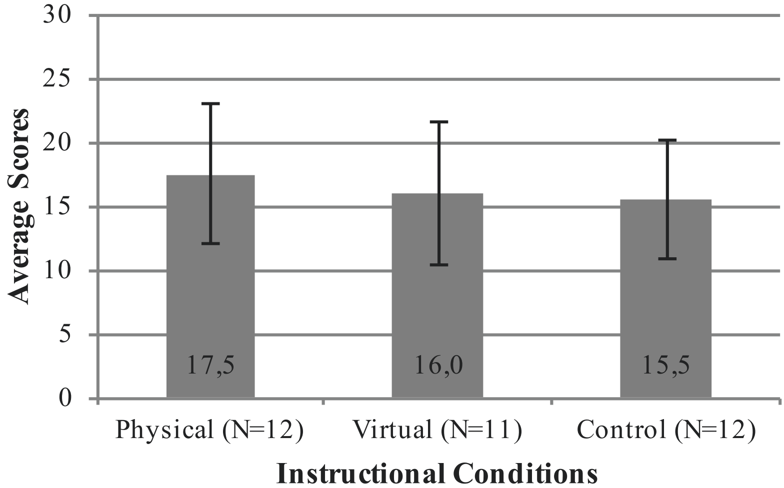


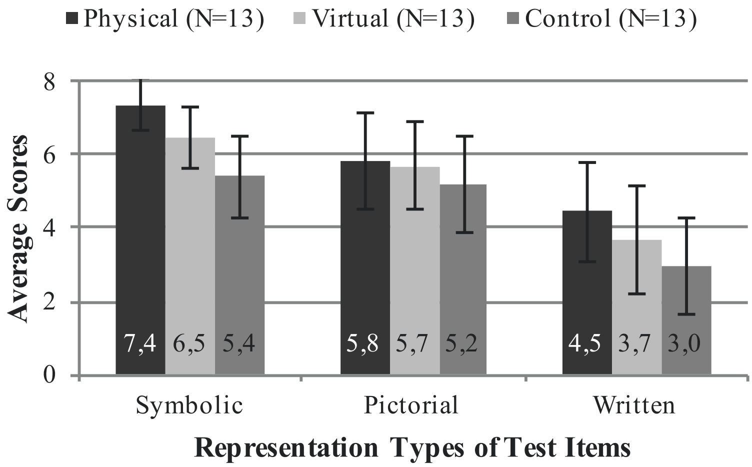
Figure 9-5. Fifth- and sixth-grade post-test average total scores (out of 30) by instructional condition (error bars = ± 1.96 SE)



In addition, we also investigated students' post-tests in three separate sections of different representation types (mathematical symbolic, pictorial, and written) to identify the influence of each instructional condition on students' performance within each post-test section. Figure 9-6 shows that the third- and fourth-grade physical manipulative groups performed best in all eight-full-score sections (Symbolic: Mean=7.4, SD=1.3; Pictorial: Mean=5.8, SD=2.4; Written: Mean=4.5, SD=2.5), relative to the virtual manipulative (Symbolic: Mean=6.5, SD=1.6; Pictorial: Mean=5.7, SD=2.2; Written: Mean=3.7, SD=2.8) and the control groups (Symbolic: Mean=5.4, SD=2.0; Pictorial: Mean=5.2, SD=2.4; Written: Mean=3.0, SD=2.4). As shown in Figure 9-7, even though the fifth- and sixth-grade physical manipulative groups did not perform best in every ten-full-score section, they performed consistently in all test sections (Symbolic: Mean=5.7, SD=2.5; Pictorial: Mean=5.9, SD=3.8; Written: Mean=5.9, SD=4.3), and better than the virtual manipulative (Symbolic: Mean=5.9, SD=2.1; Pictorial: Mean=5.5, SD=3.9; Written: Mean=4.6, SD=4.4) and the control groups (Symbolic: Mean=4.1, SD=3.5; Pictorial: Mean=5.9, SD=2.9; Written: Mean=5.5, SD=4.0).

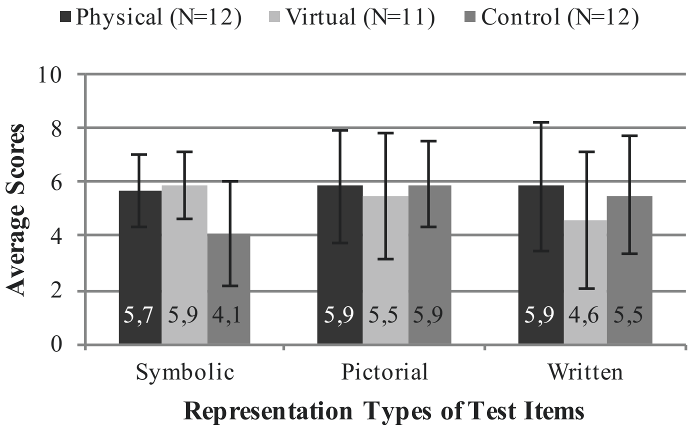
Our findings from the students' performance in different sections of the post-tests are mostly in agreement with the post-test average total scores. The third- and fourth-grade physical and virtual manipulative groups performed better than the control groups in all test sections. However, the difference in the fifth- and sixth-grade post-test performance across test sections was mixed. Although the fifth- and sixth-grade physical manipulative groups' performance in the symbolic section was lower than the virtual manipulative groups', overall, their performance was consistently close to 60% correct in all test sections. In contrast, the virtual manipulative and control groups' performance was inconsistent across test sections.

Figure 9-6. Third- and fourth-grade post-test average scores (out of 8) by representation type of test items across instructional conditions (error bars = ± 1.96 SE)



To discover whether instructional conditions influenced equations-solving strategies on the post-test, we investigated the students' post-test written solutions in terms of their strategies used for solving equations correctly. Their solutions were coded as: 1) trial-and-error substitution of values; 2) reasoning for the unknown; 3) mathematical operations (arithmetic and algebraic); and 4) other strategies. The "other strategies" code was used when students arrived at the correct answer without providing any explanation or steps for solving the equation, or when we were not able to identify their use of strategies. Our analysis did not include the situation where students did not solve the equation or solved the equation but did not arrive at the correct answer. Figure 9-8 shows that third and fourth graders solved 195 equations ($N_{physical}$=68, $N_{virtual}$=65, $N_{control}$=62) correctly by using mostly reasoning for the unknown (50.0% of physical, 55.4% of virtual, and 51.6% of control), followed by mathematical operations and trial-and-error substitution of values, respectively. As shown in Figure 9-9, fifth and sixth graders solved 139 equations ($N_{physical}$=50, $N_{virtual}$=44, $N_{control}$=45)

Figure 9-7. Fifth- and sixth-grade post-test average scores (out of 10) by representation type of test items across instructional conditions (error bars = ± 1.96 SE)



correctly by using strategies from three categories: reasoning for the unknown, mathematical operations, and other strategies. They were more likely to use mathematical operations (80.0% of physical, 79.6% of virtual, and 66.7% of control) than reasoning for the unknown or other strategies.

Our analysis of students' use of strategies to solve equations correctly reveals two main findings. First, in each grade band, the use of strategies for solving equations correctly did not differ overall between instructional conditions. Second, students in all conditions of both grade bands solved equations correctly by using mostly the strategies emphasised during the interventions (reasoning for the unknown in the third- and fourth-grade studies and mathematical operations in the fifth- and sixth-grade studies). Although we found no differences in the strategies used for solving equations correctly across the instructional conditions of each grade band, there was another potentially meaningful difference: On the third- and fourth-grade post-test, mathematical operations (which were never formally taught to third and fourth graders) were slightly more likely to be used for solving equations correctly by the physical manipulative groups than by the two other groups (Figure 9-8). Moreover, three-fifths (8/13) of the physical manipulative groups used this strategy to solve equations correctly at least once, whereas only half (7/13) of the control and one-thirds (4/13) of the virtual manipulative groups did. Likewise, on the fifth- and sixth-grade post-test, the physical and virtual manipulative groups were more likely to use mathematical operations taught during the intervention for solving equations correctly than the control groups (Figure 9-9).

Figure 9-8. Third- and fourth-grade percentage use of different strategies to solve equations correctly (out of 195) by instructional condition

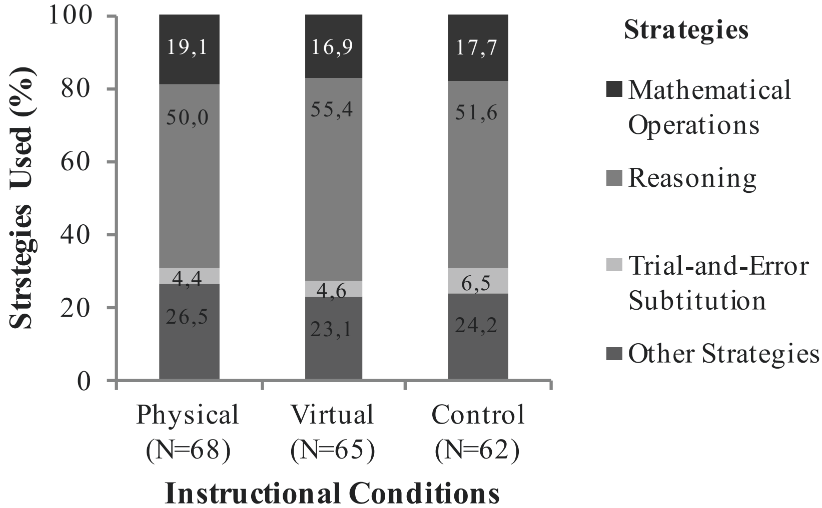
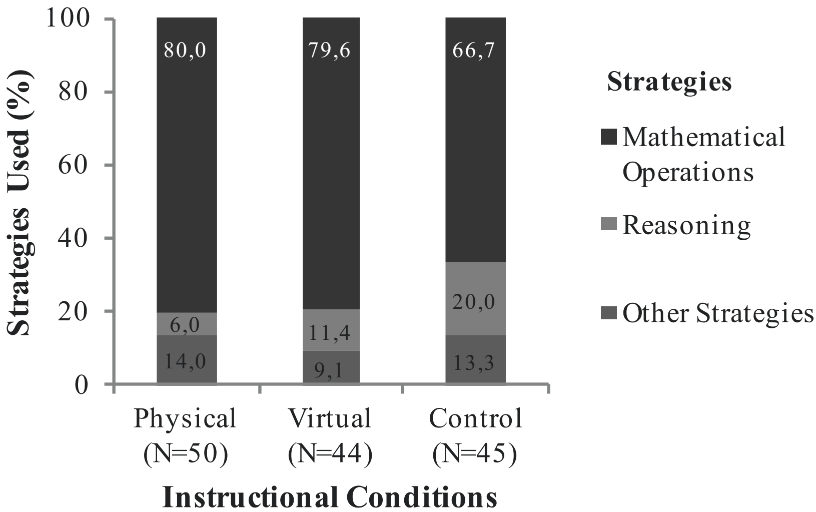


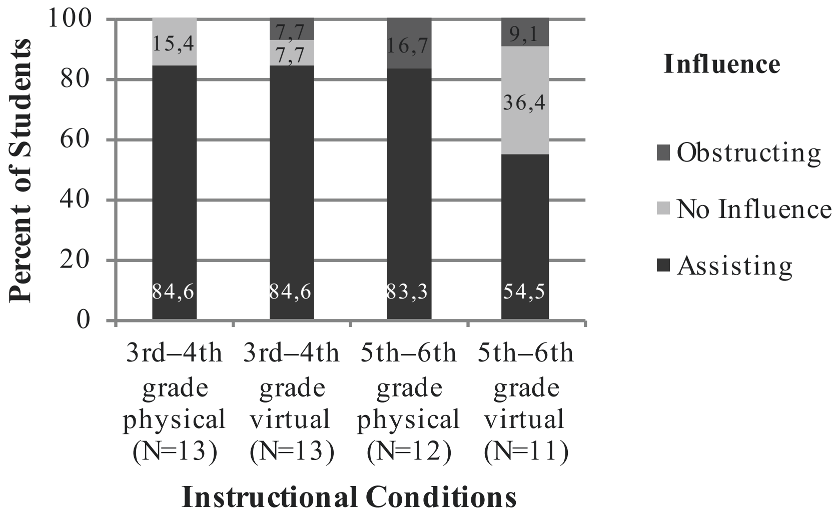
Figure 9-9. Fifth- and sixth-grade percentage use of different strategies to solve equations correctly (out of 139) by instructional condition



Students' self-evaluations

To crosscheck, against their post-test performance, how students assessed their equations learning development, we also concurrently examined their self-evaluation. In terms of learning development, students' self-evaluations generally support their post-test results. Three-fourths (10/13) of third- and fourth-graders in the physical manipulative groups considered themselves better at solving equations after the intervention, whereas less than half of the students in the other two groups (4/13 of virtual and 6/13 of control) considered that they had improved in this regard. Interestingly, although the virtual manipulative groups outperformed the control groups on the post-test, only one-third of them believed that their learning had improved, whereas the majority of them (9/13) considered that their learning had not improved. On the other hand, almost an equal portion of fifth and sixth graders across all three instructional conditions considered that their equations-solving performance had developed (6/12 of control, 7/12 of physical and 6/11 of virtual).

Figure 9-10. Percentage of students' opinions on the influence of manipulatives on their learning by grade band across instructional conditions



Next, we analysed the students' evaluation of the influence of manipulatives on their learning across instructional conditions and grade bands. Figure 9-10 shows that the majority of students in both treatment groups of both grade bands (third and fourth grades: 11/13 of physical and 11/13 of virtual; fifth and sixth grades: 10/12 of physical and 6/11 of virtual) thought that the manipulatives assisted them in learning equations. The results of the students' evaluation correspond to our findings from the post-tests and suggest that manipulatives assisted students in learning equations.

Teachers' pre- and post-interviews

We analysed teachers' pre- and post-interviews to ascertain their viewpoint on how different instructional conditions affected students' learning of equations. After the class interventions, all teachers regarded their physical manipulative group lesson as the most successful. They reasoned that the physical manipulative provided students with a concrete and tactile learning experience and also facilitated students' individual and group languaging. In their opinion, when physically handling the manipulative, students concretely constructed conceptual understanding of: 1) equation equivalence through the balance scale; 2) constants and variables through their distinct representations (number cubes and pawns); and 3) performing the same operation on both sides of the equation through actual action of removing the same elements from both sides of the balance scale. Thus, the physical manipulative groups had a better understanding of equation concepts compared to the other groups. They also believed that students in these groups would perform best on the post-test. Actually, during the pre-interviews, the fifth- and sixth-grade teachers regarded manipulatives as beneficial learning tools for younger students (who are likely to construct their understanding of new concepts through concrete experiences). Nevertheless, after the interventions, the sixth-grade teacher admitted that manipulatives could actually also assist older students (who are likely to have the capability for abstract thinking) in understanding more difficult concepts, such as equations. Moreover, the third- and sixth-grade teachers mentioned that learning how to use the physical manipulative did not take as much of their instructional time as they had expected. Rather, the physical manipulative was straight-forward and generally enabled students to learn and complete the exercise more rapidly than students in the two other groups. This finding agrees with Martin, Svihla, and Smith's findings (2012, 1), but differs from Magruder's (2012, 96).

While the physical manipulative was unanimously regarded as the most successful lesson, the teachers had mixed opinions as to which lesson should be ranked second. The third- and fourth-grade teachers considered their virtual manipulative lesson as the second best and their control lesson as the third best, whereas the fifth- and sixth-grade teachers were not confident about the second and third ranks. Both teachers mentioned that although even low-attaining students in the virtual manipulative groups were able to arrive at the correct solutions to equations during the lessons, they tended to only scroll and try different values for the unknown until arriving at a correct solution rather than developing their understanding. Therefore, these students might not actually understand the equation concepts and would thus perform worse than the control groups on the post-test. Apparently, the teachers' opinions regarding the students' learning achievement correspond to our findings from students' post-tests and self-evaluations. The physical manipulative seems to have had a positive influence on students in both grade bands, whereas the virtual manipulative had noticeable positive benefits only for third- and fourth-graders and appeared to function as an impediment to the development of fifth- and sixth-graders' equation concepts understanding.

Class intervention observations

We analysed the class intervention observations to find empirical evidence for convergent analysis. According to the observations, students in all instructional groups of both grade bands were able to represent the equivalence of the equations in various forms, solve equations, and check their solutions by themselves or with the assistance of their classmates or teachers. Nevertheless, the manipulative groups tended to work more independently, with minimal assistance from the teachers compared with the control groups.

Additionally, we found differences between the physical and virtual manipulatives. The physical manipulative groups had no difficulty in learning how to use the manipulative to model, solve, and check equations. They were more likely to work independently as well as co-operatively. When manipulating the physical manipulative, students usually said aloud (talking to themselves and their classmates) what they were doing or thinking. Consequently, they seemed to develop their understanding of equations gradually, through tactile, visual, and verbal languaging. Simultaneously, their classmates could also see and hear their mathematical thinking. Furthermore, the manipulative allowed the students to solve and check equations without strict procedure.

In contrast, the virtual manipulative was less likely to encourage verbal languaging and co-operative learning. Similar to previous research results (Moyer-Packenham et al. 2013, 36), the virtual manipulative groups tended to work silently and individually, especially when each student had their own iPad. They were more likely to hold the iPad for themselves instead of sharing. Consistent with Magruder's findings (2012, 101) and our teachers' interviews, a number of students seemed to manipulate the virtual manipulative by merely scrolling and trying until arriving at the correct solutions. Moreover, the applet's operational procedure appeared to be relatively complicated and inflexible.

Although students were able to model equations using the applet, several of them had difficulty in learning how to use it to solve and check equations. Consequently, some of them became confused and frustrated. Our findings from the class observations shows the benefits of the physical manipulative and demonstrates that the virtual manipulative functioned as a hindrance to the students' conceptual understanding, verbal languaging, and co-operative learning.

Discussion

Taken together, our convergent analyses demonstrate that students in languaging-based classrooms, across three instructional conditions, in both grade bands, learned to: 1) represent and translate equations into various forms (verbal and written, pictorial, and mathematical symbolic); 2) solve one-variable linear equations; and 3) check the solutions. These findings suggest that students in each instructional condition had developed their representational fluency, which indicated their understanding of equation concepts (NRC 2001, 119). As stated in the earlier literature review, the key to learning of mathematical concepts resides in assisting students in linking concrete and abstract symbolic representations of the same mathematical concepts. In this study, it was the languaging-based instruction that assisted students in classes, with or without manipulatives, in learning equation concepts.

In addition to languaging-based instruction, both manipulatives appeared to facilitate students' development of representational fluency and equation concepts understanding. Overall, both manipulative groups performed better than the control group on the post-tests, where no one had any access to manipulatives. This finding contradicts the claim, mentioned in previous studies, that students tend to over-rely on manipulatives without making connections to the mathematical concepts represented (Magruder 2012, 101; Uttal et al. 2013, 6). Furthermore, we found evidence that the physical manipulative-based instruction is superior to the two other instructional conditions for improving students understading of equation concepts. Students in the physical manipulative groups outperformed their classmates on the post-tests. Likewise, our findings from the students' self-evaluations, the teachers' interviews, and the classroom observations also reveal the positive impact of the physical manipulative on students' conceptual understanding, languaging, and co-operative learning. These findings—on the superiority of the physical manipulative over the virtual manipulative—do not support the previous studies that reported that virtual manipulatives are as beneficial as physical manipulatives to mathematics learning (Moyer-Packenham et al. 2013, 37; Suh and Moyer 2007, 156). This contradictory result may be because, in our study, a number of students manipulated the virtual manipulative in a rote procedural manner to get the correct solutions. Moreover, students were less likely to verbalise their mathematical thinking. These two factors may have negatively affected students' understanding of equation concepts.

In summary, the evidence from this study suggests that when making a connection between various representations constructed through manipulatives and mathematical symbols of the same concept, manipulative-based instruction is more likely to promote students' mathematical concepts understanding. This is consistent with previous research results (Suh and Moyer 2007; Teck 2013). Additionally, our findings support those of other studies in which manipulatives appear to assist students of any age (at any cognitive development level) in developing their understanding of new concepts (McNeil and Uttal 2009, 138).

The presented research results need to be interpreted with caution however, due to some of the inherent limitations—the most obvious of which being the nature of this research as an empirical study conducted in the real contexts of the classroom rather than a laboratory environment. However, the results of research conducted in an authentic teaching and learning context may have provided a better understanding of the real world compared with the findings of research carried out in a laboratory environment. Second, in spite of the teachers' similar qualifications, their different backgrounds and experience as well as their freedom to adjust their lessons may have affected the research results. However, we believe that this had no critical influence on our findings because in each classroom the same teacher taught the same content under all of the instructional conditions. Third, teachers and students may have acted unusually when being observed and video-recorded. Nonetheless, being in a familiar environment (one's own classroom) would likely help them to act more naturally. Fourth, the explanation on the post-test instructions and the encouragement provided during the tests may have had some influence on the students' post-test performance. Still, the explanation and encouragement were, in fact, necessary for students to gain a toehold because the test items were distinctly different from normal school tests and some students became nervous about taking a test after one 45–minute lesson. Fifth, because there was no pre-test before the class interventions, one could argue that the post-test results may have been skewed by the differences in the students' prior mathematics performance levels. However, each class teacher randomly assigned an equal number of students with different prior mathematics performance to each instructional group and so the concern regarding skewed post-test results could be ruled out. Lastly, when conducting cross-sectional case studies, a trade-off between breadth and depth of the study is an unavoidable issue. Due to our limited resources and time as well as the school's constraints (e.g. the number of students per class per teacher), the sample size was rather small and the duration of each class intervention was relatively short. As a result, it is difficult to extend these research findings to other educational contexts.

Conclusion and implications

Our research results highlight the benefits of manipulatives in classrooms for mathematical concepts understanding. These research findings not only provide implications for practice but also for policy-making and future research.

Regarding the question of whether manipulatives should be adopted into practice: our findings support the recommendations, mentioned in our literature review, regarding how to benefit from manipulatives; We recommend that manipulatives be used for facilitating students' understanding of new mathematical concepts. Additionally, manipulatives should be used to assist students in developing their representational fluency (i.e. making a connection between concrete representations constructed through the manipulatives and mathematical symbols of the to-be-learnt concepts) through languaging.

Two implications for policy can be drawn from the presented research results. Our first recommendation resonates with the mathematics instruction objectives of the Finnish National Core Curriculum for Basic Education (FNBE 2015, 128, 235): mathematics curricula should encourage instruction utilising manipulatives in collaboration with languaging to enhance students' representational fluency, which leads to their understanding of mathematical concepts. Our second recommendation is that teacher training should prepare pre- and in-service teachers to effectively benefit from manipulatives.

Despite the fact that this research provides valuable insights into the benefits of manipulatives in classrooms for equation concepts understanding, the limitations of this research make it difficult to generalise our findings to other classroom settings. Therefore, future studies should: investigate larger sample sizes, employ a longer period of class intervention, and add pre- and delay-tests to the research design. During the post-interviews, three out of four teachers mentioned that they plan to use both physical and virtual manipulatives to teach equations in the future. Therefore, it would be valuable to add another treatment group using physical and virtual manipulatives to further research on this topic. Furthermore, to better understand how the mathematics classroom can fully benefit from manipulatives, future research should consider investigating the benefits of manipulatives for diverse learners, different educational levels, and other mathematics content.

Acknowledgements

This paper is based upon the first author's doctoral research under the second author's supervision. The research was supported by the Finnish Cultural Foundation, Häme Regional Fund 2015. The authors thank the teachers and the students for being part of our research. Illustration and icons were designed by Freepik.

References

Attorps, I. “Mathematics Teachers’ Conceptions about Equations”. PhD diss., University of Helsinki, 2006.

Australian Curriculum, Assessment and Reporting Authority. *The Australian Curriculum: Mathematics F-10. 2015. www.australiancurriculum.edu.au/download/f10*

Bauersfeld, H. “Language Games in the Mathematics Classroom: Their Function and Their Effects". The Emergence of Mathematical Meaning: Interaction in Classroom Cultures. Eds. P. Cobb and H. Bauersfeld. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., 1995: 271–94.

Common Core State Standards Initiative. Common Core State Standards for Mathematics. 2010. www.corestandards.org/wp-content/uploads/Math_Standards1.pdf

FNBE (Finnish National Board of Education). *Perusopetuksen Opetussuunnitelman Perusteet 2014*. Tampere: Juvenes Print-Suomen Yliopistopaino, 2015.

FNBE (Finnish National Board of Education). *International Comparisons of Some Features of Finnish Education and Training.* 2012. www.oph.fi/download/141160_international_comparisons_of_some_features_of_finnish_education_and_training_2011.pdf

Goldin, G., and N. Shteingold. “Systems of Representations and the Development of Mathematical Concepts". *The Roles of Representation in School Mathematics NCTM Yearbook 2001.* Eds. A. Cuoco and F. Curcio. Reston, Virginia: NCTM, 2001: 1–23.

Joutsenlahti, J., and P. Kulju. “Kieliteoreettinen Lähestymistapa Koulu- matematiikan Sanallisiin Tehtäviin ja Niiden Kielennettyihin Ratkaisuihin". *Toisensa Kohtaavat Ainedidaktiikat. Ainedidaktiikan Symposiumi, Tampere, 2009.* Eds. E. Ropo, H. Silfverberg, and T. Soini. Tampere: Tampereen yliopisto, 2010: 77–89.

Joutsenlahti, J., and K. Rättyä. “Kielentämisen Käsite Ainedidaktisissa Tutkimuksissa". *Rajaton Tulevaisuus - Kohti Kokonaisvaltaista Oppimista.* Eds. M. Kauppinen, M. Rautiainen, and M. Tarnanen. Jyväskylä: Jyväskylän yliopistopaino, 2015: 45–61.

Lesh, R., M. Landau, and E. Hamilton. “Conceptual Models in Applied Mathematical Problem Solving Research". *Acquisition of Mathematics Concepts and Processes. Eds. R. Lesh and M. Landau. New York: Academic Press, 1983: 263–343.*

Magruder, R. “Solving Linear Equations: A Comparison of Concrete and Virtual Manipulatives in Middle School Mathematics". PhD diss., University of Kentucky, 2012.

Marshall, L. and P. Swan. “Exploring the use of mathematics manipulative materials: Is it what we think it is?” *The EDU-COM 2008 International Conference. Sustainability in Higher Education: Directions for Change.* Perth, 2008: 338–50.

Martin, T., V. Svihla, and C. Smith. “The Role of Physical Action in Fraction Learning". *Journal of Education and Human Development 5, 1 (2012): 1–17.*

McNeil, N., and L. Jarvin. "When Theories Don't Add Up: Disentangling the Manipulatives Debate". Theory into Practice 46, 4 (2007): 309–16.

McNeil, N., and D. Uttal. "Rethinking the Use of Concrete Materials in Learning: Perspectives from Development and Education"**.** *Child Development Perspectives 3, 3 (2009): 137–39.*

Moyer-Packenham, P., J. Baker, A. Westenskow, K. Anderson, J. Shumway, K. Rodzon, and K. Jordan. "A Study Comparing Virtual Manipulatives with Other Instructional Treatments in Third- and Fourth-Grade Classrooms". Journal of Education 193, 2 (2013): 25–39.

NRC (National Research Council). Adding It Up: Helping Children Learn Mathematics. Eds. J. Kilpatrick, J. Swafford, and B. Findell. Washington DC: National Academy Press, 2001.

OECD (Organisation for Economic Cooperation and Development). Education Policy Outlook Finland. 2013. www.oecd.org/edu/EDUCATION%20POLICY%20OUTLOOK%20FINLAND_EN. pdf

Ojose, B., and L. Sexton. "The Effect of Manipulative Materials on Mathematics Achievement of First Grade Students". *The Mathematics Educator 12, 1 (2009): 3–14.*

Puchner, L., A. Taylor, B. O'Donnell, and K. Fick. "Teacher Learning and Mathematics Manipulatives: A Collective Case Study about Teacher Use of Manipulatives in Elementary and Middle School Mathematics Lessons". School Science and Mathematics 108 (2010): 313–25.

Suh, J., and P. Moyer. "Developing Students' Representational Fluency Using Virtual and Physical Algebra Balances". Journal of Computers in Mathematics and Science Teaching 26, 2 (2007): 155–73.

Swain, M. "Languaging, Agency and Collaboration in Advanced Second Language Proficiency". Advanced Language Learning: The Contribution of Halliday and Vygotsky. Ed. H. Byrnes. London: Continuum, 2006: 95–108.

Teck, W. "Promoting Conceptual Understanding of Algebra Using Virtual Manipulative: A Study of Form 4 Students". Jurnal Penyelidikan IPG KBL 11 (2013): 1–14.

Uttal, D., M. Amaya, M. Maita, L. Hand, C. Cohen, K. O'Doherty, and J. DeLoache. "It Works Both Ways: Transfer Difficulties between Manipulatives and Written Subtraction Solutions". Child Development Research 2013: 1–13. http://dx.doi.org/10.1155/2013/216367

PUBLICATION
III

The Potentials of Tangible Technologies for Learning Linear Equations

Daranee Lehtonen, Lucas Machado, Jorma Joutsenlahti, & Päivi Perkkilä

Multimodal Technologies and Interaction, 4(4), Article 77
https://doi.org/10.3390/mti4040077

Multimodal Technologies and Interaction

Article

The Potentials of Tangible Technologies for Learning Linear Equations

Daranee Lehtonen [1,*], Lucas Machado [2,3], Jorma Joutsenlahti [1] and Päivi Perkkilä [4]

1 Faculty of Education and Culture, Tampere University, Åkerlundinkatu 5, P.O. Box 700, 33014 Tampere, Finland; jorma.joutsenlahti@tuni.fi
2 Faculty of Information Technology and Communication Sciences, Tampere University, Kanslerinrinne 1 Pinni B, P.O. Box 300, 33014 Tampere, Finland; lucas@demola.net
3 Demola Global, Åkerlundinkatu 8, 33100 Tampere, Finland
4 Faculty of Education and Psychology, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland; paivi.m.perkkila@jyu.fi
* Correspondence: daranee.lehtonen@tuni.fi

Received: 13 August 2020; Accepted: 21 October 2020; Published: 23 October 2020; Corrected: 10 January 2022

Abstract: Tangible technologies provide interactive links between the physical and digital worlds, thereby merging the benefits of physical and virtual manipulatives. To explore the potentials of tangible technologies for learning linear equations, a tangible manipulative (TM) was designed and developed. A prototype of the initial TM was implemented and evaluated using mixed methods (i.e., classroom interventions, paper-based tests, thinking aloud sessions, questionnaires, and interviews) in real classroom settings. Six teachers, 24 primary school students, and 65 lower secondary school students participated in the exploratory study. The quantitative and qualitative analysis revealed that the initial TM supported student learning at various levels and had a positive impact on their learning achievement. Moreover, its overall usability was also accepted. Some minor improvements with regard to its pedagogy and usability could be implemented. These findings indicate that the initial TM is likely to be beneficial for linear equation learning in pre-primary to lower secondary schools and be usable in mathematics classrooms. Theoretical and practical implications are discussed.

Keywords: manipulatives; multimodality; tangible user interface; educational technology; mathematics learning; linear equations; basic education

---

1. Introduction

Algorithmic skills have typically been the main focus of school mathematics [1,2]. Unfortunately, mathematics teaching that focuses only on procedural skills usually results in learning to memorise without understanding [2,3]. Linear equation solving, one of the important algebraic concepts, has commonly been taught in terms of rules and procedures rather than encouraging an understanding of the concepts leading to those rules [2,4]. Research has indicated that rule-based rote learning leads to misconceptions, inability to transfer procedures to other contexts, and forgetting rules [5].

Physical manipulatives (i.e., concrete objects that enable students to explore mathematical concepts through various senses) have traditionally been used to promote students' understanding of abstract mathematical concepts. Physical manipulatives have various benefits, including representing an abstract concept [6], encouraging physical action to facilitate learning [7], enhancing memory through physical action [8], and supporting embodied cognition [9]. Nevertheless, students who learn through physical manipulatives may not be able to connect the concrete representation with the symbolic representation of the same concept [10]. Over the last two decades, the use of virtual manipulatives (i.e., interactive pictorial representations of virtual objects on computers or tablets) has gained attention in mathematics education. Different advantages of virtual manipulatives include providing the simultaneous link between pictorial and symbolic representations, step-by-step scaffolding, immediate feedback [11], and drawing learners' attention to the to-be-learned mathematics [12]. However, there has been concern about the disadvantage of replacing rich physical interactions using physical manipulatives with clicking a mouse or tapping and scrolling a touch screen when manipulating virtual manipulatives [13,14]. A body of research signals that the combination of physical and virtual manipulatives may facilitate students' mathematics learning [11,12,15].

Recently, tangible manipulatives (TMs, i.e., a combination of physical and digital manipulatives) have been introduced. TMs offer a new form of interaction by allowing learners to naturally manipulate physical objects, which then provide output, typically through a graphical user interface (GUI). Thus, TMs may be the possible solution to the disagreement regarding the advantages and disadvantages of physical and virtual manipulatives.

To date, a considerable amount of research has focused on usability and engagement of TMs from the learners' perspective, thereby leaving the contribution of TMs to learning as well as teachers' perspective on TMs under-researched [16]. Additionally, the available manipulatives for equation solving are usually either physical (e.g., algebra tiles and Hands-On Equations [17]) or virtual (e.g., the Hands-On Equations applet [18] and virtual algebra balance scale applet on the National Library of Virtual Manipulatives website [19]). Recently, the Multimodal Algebra Learning (MAL) project [20] has attempted to develop virtual and tangible manipulatives for solving equations. Nevertheless, the system primarily focuses on pedagogy.

To holistically explore the potentials of TMs in mathematics classrooms, we proposed an initial TM for primary students to learn about linear equations. We developed the TM by taking pedagogy, usability, and practicality into account to ensure its successful classroom adoption. Then, we conducted a mixed-methods study in real classrooms in Finland to evaluate a prototype of the TM based on students' and teachers' feedback. This paper reports the design, development, and implementation of the initial TM as well as the classroom evaluation of its prototype. The classroom evaluation addresses the following research questions in particular:

1. What are the impacts of the proposed manipulative on students' learning achievement?
2. Does the proposed manipulative promote students' learning in terms of understanding equation-solving concepts, languaging, and learning through discovery and social interaction, and, if so, then how?
3. How do students perceive the usability of the proposed manipulative, and how well can they use it?

In the following section, we highlight the theoretical background of the study. Then, we discuss the design, development, and implementation of the initial TM. Next, we describe the classroom evaluation of the TM prototype. After that, we report and discuss our findings, reflect on the limitations of this research, and provide suggestions for future development and research. Finally, we conclude our study and explore its implications.

2. Theoretical Background

2.1. Equation Solving

Mastering equation solving is often challenging for students, e.g., [2,21–23]. To learn how to solve equations, students need to understand several concepts. Understanding mathematical equivalence includes an understanding of an equal sign as a relational symbol, of each side of the equation as an entity, and of various interchangeable ways of representing numbers and expressions [21]. An equation is composed of different terms (i.e., constants, variables, and coefficients). Thus, understanding the meaning of the mathematical symbols that represent those terms is essential [22]. Moreover, it is important to understand that the purpose of equation solving is to find the values of the variables that make the equation true [23] or to show that the equation has no real number-value solution. All these concepts informed our TM design for promoting students' understanding of equation solving.

2.2. Multimodal Mathematics

Mathematics is inherently multimodal. In mathematics, natural language (i.e., verbal and written), mathematical symbols (i.e., numbers and symbols), and visual representations (e.g., pictures, graphs, and diagrams) are typically used for meaning making. While the three resources intertwine to construct mathematical meanings as a whole, each resource has its own unique task [24–26]. According to O'Halloran [25,27], language assists in reasoning about the mathematical process and its results, symbols describe mathematical relations, and visuals present images to concretise mathematical relations. Thus, multimodality (i.e., multiple ways of communication) plays an important role in mathematics learning. Students are required to be able to interpret and make use of all these resources simultaneously [26]. This meaning-making process contributes to students' mathematical thinking and knowledge construction [26,28–30].

According to Bruner [6], a person constructs their own knowledge through physical actions, images, and abstract symbols. He proposed three modes of representation: (1) enactive representation (e.g., direct manipulation of objects), (2) iconic representation (e.g., pictures and graphics), and (3) symbolic representation (e.g., language and mathematical symbols) [6]. The Multimodal languaging model (referred to as languaging in this paper) was informed by Bruner's three modes of representation. Languaging can be defined as one's expression of their own mathematical thinking through four languages: natural, mathematical symbolic, pictorial, and tactile (e.g., manipulative manipulation) [30,31]. The use of manipulatives as mathematical representations is also recommended, for example, by Lesh et al. [32]. Using different languages to solve a mathematical problem or present the solution to a mathematical problem assists a student and their peer group in organising their own mathematical thinking and eventually gaining a better understanding of that mathematical concept or procedure [30,31,33].

Representational fluency—the ability to understand and construct multiple external representations of the same mathematical content and the ability to connect different modes of representation with each other—plays an important role in mathematics learning [32,34,35]. Research has indicated that representational fluency can contribute to mathematical knowledge and understanding [2,11,36]. Thus, representational fluency provides support for the languaging model. Languaging was used to guide our TM design and study (e.g., what modes of representations the TM should provide and how it should be used in classrooms) to ensure that the TM benefits students' equation learning.

2.3. Tangible Technologies for Learning

Tangible user interfaces (TUIs) emerged in the 1990s from Ishii and Ullmer's [37] pioneering work. TUIs enhance human–computer interaction by enabling users to physically operate (i.e., input) digital information through the manipulation of physical objects, thereby seamlessly interlinking the physical and digital world [38]. TUIs have been utilised in various application domains to date, including education.

The application of TUIs for learning is underlined by the integrated advantages of physical and virtual manipulatives. TUIs enable physical interaction with concrete objects, which provides a sense of physicality and embodiment, allows for natural bodily interaction [38], and engages multiple senses [39]. The multimodal interaction of TUIs enables mappings between physical and digital representations (i.e., physical touch and gestures with pictorial, symbolic, and other representations) [40] and thereby promotes knowledge transfer. TUIs allow parallel multi-user interaction, which encourages co-located and distanced collaborations [16,38,40]. Additionally, TUIs also provide immediate feedback; encourage independent exploration; promote facial, gestural, and verbal communication [16]; allow accessibility to various learners; and motivate learning [39].

The embedded embodied cognition perspectives provide support regarding the potential benefits of learning with TMs. From the embedded cognition viewpoint, manipulatives may work as students' external working memory, thereby allowing them to allocate their cognitive resources to learning during their interaction with manipulatives (i.e., online cognition) [9,15]. Moreover, the theory of Physically Distributed Learning (PDL) [7] suggests that a physical environment (e.g., manipulatives) that enables students' exploration can benefit their learning. From the embodied cognition viewpoint, students' previous sensorimotor experiences of interacting with manipulatives (i.e., offline cognition) may support their transfer of learning [9,41]. Gradually, students' dependence on manipulatives often decreases, while their internal cognition increases [9].

Studies have shown how TUIs have enhanced mathematics learning in different content domains and various educational levels. For example, number composition for primary school [14], fractions for primary school [42], geometry for pre-primary and primary school [43,44], trigonometry for undergraduate school [45], and algebra for lower and upper secondary school [20] have been studied. The potential benefits of TMs were used to inform our TM design and to guide our result analysis and discussion.

2.4. Learning through Discovery and Social Interaction

The use of manipulatives as hands-on learning tools for constructing abstract mathematical concepts has been recommended based on the work of Piaget, Bruner, and Montessori among the early theorists [8,10]. However, simply using manipulatives does not automatically contribute to mathematics learning [13,46,47]. Meaningful learning using manipulatives requires students to think and reflect on what they have experienced [47] and discuss their discoveries with others [46].

Discovery learning, proposed by Bruner [48], provides the theoretical foundations for the use of manipulatives to support mathematics learning through first-hand experience and reflection. Discovery learning is a process in which learners interact with the environment (e.g., manipulatives) and actively construct their own knowledge through inductive reasoning [48]. During discovery learning, learners are not left unaided but rather are assisted through guidance or scaffolding [49]. Previous studies [11,45] found that educational technology (e.g., virtual manipulatives and TMs) can be used to provide learners with first-hand experiences and guide or scaffold their inductive reasoning process during discovery learning [49].

According to Vygotsky [50], learning is a collaborative process in which learners co-construct knowledge through social interaction within the zone of proximal development. Thus, learners should work in small groups that have heterogeneous members [51] and should be encouraged to share and listen to one another's thoughts [52,53]. From the cognitive perspective, explaining the material to peers allows learners to retain that information in their memory and relate it to prior information stored in their memory [53]. Therefore, peer tutoring can benefit both the tutor and the tutee [54]. Hence, learning through social interaction not only supports languaging in mathematics classrooms [30] but also promotes meaningful learning with manipulatives (cf. discovery learning) [36,43,45]. The design of our TM and its pedagogic utilisation in classrooms was built on social constructivism, e.g., [55], particularly learning through discovery and social interaction, to ensure that the TM benefits students' mathematics learning.

2.5. The Finnish National Core Curriculum for Basic Education (NCC) 2014

The current Finnish National Core Curriculum (NCC) for Basic Education [56] has placed emphasis on teaching and learning for understanding mathematical concepts. Equation solving has been included in the NCC [56] as one of the key areas of mathematical content for Grades 3–9. Student of Grades 3–6 should be introduced to the concept of the unknown as well as be made to examine and solve linear equations through reasoning and experimentation (i.e., trial-and-error substitution of values for the unknown). Students of Grades 7–9 should be able to form and solve linear equations algebraically.

Multimodality and languaging have been incorporated into mathematics instruction in the NCC [56]. Students are encouraged to use concrete tools, spoken and written natural language, and drawings to present their own conclusions and solutions to others [56]. Thus, students should be provided with the opportunity to make mathematical meaning in different ways. For example, instead of only doing exercises by writing mathematical symbols on a worksheet, they can also do similar exercises by drawing, using manipulatives, and/or having discussions with peers.

The NCC [56] has emphasised learning through exploration and discovery, collaboration and social interaction skills, and the use of information and communication technology to enhance learning. Additionally, the differentiation of instruction based on students' personal needs and developmental differences has been highlighted to support their diversity (i.e., personal needs and developmental differences) [56]. Our TM and how it was to be used in classrooms were designed to align with the NCC. Later, its prototype was tested in real Finnish classrooms to evaluate whether the design conforms with the NCC.

3. Design, Development, and Implementation of the X-is Tangible Manipulative (TM)

3.1. Design Objectives and Principles

X-is ('X is equal to') is an initial TM designed and developed as a learning tool for primary school students to learn the concepts of linear equation solving. For this purpose, a set of design principles (DPs) was established based on the theoretical background (Section 2) as well as the literature review and empirical results derived from our initial research [57] and design concept evaluation [58]. Previously, we conducted research in primary schools to evaluate existing manipulatives [57] and our manipulative design concepts [58] in terms of their practicality.

- *DP 1.* *Promote understanding of equation solving:* A manipulative should assist students in learning equation solving by concretising the concepts of mathematical equivalence [21], different terms in an equation [22], and equation solving [23].
- *DP 2.* *Be in agreement with school curriculum:* Finnish teachers of basic education plan their teaching based on the Finnish NCC. Therefore, a manipulative should conform to the NCC [56] to ensure its use in classrooms.
- *DP 3.* *Support multimodality and languaging:* A manipulative should help students to link multiple representations of equation concepts and to express their mathematical thinking through various modes of meaning making [30,31].
- *DP 4.* *Enable learning through discovery:* A manipulative should enable students to learn through their first-hand experience and provide them with appropriate guidance and scaffolding [48,49].
- *DP 5.* *Assist social interaction:* A manipulative should encourage students to co-construct their knowledge through peer interaction while suppressing silent and individual activities [50–53].
- *DP 6.* *Be suitable for diverse learners:* A manipulative should provide differentiation of instruction based on students' diversity [56] by assisting students who are at different achievement levels in learning equation solving.
- *DP 7.* *Be easy to use:* An easy-to-use manipulative is more likely to be adopted in classrooms. According to our empirical research [57,58], the ease of use of a TM can be optimised through the following:
  - DP 7.1. Single point of interaction: The input and output of a TM should occur at the same point of interaction (i.e., a 'co-located' design [16]) to allow students to manipulate physical objects and look at a GUI without moving their sight back and forth.
  - DP 7.2. Use of base-10 blocks as physical objects: Base-10 blocks are widely used manipulatives for learning of number sense, place value, and operation in various countries, including Finland. All the schools that participated in our previous

studies [57,58] were familiar with the base-10 system. Thus, it can be conveniently used as physical manipulative objects to reduce students' cognitive friction [59].

DP 7.3. Straightforward user interface (UI): A simple UI enables teachers and students to use it with ease. Consequently, it saves time spent on its utilisation and prevents frustration.

DP 8. *Be feasible for classroom and school practice:* A manipulative design that takes the following factors related to classroom and school practice into account is more likely to be adopted and used in classrooms:

DP 8.1. Affordability: Our empirical research [57,58] and previous studies [60–62] have revealed that schools' tight budgets have a highly negative impact on the acquisition of manipulatives. Therefore, an affordable manipulative is more likely to be acquired under this financial pressure.

DP 8.2. Practicality and convenience: According to our research [57,58] and that of others [60–62], time constraints, manipulative preparation and organisation issues, and limited storage space are among the possible hindrances to manipulative acquisition and utilisation. Based on our concept evaluation [58], the following properties can increase the practicality and convenience of manipulatives:

- A straightforward manipulative requires less time and effort spent on preparation, instruction, operation, and clean-up.
- A portable manipulative can be easily circulated in the classroom and around the school.
- A proper size and sensible number of parts allows for easy storage and prevents pieces from becoming lost or mixed up.
- A manipulative should be compatible with Android tablets or iPads due to (1) the increasing number of these devices in Finnish schools as a result of the digitalisation of learning environments that is encouraged by the current NCC [56] and (2) the growing number of Finnish teachers and students who have these devices.

DP 8.3. Durability: According to our concept evaluation [58] and previous research [61], teachers are concerned that manipulatives could get broken when used in classrooms. The durability of manipulatives increases when there are no fragile or moving parts.

DP 8.4. Utility: Our concept evaluation [58] has revealed that high utility is one of the teachers' criteria for acquiring manipulatives. A manipulative that can be used for different grade levels or content areas is preferable. Its compatibility with schools' existing infrastructure and equipment is also important.

X-is was developed with the goal of meeting the above-mentioned design principles to ensure its success in classrooms. Sections 3.2–3.4 present the X-is system and the design principles that guided its design.

3.2. The Implemented Architecture

X-is is comprised of two parts: physical objects and a tablet application (DP 8.2, 8.4). The application working prototype was developed from 2018–2019 by a team of undergraduate and graduate students and their supervisor at the Faculty of Information Technology and Communication Sciences in close cooperation with the first author.

Our current solution is rather complicated in terms of components and set-up. The reasons for this are clarified in this section, while the associated limitations and possibilities for improvement are discussed in Section 5.4. The input interactions with the tablet application occur though the placement and removal of physical objects (Figure 1i) on a tablet screen (Figure 1ii) that functions as an external display of a computer running the application (Figure 1iii), which was developed using the Unity development platform. The free software Spacedesk [63] was used to enable a wireless video and audio connection (Figure 1iv) between the computer and the tablet. An external USB web camera that supports a 720 p resolution (Figure 1v) is connected to the computer and positioned using a tripod (Figure 1vi) so that the physical objects can be seen on the tablet screen. Image recognition algorithms making use of OpenCV for Unity Library [64] use the web camera image to detect the objects' positions and their amounts. This detection is based on distinction of the objects' sizes and colours. According to the information derived by the recognition algorithms, the computer provides visual and audio outputs to the tablet.

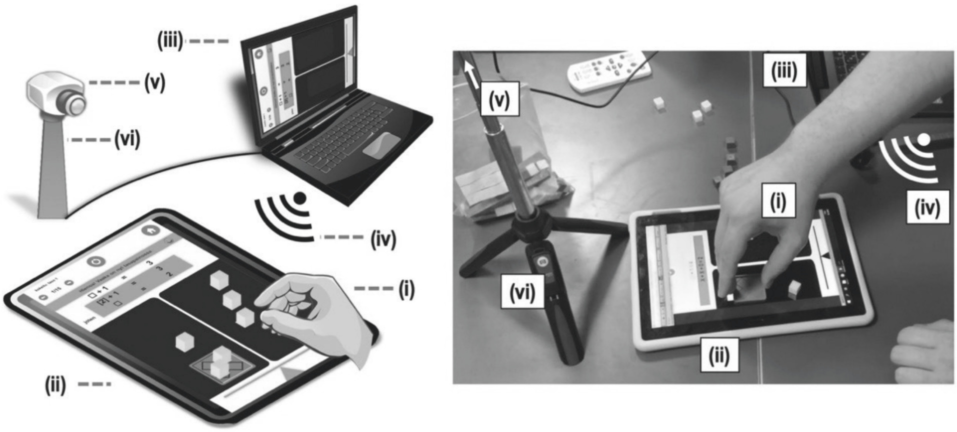


Figure 1. Architecture and components of X-is. (**i**) Placement of physical objects; (**ii**) Tablet screen; (**iii**) Computer; (**iv**) Wireless connection; (**v**) Web camera; (**vi**) Tripod.

The original conceptualisation involved the use of only an Android tablet with a minimum display size of 10 inches, which is large enough to enable two students to work at the same time (DP 5) and to place physical objects on the screen, combined with a web camera. However, during development, the running platform was changed to Windows 10 64-bit due to an issue with external devices running on Android. For an external web camera to work directly with an Android tablet, it must comply with a specification for external image devices. Android then recognises those devices and make them available to apps through its internal application programming interface (API), which is different on every version of Android and may break compatibility. However, OpenCV for the Unity Library was not able to access external image devices with the Android API used during the development period. Thus, a computer that runs Microsoft Windows or macOS and is able to connect the camera to the Unity Library was needed. In that fashion, an iPad running Spacedesk Software could also be used as a tablet because it works as an output device (i.e., secondary machine) and the actual processing happens in the computer (i.e., primary machine).

Another alternative set-up is to run the application directly on a tablet with a Windows 10 64-bit system, which is then connected to an external web camera (Figure 2). This would be more straightforward for teachers (DP 8.2). However, we did not have access to a tablet running Windows during our research.

Figure 2. Alternative set-up of the system with a tablet running Windows.

3.3. Object Tracking Alternatives

During development, various possibilities for object tracking were considered to determine the best balance between our target design principles and technological constraints. The initial design approach was to directly detect objects (i.e., types, amounts, and positions) on the tablet touch screen without requiring an external camera. This solution would not only enable seamless and direct interaction but also increase the product portability and practicality as well as decrease error factors. However, implementation proved to be unfeasible due to the limitations of the technology currently employed in touch screens. First, the vast majority of commercial touch screens currently available use projected capacitive sensing technology [65]. This technology can detect touch by capacitance variations in the screen surface, which occur, for example, with human finger contact but would not sense our plastic objects without conductive markers attached to them. Furthermore, it is not possible for the touch screens to distinguish features such as colour or size through touch. Second, the screen touch controller is optimised for finger contact and uses techniques such as 'grip suppression' and 'palm rejection' to filter out noise and undesired touches [66]. In our case, these would probably disable detection of stationary objects, limit the number of detectable objects, or merge them incorrectly as an adaptation to large fingertips.

Other tracking techniques—such as magnetic sensing (e.g., GaussSense [67]), near-field communication (e.g., Nintendo Amiibo [68]), and smart objects (e.g., MAL-Smart tiles [69])—would at least require embedding magnets, tags, or circuit boards to detectable objects. Embedment would be challenging in our case because a base-10 unit—our smallest object—is a 1 cm cube. These techniques would also increase the cost and fragility of our manipulative. Moreover, they typically require custom hardware for a specific purpose, thereby limiting utility. Neither optical tabletop systems (e.g., tanGible Augmented INteraction for Edutainment (GAINE) [70]) nor mixed reality tabletop systems (e.g., Augmented Reality (AR) enhanced tabletop system [44]) were possible because these typically require a large amount of stationary installation (e.g., computer, camera, projector, and screen) beneath or over the tabletop. Thus, tabletop systems are neither portable nor affordable. Another possibility was a vision-based tracking mixed reality system, where a mirror is placed in front of a tablet camera and enables the system to detect objects on a flat surface in front of the tablet (e.g., Ceibal Tangible (CETA) [14]). This system is inexpensive, portable, and would be able to detect our objects. However, its TUI input happens in front of a tablet, while its GUI output is displayed on a tablet screen (i.e., a 'discrete' design [16]). Thus, it does not allow for a single point of interaction, which is one of our most important design principles.

All in all, our selected object tracking solution enables the application system to meet many of our target principles (DP 7.1, 8.1–8.4). However, practicality and convenience are somewhat limited compared to the initial design approach.

3.4. Features and Interactions

X-is uses two types of physical objects (Figure 3a,b)—base-10 blocks representing constants (DP 7.2) and X-Boxes, specially designed for X-is, representing unknowns—for its input to assist students in recognising the distinction between different terms in an equation (DP 1). The application contains exercises divided into two levels for primary students in different grades (DP 8.4) to learn the concepts of linear equation solving using different strategies aligned with the Finnish NCC (DP 2). The goal of Level 1 is to solve equations by substituting values for the unknown so that both sides of an equation are equal (Video S1). The goal of Level 2 is to isolate a single unknown on one side of an equation and the constants on the other side by subtracting the same quantity of constants and unknowns from both sides of an equation (Video S2). This algebraic strategy was chosen because of its emphasis on the equivalence principle (DP 1). The balance model is used as a didactic model to assist students' understanding of linear equation solving [23]. Either side of the scale represents either side of the equation, while the movement of the scale (tilting or balance) represents the equality of the mathematical expressions on either side of the equation.

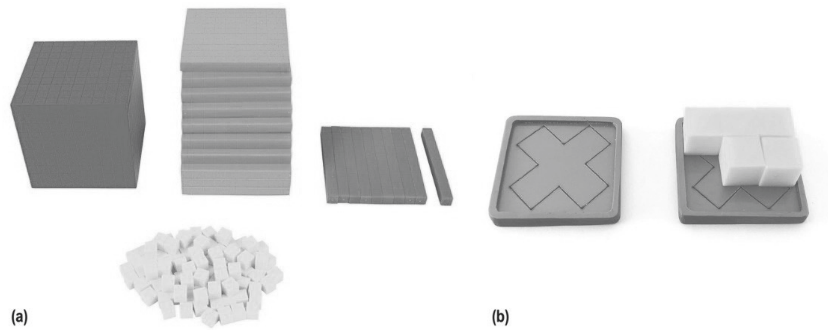


Figure 3. Physical objects used in X-is. (**a**) A base-10 block system, an existing and widely used manipulative, consisting of four different sizes and colours representing their individual place values: units (one's place), rods (ten's place), flats (hundred's place), and cubes (thousand's place). The system is familiar to teachers and students. Reprinted with permission [71]; (**b**) X-Boxes (left) specially designed for X-is, which can be used with base-10 units when solving Level 1 equations (right).

Each level starts with an introductory animation demonstrating how to model and solve an equation using the specific strategies for that level (DP 4). After that, there are exercises with gradually increasing difficulty. First, students have to model the given equation by placing base-10 blocks and X-Boxes on either side of the scale on the tablet screen (Figures 4a and 5a). This task requires students to translate the symbolic representation of the equation into a physical representation, thereby developing their representational fluency. Then, they have to solve the equation (i.e., find the value of an X-Box) by adding base-10 blocks to the X-Boxes to balance the scale in Level 1 (Figure 4b) or by removing the same number of physical objects from both sides of the scale to maintain its balance in Level 2 (Figure 5b). Solving the equations by balancing the scale emphasises the equal sign as a relational rather than operational symbol (DP 1).

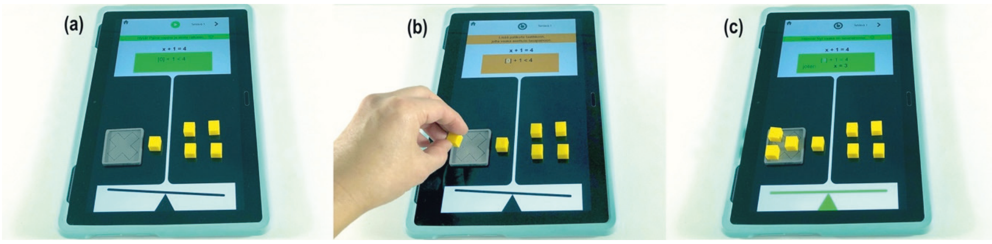


Figure 4. The steps to complete the Level 1 exercise x + 1 = 4. (**a**) Modelling the given equation; (**b**) Solving the equation by substituting the unknown's value; (**c**) The solved equation and balanced scale.

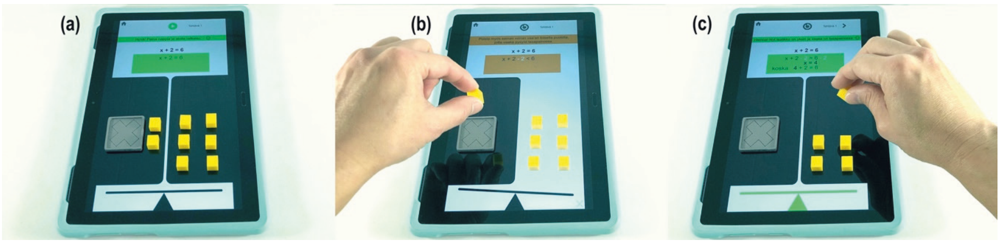


Figure 5. The steps to complete the Level 2 exercise x + 2 = 6: (**a**) Modelling the given equation; (**b**) Solving the equation by doing the same operation on both sides of the scale; (**c**) The solved equation and balanced scale.

The X-is TUI is comprised of multimodal inputs and outputs to help students link multimodal representations of equations and express their mathematical thinking through various modes of meaning making (DP 3). Moreover, the application provides instant scaffolding (i.e., guidance and feedback) in the form of pictures, text, mathematical symbols, and sounds to guide students on what to do or to inform them of the correctness of their actions (DP 4). Instead of tapping and scrolling a touch screen, students interact with the application by placing physical objects on or removing them from a working zone (Figure 6ii) above a scale (Figure 6i). The working zone is separated into the left and right sections, which are reserved for each side of the scale. When the 'weights' are unequal, the scale tilts towards the heavier side; in contrast, when both sides of the scale are equal, the scale is balanced. To emphasise that an equation is solved, in addition to being balanced, the scale turns green, which is accompanied by a ding sound and a textual compliment (Figures 4c and 5c). The given equation is situated at the top of a mathematical symbol zone, while the math sentence for the current equation-solving process is presented in a math expression window below the given equation (Figure 6iii). A text window provides textual instructions, guidance, and feedback (Figure 6iv). For example, when solving the equation x + 2 = 6, after students have removed two base-10 units from the left side of the scale, the text window provides the textual message: 'Remove the same quantities from the other side of the scale to keep the scale balanced!' The right side of the working zone blinks, and the scale tilts towards the right side (Figure 6).

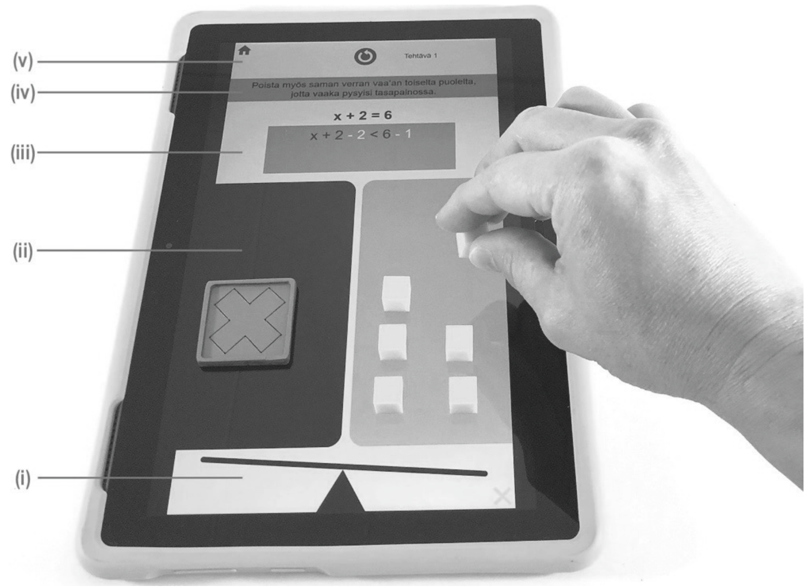


Figure 6. GUIs on the screen from bottom to top. (**i**) Scale; (**ii**) Working zone; (**iii**) Mathematical symbol zone; (**iv**) Text window; (**v**) Navigation and operation menu.

X-is employs a user-friendly UI to enable students' natural interactions (DP 7.3). The overall graphic design is clear and simple. The screen contains only the necessary elements to avoid students' distraction from their learning process. The navigation and operations menu uses minimal, easy-to-understand graphical icons (Figure 6v). Concise and clear sentences appear one at a time. Button tapping is kept to a minimum. After a delay, if the physical objects on the screen have not been moved, the application responds according to the last manipulation (rather than requiring the student to tap in the affirmative). Occasionally, students are required to tap buttons on the screen, for example, to start solving the equation after modelling it or to proceed to the next exercise after solving an equation.

3.5. Prototyping

We initially intended to present a working (fully interactive) prototype in the classroom to validate its technical feasibility and evaluate its potential. However, due to time constraints, the development team was only able to get Level 1 of the application fully functional. This made it problematic to use the working prototype as a task-oriented prototype [72] for classroom evaluation because the evaluation requires students to learn to solve equations at both levels. Moreover, at the moment, the performance of the image recognition algorithm depends largely on external factors, including environmental lighting conditions and the position setting of the external camera during system calibration. If the current prototype were used in a real classroom setting, it is likely that the image recognition may not function reliably. Thus, another prototype that allows students to uninterruptedly perform both levels of equation solving was required.

Wizard of Oz [72] is a rapid prototype that was created using Microsoft PowerPoint for the classroom evaluation. The prototype contains all the needed UIs for Levels 1 and 2. To create the illusion of a functioning manipulative, while the user interacted with physical objects and the tablet (i.e., secondary machine), the researcher (i.e., the Wizard) unnoticeably operated a PowerPoint slideshow by hand on a computer (i.e., primary machine) to respond to the user's actions.

4. Classroom Evaluation of X-is

4.1. Participants

We recruited participants from primary schools and lower secondary schools in southern Finland, and the equity of the Finnish education system [73] allowed for convenience sampling. Ethical practices were assured throughout the study in accordance with the European Code of Conduct for Research Integrity [74]. For instance, the participants participated in the study on their own wills, informed consent was obtained from all participants or their legal guardians prior the study, and the participants' privacy was protected. The student participants included 12 fourth graders (ages 10–11), 12 fifth graders (ages 11–12), 35 seventh graders (ages 13–14), and 30 eighth and ninth graders (ages 14–16). Additionally, one fourth- and one fifth-grade class teacher (teaching experience: 10–11 years; moderate experience using physical and virtual manipulatives), three lower secondary school mathematics teachers (teaching experience: 10–27 years), and a special education teacher (teaching experience: 4 years in primary and lower secondary schools) participated in the study. The students at each grade level had mixed achievement in mathematics.

4.2. Research Design and Procedures

The classroom evaluation was conducted to examine the potentials of X-is in classrooms in terms of leaning achievement, learning support, and usability. The convergent design (previously known as concurrent or parallel design) of the mixed methods approach [75] was used as the strategy of inquiry. First, both qualitative and quantitative data were collected simultaneously from students and teachers using various methods (see Table 1). The purposes and research design of the classroom evaluation are outlined in Table 1.

Table 1. Purposes and research design of the classroom evaluation.

Purposes	Methods			
	Class Intervention	Paper-Based Test	Questionnaire and Interview	Thinking Aloud
Learning Achievement		• Both groups of fourth graders (n = 12) • Both groups of fifth graders (n = 12) • Seventh graders (n = 35) • Eighth and ninth graders (n = 30)		• X-is group of fourth graders (n = 6) • X-is group of fifth graders (n = 6)
Learning Support	• Both groups of fourth graders (n = 12) • Both groups of fifth graders (n = 12) • Fourth- and fifth-grade class teachers (n = 2)		• X-is group of fourth graders (n = 6) • X-is group of fifth graders (n = 6) • Fourth- and fifth-grade class teachers (n = 2) • Lower secondary school mathematics teachers (n = 3) • Special education teacher (n = 1)	• X-is group of fourth graders (n = 6) • X-is group of fifth graders (n = 6)
Usability	• Both groups of fourth graders (n = 12) • Both groups of fifth graders (n = 12) • Fourth- and fifth-grade class teachers (n = 2)		• X-is group of fourth graders (n = 6) • X-is group of fifth graders (n = 6)	• X-is group of fourth graders (n = 6) • X-is group of fifth graders (n = 6)

Note: Fourth and fifth graders were divided into two groups: the paper-based intervention and the X-is intervention.

The fourth and fifth graders, who had never received any formal instruction regarding equation solving before this study, participated in one of the languaging-based class interventions, either paper-and-pencil or X-is. After the intervention, both groups individually completed a paper-based test. Typically, Finnish students of basic education (first to ninth grades) only obtain their academic knowledge in school. For this reason, it could be assumed that the fourth and fifth graders in this study had no prior knowledge of equation solving. Therefore, a pre-test was unnecessary, and only the paper-based test was conducted after the intervention to evaluate the learning achievement of new knowledge. After the test, each X-is student was asked to think aloud as they solved equations using X-is. At the end, a questionnaire, accompanied by an interview, was completed with individual students from the X-is groups.

The seventh through ninth graders took the same paper-based test, but did not participate in any class intervention. Prior to the test, they had received several equation-solving lessons (their first, second, or third equation course respectively) as a part of their normal school curricula, thereby serving as a comparison group (i.e., students in traditional classrooms). For the same reason mentioned in the previous paragraph, it could be assumed that the seventh through ninth graders in this study had only obtained their knowledge of equation solving in school. Thus, a pre-test was unessential, and the paper-based test would evaluate their learning achievement after attending traditional classrooms. This research design was employed, because the Finnish basic education is organised according to the NCC [56], where algebraic linear equation solving is only taught in Grades 7–9. Therefore, a control group from fourth and fifth grades that had received traditional classroom instruction in algebraic equation solving could not be found. Students from lower secondary schools in the same area were recruited as the comparison group to ensure the participants' homogenous socioeconomic and academic background.

A lesson plan and a worksheet developed for the class interventions were provided to the fourth- and fifth-grade class teachers before the interventions. After the interventions, a questionnaire and an interview were administered for both class teachers. To discover the possible utilisation of X-is in other classroom contexts, X-is was also evaluated by three lower secondary school mathematics teachers and a special education teacher. First, the X-is Wizard of Oz prototype was used to demonstrate to the teachers how to solve equation exercises at both levels (see Section 3.4). Then, the teachers tried X-is. Next, they were asked to complete the same teacher questionnaire and interview.

Finally, the collected data were concurrently analysed before being compared and combined to cross-validate the findings and holistically understand the results with regard to the following aspects:

- To determine the impacts of the languaging-based instruction (with or without X-is) on students' learning achievement based on the results from paper-based test and thinking aloud sessions.
- To discover whether and how X-is promoted students' learning (i.e., their understanding of equation-solving concepts, languaging, and learning through discovery and social interaction) based on the results from class interventions, student and teacher questionnaires and interviews, and thinking aloud sessions.
- To investigate how students perceived the usability of X-is and how well they could use it based on the results from class interventions, student questionnaire and interview, and thinking aloud sessions.

4.3. Data Collection and Analysis

4.3.1. Class Intervention

The class interventions were carried out to compare between the two languaging-based instructional conditions—paper-and-pencil (n = 12) or X-is (n = 12) interventions—with regard to their usefulness in terms of supporting students' learning. Each fourth and fifth grader participated in one 45-min class intervention led by their class teachers in their classrooms during regular school hours. For each condition, the teachers divided their students into three pairs—one high- and one medium-attaining student, two medium-attaining students, and one medium- and one low-attaining student—based on the students' mathematics achievement throughout the school years. The pair assignments also took into account students' relationships and ability to work together to ensure collaboration within pairs. Pair work was used for the intervention instead of group work to reduce group collaboration development time, while the combination of students in each pair was employed to assure their learning through social interaction [51,76].

At the beginning of the intervention, the teachers taught the concepts of equations, equivalence, the unknown, and equation solving (by substituting values of the unknown in the fourth-grade interventions and by doing the same operation on both sides of an equation in the fifth-grade interventions) to the entire class for 10–15 min. After that, the students worked in pairs to learn to solve equations under their teacher's supervision. During the pair work, the paper-and-pencil groups used only the provided worksheet (Figure 7a), whereas the X-is groups used the X-is Wizard of Oz prototype in addition to the worksheet (Figure 7b).

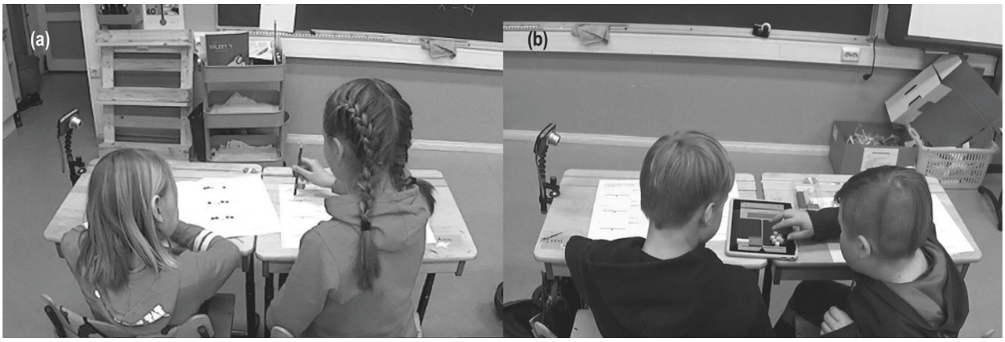


Figure 7. Pair work during the classroom interventions. (**a**) Paper-and-pencil condition; (**b**) X-is condition.

The worksheet was designed to promote students' understanding of equation concepts through translation and connections between multiple representations (i.e., verbal and written, pictorial, and mathematical symbolic) of equations. It was developed based on mathematics textbooks used in Finland and then evaluated by the third and fourth authors, who are experienced teacher educators. The worksheet contained eight equations presented in one of the following formats: $x + A = B$, $A + x = B$, $A = x + B$, or $A = B + x$, where A and B are positive integers and x is the unknown.

All 12 pair work sessions were video recorded from two different angles (530 min in total) to warrant well-captured data. The video materials were transcribed in terms of the students' actions (i.e., who did what, with whom). Due to the noisy background of the classrooms, the students' dialogues could not be transcribed. The pair work analysis particularly focused on how each instructional condition supported students' on-task peer communication (i.e., multi-representation translation, equation solving, and providing the unknown value) regarding directions and types. Communication directions and types were used as indicators of languaging and peer interaction, which encouraged learning through social interaction. Students' off-task communication (e.g., chitchatting, laughing, or cleaning up physical objects from the tablet screen before proceeding to the next exercise) and communication with the teacher were not included in the analysis. The video transcription was analysed using a qualitative deductive content analysis method [77]. A categorisation matrix was developed, and all the data were coded according to the categories in the matrix. Communication directions were categorised into one-way or two-way communication, while communication types were categorised into verbalisation, physical actions, or verbalisation and physical actions. Table 2 lists the coding categories with examples.

Table 2. Categories for communication directions and types.

Category	Description	Examples
Communication Directions		
One-way	Sending information through speaking, writing, or gestures without response from peer	• Speaking out loud while doing something • Asking for help, no peer response • Giving advice, no peer response
Two-way	Sending and receiving information through speaking, writing, or gestures	• Discussing or negotiating • Giving and taking advice/assistance • Manipulating X-is together
Communication Types		
Verbalisation	Communicating through speech	Asking or discussing
Physical actions	Communicating through gestures	Pointing or showing
Verbalisation and physical actions	Communicating through speech and gestures	Manipulating X-is and explaining at the same time

Then, the coded data were merged into communication episodes. One episode is a unit of complete actions for a specific purpose, for example, a discussion about how to solve an equation or the whole process of solving the equation. In total, 287 episodes of on-task peer communication were discovered. The frequency of individual episodes was counted (i.e., quantified) and divided into categories for descriptive statistical analysis. A Pearson's chi-squared test was performed to statistically investigate the relationship between instructional conditions and peer communication.

Additionally, the video data were analysed to seek evidence of support for each instructional condition for students' learning (i.e., understanding of equation-solving concepts, languaging, and learning through discovery and social interaction) and the usability of X-is. The discovered evidence was used to complement the findings from the teachers' questionnaires and interviews.

4.3.2. Students' Paper-Based Tests

The same 45-min paper-based test (Appendix A) was administered for all students to compare the learning achievement of students in languaging-based intervention conditions (the fourth and fifth graders in the paper-and-pencil and X-is groups: $n = 24$) to that of students in typical classrooms (the seventh through ninth graders: $n = 65$). The test was almost identical to the intervention worksheet. The only difference was that the test contained only six equations. Students could earn a maximum of three points for each equation for the correct representation translation, equation solving, and value of the unknown, for a total score of 18. Cronbach's alpha of the test was 0.82. As it is appropriate for small sample sizes, the Mann–Whitney U test was used to investigate the difference in students' learning achievement for languaging-based intervention groups (with or without manipulatives) and the comparison groups. The learning achievement difference between the two intervention groups (paper-and-pencil and X-is) and between each intervention group and the comparison groups was not statistically analysed due to the sample size being too small.

4.3.3. Students' Thinking Aloud

Thinking aloud [78] was conducted to assess the usability of X-is and its contribution to students' learning process and achievement. Only the students in the X-is groups ($n = 12$) were individually asked to model and then solve one to two equations and at the same time verbalise their actions, thoughts, and opinions. When students faced difficulties, they were provided with minimal hints and guidance to assist them in proceeding with the task.

All 12 thinking aloud sessions (18 min in total) were video recorded and transcribed. The video transcription was analysed using a qualitative inductive content analysis [77]. The data were open coded and then grouped into sub-categories, which were further grouped into categories and the main categories related to the research focus.

4.3.4. Student Questionnaires and Interviews

All the X-is students (n = 12) individually completed a questionnaire and participated in a face-to-face interview to evaluate the perceived usefulness and usability of X-is. A 4-point Likert-type scale (ranging from 1 = fully disagree to 4 = fully agree) was adapted from the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire [79], as it is concise and contains the usability dimensions that we investigated. We only used one item from each sub-scale of the USE questionnaire to build our questionnaire, since the selected items best describe each studied dimension. When completing the face-to-face questionnaire, we also verbally other listed items that belonged to the same usability dimension for the participants. The scale was comprised of four items assessing four factors: 'X-is is easy to use' (ease of use), 'It is pleasant to solve equations with X-is' (enjoyment), 'X-is helps me to understand how to solve equations' (usefulness), and 'I would like to solve equations with X-is' (intention for future use). Cronbach's alpha of the scale was 0.69. The interview was conducted at the same time as the questionnaire. The students were asked to give the reason for their response to each questionnaire item. They were also asked for suggestions regarding how X-is could be improved.

All 12 interviews (102 min in total) were video recorded and transcribed. Frequencies of the questionnaire responses were used to determine students' perceptions of X-is. The video transcription was analysed using a qualitative inductive content analysis method. The qualitative findings were used to complement the quantitative findings.

4.3.5. Teacher Questionnaires and Interviews

All teachers (n = 6) completed a questionnaire (Appendix B) and a face-to-face interview to assess how teachers perceived the utility of X-is compared to paper-and-pencil instruction (i.e., using worksheets or textbooks). Three 4-point Likert-type scales (ranging from 1 = not at all to 4 = very well) were designed specifically for evaluation of how X-is compares to paper-and-pencil instruction regarding support for:

- Students' understanding of equation-solving concepts (three items; X-is: $\alpha = 0.90$, paper-and-pencil instruction: $\alpha = 0.90$),
- Students' languaging (five items; X-is: $\alpha = 0.71$, paper-and-pencil instruction: $\alpha = 0.67$), and
- Students' learning through discovery and social interaction (three items; X-is: $\alpha = 0.53$, paper-and-pencil instruction: $\alpha = 0.81$).

Although, the Cronbach's alpha value ($\alpha = 0.53$) for the responses to students' learning through discovery and social interaction for X-is was below the generally acceptable level [80]. It is noteworthy that the low alpha value could be due to the small number of items ($N = 3$) in the scale and the small number of the participants ($N = 6$) [80]. Moreover, the Cronbach's alpha value ($\alpha = 0.81$) for the responses to the same scale for the paper-and-pencil instruction was acceptable [80]. Thus, the responses to the scale for X-is were used and interpreted with caution.

During the session, the teachers were required to respond to the questionnaire items as well as provide an explanation for each response. In addition, they were also asked about appropriate grade levels and learning-attaining levels in regard to working with X-is, suggestions for improving X-is, and differences between physical (TUI) and digital (GUI) block manipulation. All six interviews (175 min in total) were audio recorded and transcribed. Descriptive statistics (e.g., frequencies, cumulative sums, and cumulative means) of the questionnaire responses were used to present teachers' perceptions of X-is. A Wilcoxon matched-pairs signed rank test was conducted to determine the difference between X-is and paper-and-pencil instruction according to the teachers. The audio transcription was analysed using a qualitative inductive content analysis. The qualitative findings were used to complement the quantitative findings.

5. Results and Discussion

We next report and discuss the findings in three sections according to our research questions regarding how well X-is promotes learning achievement, learning support, and usability. Then, we reflect on the limitations of the research and give recommendations for future research.

5.1. Learning Achievement

5.1.1. Students' Paper-Based Tests

The Mann-Whitney U test was performed to determine the differences between the total test scores for the languaging-based instructional intervention (with paper-and-pencil or with X-is) groups and the comparison groups. Figure 8 illustrates that there was no significant difference in the test scores for the fourth and fifth graders who received the languaging-based instructional intervention ($M = 11.09$, $SD = 3.69$) compared to the seventh graders in the comparison group ($M = 12.29$, $SD = 5.28$; $U = 314$, $p = 0.10$). However, there was a statistically significant difference between the total test scores for the intervention groups compared to the eighth and ninth graders ($M = 15.57$, $SD = 2.40$; $U = 95$, $p < 0.001$) and for the seventh graders compared to the eighth and ninth graders ($U = 332.5$, $p = 0.01$). In addition, we examined students' low achievement on the test. Passing the test required a student to earn 50% of the maximum score (cut-off score = 9/18). A similar portion of the students in the intervention groups and the seventh graders failed the test (25% and 26%, respectively), whereas none of the eighth or ninth graders did.

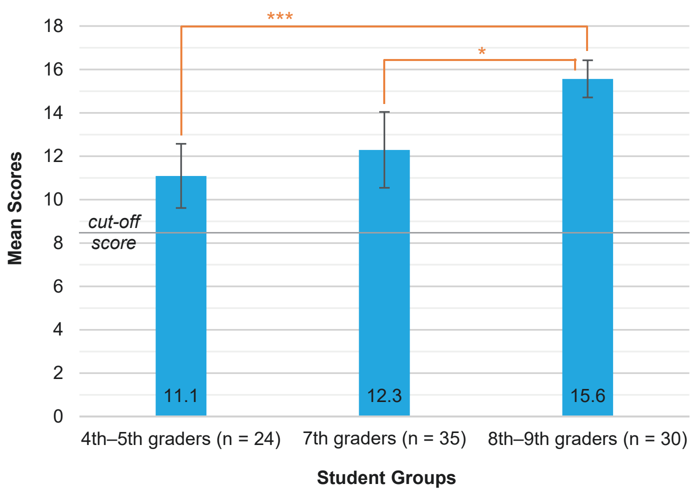


Figure 8. Mean total scores (out of 18) for the intervention and comparison groups. Error bars show 95% CI = Mean ± (1.96 SE). * $p < 0.05$, *** $p < 0.001$.

To determine the impact of each languaging-based instructional condition on students' learning achievement, we examined the average total test scores of the students in both intervention groups. Overall, the paper-and-pencil group ($M = 11.29$, $SD = 4.08$) and the X-is group ($M = 10.90$, $SD = 3.44$) presented a rather similar test performance. We also investigated the strategies that each intervention group used on the test when solving equations correctly to discover whether the instructional conditions influenced students' use of taught strategies (i.e., reasoning for the unknown in the fourth-grade intervention and doing the same operation on both sides in the fifth-grade intervention) vs other strategies. The analysis did not include any situations in which students arrived at the correct answer without providing an explanation or their steps for equation solving or in which their used strategies could not be identified. Figure 9 shows that the X-is group (31/41) was more likely to use the strategies taught during the interventions for solving equations correctly than the paper-and-pencil group (24/42). Moreover, two-thirds (8/13) of the X-is group used the taught strategies to solve equations correctly at least once, whereas only half (6/12) of the paper-and-pencil group did.

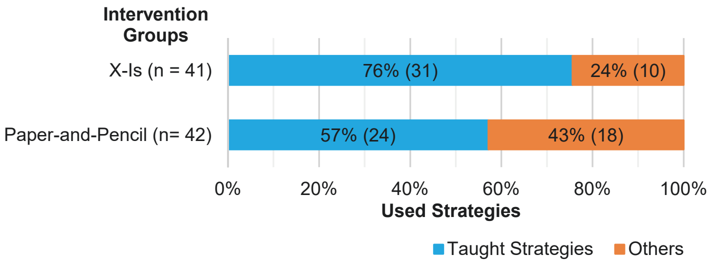


Figure 9. Percentage (number of cases in parentheses) of different strategies used to solve equations correctly on the test by instructional condition.

5.1.2. Students' Thinking Aloud Sessions

In addition to the test analysis, we analysed the thinking aloud transcriptions to examine learning achievement of the X-is group. First, we investigated the students' representational fluency (i.e., how well they could model the math symbolic equations with X-is physical objects), which is an indicator of their equation concept understanding. All students were able to use physical objects to model the given equations successfully. Ten students completed the equation modelling on their own in under 15 s, whereas two students required more time and some guidance from X-is or us. We further examined how well the students could solve the given equations with X-is. All of them were able to solve the equations correctly. Four of them also provided clear argumentation to support their equation-solving process, which indicated their understanding of the equivalent concept and equation-solving principles. For example, after almost 2 min of equation solving, a low-attaining fourth grader was able to explain how she solved the equation by reasoning for the unknown (Figure 10).

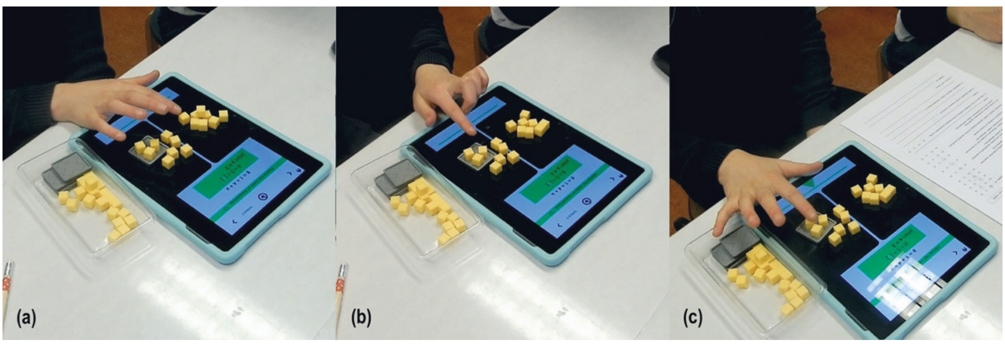


Figure 10. Student solving an equation ($8 = 1 + 4 + x$) at Level 1 (reasoning for the unknown). (**a**) 'Here are eight' (pointing at eight base-10 units on the left side of the scale); (**b**) 'And here has to weigh the same' (pointing at the right side of the scale); (**c**) 'You have to add [base-10 units] here, so that they [units on the right side] are altogether eight' (pointing at three base-10 units inside the X-Box on the right).

Figure 11 shows how a medium-attaining fifth grader verbalised how he solved the equation by doing the same operation on both sides of an equation step by step. The student provided reasons supporting his actions.

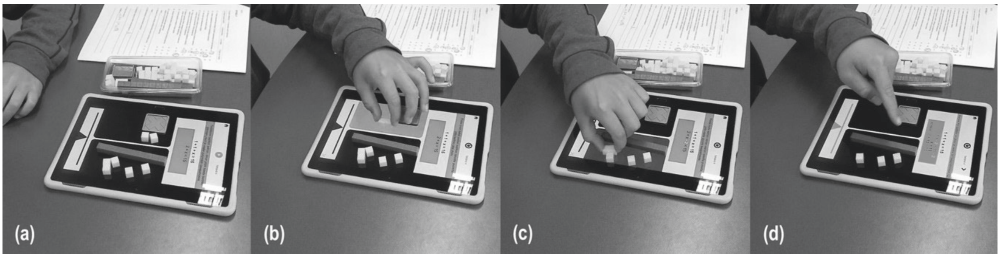


Figure 11. Student solving an equation (1 + 1 + x = 15) at Level 2 (doing the same operation on both sides of an equation). (**a**) Correct modelling of the equation; (**b**) 'You have to take two away from here, so that x will stay alone' (taking two base-10 units from the left side of the scale); (**c**) 'Then you have to take also two from here' (taking two base-10 units from the right side of the scale). 'Because if you take [blocks] from one side, then you have to take [blocks] also from the other side'; (**d**) 'So, here is 13, which is equal to x' (pointing at the rest of the blocks on the right).

Hight-attaining fourth and fifth graders used X-is to model and solve an equation correctly within 30 s. It is worth mentioning that the fourth grader demonstrated that he could reason for the unknown by himself without X-is. During the thinking aloud, he calculated the value of the unknown mentally before manipulating X-is.

5.1.3. Discussion of Learning Achievement

In summary, the paper-test result analysis demonstrate that languaging-based instruction (with paper and pencil or with X-is) had a significantly positive impact on equation learning achievement of the fourth and the fifth graders, who had received one intervention lesson, compared to the seven-grade comparison group, who had received approximately 10 normal equation-solving lessons. These quantitative findings suggest that the experimental groups benefited from the intervention instruction, thereby encouraging further development of languaging-based instruction and X-is to support learning equation solving.

The test performance of the X-is students illustrates students' embodied cognition in the absence of X-is, thereby indicating their independence from the manipulative [9]. A favourable impact of X-is on students' learning performance and understanding of equation solving was evidenced by (1) students' likelihood to use the strategies taught during the interventions to solve the equations correctly, (2) their ability to model and solve equations correctly during the thinking aloud sessions, and (3) their clear argumentation to support their equation-solving process. Particularly, their argumentation indicated that they had a good understanding of mathematical equivalence (e.g., the equal sign as a relational symbol) and the equation-solving process. The evidence that X-is is advantageous for students' learning achievement corroborates the findings of earlier work [14,45,81] in tangible technology-enhanced mathematics learning. Moreover, the findings from the thinking aloud data suggest that X-is facilitated the achievement of equation learning (i.e., modelling and solving equations successfully) among diverse students, particularly low and medium achievers. These findings are consistent with those of Pires et al. [14], who found that children with no proficiency in number combinations benefited more from their TM compared to their virtual manipulative.

5.2. Learning Support

5.2.1. Class Intervention

The video data reveal that, during the pair work sessions, the students mostly concentrated on completing their tasks. Off-task activities were rarely observed. Generally, fourth graders in both conditions could work out the problems on their own. The teacher's assistance was needed mainly for translation of the equations in word problems into pictures and mathematical symbols. For the fifth-grade interventions, paper-and-pencil students required more help from the teacher than their X-is peers. The teacher assisted paper-and-pencil students in solving equations by doing the same operation on both sides and translating the equations presented as word problems. X-is students mainly needed the teacher's support for the translation of the equations presented as word problems. Overall, paper-and-pencil students worked silently and separately on their own worksheets at different paces. Thinking aloud while writing, discussing, giving or asking advice, and looking at what the groupmate doing were occasionally observed. In contrast, the X-is group usually manipulated X-is together or took turns (i.e., alternately manipulated X-is and watched when their groupmate manipulated the manipulative). Thinking aloud, discussion, and giving and taking advice were usually observed. After manipulating X-is, each student silently recorded their equation-solving processes on their own worksheet. Most of them also looked at the math sentences on the tablet while recording their work.

Table 3 presents the frequencies and percentages of peer communication episodes regarding directions and types of communication observed during the pair work. It should be noted that all of the students participating in the interventions were unfamiliar with languaging, particularly verbal and written. From the total of 287 observed episodes, most of the peer communication happened during the X-is group's pair work (70%). The average numbers of peer communication episodes were 14 episodes/pair (*SD* = 7.7) for the paper-and-pencil group and 34 episodes/pair (*SD* = 13.9) for the X-is group. Regarding communication directions, most of the paper-and-pencil group communication (60%) was one-way, while most of the X-is group communication (70%) was two-way, indicating increased peer interaction among the latter. In terms of communication types, most of the paper-and-pencil group communication (73%) was verbal followed by verbal and physical (19%), while the verbal communication (48%) and verbal and physical communication (44%) occurred at similar rates for the X-is group. We further examined the types of communication used for each communication direction. Both instructional conditions used mainly verbal communication, particularly, thinking aloud during one-way communication (78% of paper-and-pencil group and 92% of X-is group). For the paper-and-pencil group, most of the two-way communication was verbal (65%) followed by verbal and physical (35%). In contrast, for the X-is group, most of the two-way communication was verbal and physical (63%) followed by verbal (30%). It should be noted that two-way physical communication (7%) was only observed in the fourth-grade X-is intervention group. This kind of communication happened when students used X-is to model or solve equations together without talking to each other. The Pearson's chi-squared test revealed that there were statistically significant associations for instructional conditions with students' peer communication directions [X^2 (1, N = 287) = 23.15, $p < 0.001$] and types [X^2 (2, N = 287) = 17.13, $p < 0.001$].

Table 3. Observed frequencies and percentages of peer communication episodes regarding directions and types by instructional condition.

Communication	Paper-and-Pencil	X-is	Total
	n (%)	*n* (%)	*n* (%)
Directions (*N* = 287)			
One-way	51 (17.8)	60 (20.9)	111 (38.7)
Two-way	34 (11.8)	142 (49.5)	176 (61.3)
Types (*N* = 287)			
Verbalisation	62 (21.6)	97 (33.8)	159 (55.4)
Physical actions	7 (2.4)	16 (5.6)	23 (8.0)
Verbalisation and physical actions	16 (5.6)	89 (31.0)	105 (36.6)

5.2.2. Student Questionnaires, Interviews, and Thinking Aloud Sessions

Most of the students (10/12) in the X-is intervention group felt that X-is assisted them in understanding equation solving because of its TUI (i.e., allowing physical input and providing multimodal output), scaffolding (e.g., the tilting scale and changing background colours of the math expression window), and the balance model used. Pointing at the math expression window, one medium-attaining fifth grader stated, 'From here, you can see what has been done. You can see also from the scale', (pointing at the unbalanced scale), 'whether there is too much or too little. So it helps'. According to a low-attaining fifth grader, 'Being able to put and move these blocks with my hands is better for me and somehow I understand better'.

It is noteworthy that only one high-attaining and one medium-attaining fourth grader disagreed on the supportiveness of X-is because they were able to solve equations without X-is. Moreover, the fifth graders had a higher level of agreement (five strongly agreed, one agreed, and none disagreed or strongly disagreed) than the fourth graders (one strongly agreed, three agreed, and two disagreed or strongly disagreed) regarding the learning support provided by X-is. A possible reason for this might be that the reasoning for the unknown strategy (Level 1) was rather easy for fourth graders, while the strategy for doing the same operation on both sides (Level 2) was appropriately challenging for fifth graders.

Some students found new ways to use X-is in addition to the original intention. Students who were able solve equations by themselves pointed out that they first solved an equation and recorded their solution on the worksheet, and after that, they checked their answer using math sentences in the math expression window on the tablet screen. Thus, the math sentences worked as their answer key instead of as step-by-step scaffolding. During the thinking aloud session, a medium-attaining fourth grader presented her own strategy for how to use X-is by first separating base-10 blocks into spatial groups to find the value of the unknown (Figure 12).

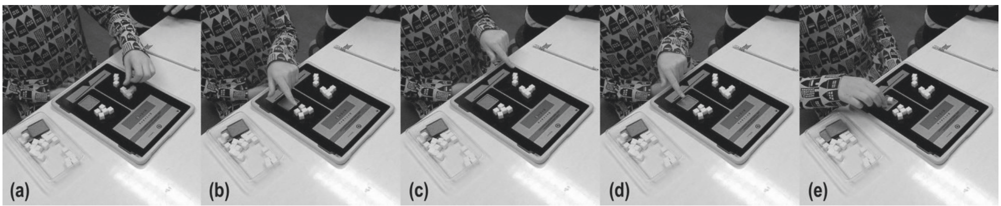


Figure 12. How one student solved an equation (8 = 1 + 4 + x) using her own invented strategy. (**a**) 'First, I move them here' (moving five base-10 units on the left side of the scale further away from three base-10 units on the same side); (**b**) 'So that they' (pointing at the five separate base-10 units on the left) 'are equal to what is here' (pointing at five base-10 units on the right); (**c**) 'Then I look at here' (pointing at three base-10 units on the bottom of the left).); (**d**) 'They are the same [number of units] that I have to put here' (pointing at the empty X-Box on the right); (**e**) The student then added three base-10 units to the X-Box.

5.2.3. Teacher Questionnaires and Interviews

Based on the questionnaire, the teachers clearly rated X-is as being better at supporting students' equation concept understanding, languaging, and learning through discovery and social interaction than paper-and-pencil instruction (Table 4). A Wilcoxon matched-pairs signed rank test also indicated that these differences were statistically significant ($Z = 2.10$, $p = 0.04$).

Similar to the students, all six teachers agreed that X-is promoted students' overall understanding of equation-solving concepts better than paper-and-pencil instruction. They pointed out that X-is concretises the equation concepts by allowing students to learn by doing and by providing scaffolding through the dynamic tilting scale and math expression window. For example, one lower secondary school mathematics teacher said, 'I think the manipulative supports students' understanding of equation concepts very well. It is very concrete. Students can try it with their own hands and see it with their own eyes. It is active learning'. The fifth grade class teacher also observed that during the interventions, students working with X-is concentrated more on the process of isolating x by doing the same operation on both sides of the scale instead of just calculating the value of the unknown in their heads.

Table 4. Teachers' ratings of the learning support of X-is compared to paper-and-pencil instruction.

Scales	Cumulative Sum Mean			Z
	Scale [min, max]	X-is	Paper-and-Pencil Instruction	
Supports students' understanding of equation-solving concepts	[3, 12]	11.0	7.7	2.10 *
Supports students' languaging	[5, 20]	17.3	12.8	2.10 *
Supports students' learning through discovery and social interaction	[3, 12]	10.5	6.5	2.10 *

Note: * $p < 0.05$.

All the teachers believed that X-is better encouraged students to express their mathematical thinking compared to paper-and-pencil instruction, particularly through physical and mathematic symbolic representations. They highlighted that physical block manipulation and math sentences in the math expression window contributed to these physical and symbol representations, respectively.

"*The manipulative is very action based and visual. I would say that these help students to explain [the concept] to peers. When students have solved it with their hands, it is easier for them to talk about [the process]. Textbooks and e-textbooks are also very visual, full of pictures and videos. However, textbook exercises usually urge students to move forward too fast instead of talking about the current exercise*."(Lower secondary school mathematics teacher).

The fifth-grade class teacher also noticed that during the class interventions, paper-and-pencil students were particularly silent and worked separately, even though he had encouraged them to work together and discuss (Figure 13a). In contrast, the X-is students were more active to discuss and required less encouragement for verbalisation (Figure 13b). The teacher felt that this was because the X-is students had to think about how to manipulate the physical blocks, thereby encouraging thinking aloud and discussion. He also observed that the X-is students learned mathematical symbolic representation from the math expression window. Therefore, they were able to write math sentences explaining their step-by-step solutions more clearly than the paper-and-pencil students.

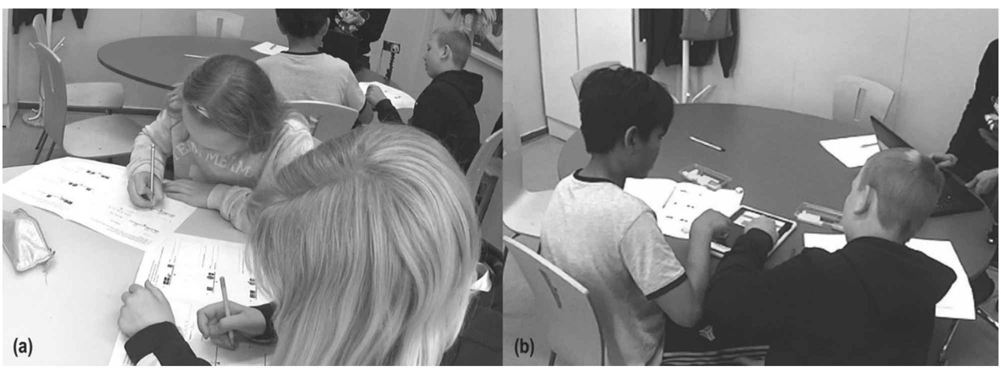


Figure 13. Fifth-graders' communication during the intervention. (**a**) Paper-and-pencil students solving an equation silently and separately; (**b**) X-is students discussing while solving an equation.

All the teachers agreed that X-is provided better encouragement for students' learning through discovery than paper-and-pencil instruction. In their opinion, because X-is works automatically and provides real-time feedback, it allowed students to experiment by themselves and learn independently at their own pace. The fifth-grade teacher noted that X-is students needed less assistance from him during the interventions compared to the paper-and-pencil students. The fourth-grade teacher had a similar finding:

> "*When students got stuck, they could not get through on their own with the worksheet. Whatever weights students added on the scale [image on the worksheet], the scale wouldn't move. So students might proceed with the wrong solution. But with the manipulative, students could add and remove blocks and got feedback from the manipulative*."

The teachers had mixed perceptions of how X-is supports students' learning through social interaction compared to paper-and-pencil instruction. Most of the teachers thought that X-is better encouraged students' social interaction by allowing them to share the same tablet, thus enhancing peer interaction naturally:

> "*When working with manipulatives, there are steps that the students can easily talk about. One student can tell another one, for example, 'First, put it there. Do you notice how the scale moves?' But when working with a textbook where there is, for example, $x + 2 = 6$, there is not much to discuss, only $x = 6 - 2$. It is difficult for students to invent what to talk about*."(Special education teacher).

However, two mathematics teachers had different views from the majority. They rated both X-is and paper-and-pencil instruction equally. They thought that the teacher and group dynamic played a more important role in peer interaction than the learning materials. Moreover, they felt that some students might try to take control of the manipulative without sharing it with others. Nevertheless, a few incidents of unbalanced participation occurred in the X-is interventions.

According to the teachers, X-is could be used to promote equation concept understanding among pre-primary through ninth-grade students. They believed that Level 1 would be suitable for pre-primary through fourth-grade students, while Level 2 would be suitable for fifth graders onwards. All teachers indicated that they would use X-is with students of all attainment levels in a small group to introduce equation concepts at the beginning of their class. Later on, X-is could be used for differentiation. High achievers could do more challenging exercises with X-is on their own, whereas students who had difficulty with equation solving could use X-is as a recap of what had been taught.

Four teachers were satisfied with X-is as it was. Two teachers provided suggestions for improving the pedagogical benefits of X-is. Instead of the letter x, the unknown could be represented with symbols, pictures, or various letters to emphasise that anything could be used to represent the unknown. Furthermore, after students have learnt how to write math sentences from the math expression window, there could be exercises in which they write math sentences on their own.

When asked about the difference between TUI and GUI—manipulation of physical and digital blocks, respectively—all teachers preferred physical block manipulation. They argued that manipulating physical blocks supports students' learning by linking their actions with their thinking:

> "*When everything is digital, students may perceive it as a game. So, they will act like [they do when] playing games [and] just rush to do everything. A good example is when I asked students to learn how to draw points in GeoGebra [an interactive mathematics application]. Some students just kept on clicking [their mouse], so that their screen was full of points. Their minds were in a racing track. I think physical blocks could slow them down to think*."(Lower secondary school mathematics teacher).

Moreover, physical block manipulation is likely to interest students more because the digital world is too familiar for them:

> "*I think physical blocks are definitely better than digital blocks. They are more interesting for students because nowadays, they have been doing things all the time with the digital world. I noticed that*

last semester, [lower secondary school] students were very enthusiastic about playing board games during a math class. I don't see working with physical objects as too childish for lower secondary school students."(Lower secondary school mathematics teacher).

5.2.4. Discussion of Learning Support

With respect to understanding equation-solving concepts, our findings are in line with those of other studies [14,43,45] that found that TUIs are likely to assist students in developing their understanding of mathematical concepts. We found that X-is could benefit diverse-attaining pre-primary through ninth-grade students in understanding equation concepts. Our observations suggest that the unique attributes of X-is, which combines the benefits of physical and virtual manipulatives, might contribute to students' conceptual understanding. First, similar to our initial research [36], physical interactions of grasping and moving base-10 blocks and X-Boxes helped students to concretely develop their understanding of the different terms in an equation and equation-solving concepts. Second, according to Moyer-Packenham and Westenskow [12], the focused constraint of virtual environments promotes students' learning in mathematics by constraining and focusing their attention on certain mathematical objects and processes. In the current study, the scale's tilt and the math expression window's changing colours drew students' attention to the mathematical equivalence concept. Moreover, the GUI provided simultaneous linking [12], that is, linking the visual and mathematical symbolic representations of an equation. Third, tangible technologies concurrently connect multimodal representations (i.e., physical, visual, and symbolic representations) of the same concept. The TUI explicitly bridged students' actions and the effects of their actions in different forms, which not only helped students to perceive the relationship between concrete and abstract representations of equation equivalence but also to present their equation-solving processes with mathematical symbols. One anticipated finding was that some students discovered their own way to utilise X-is. This finding corresponds to Clements' [82] definition of a good manipulative, which is one that allows the learner to control and use it flexibly. Our observation of creating spatial groups of base-10 blocks to find the value of the unknown supports the findings from our initial research [36] and that of others [13,14,83] that physical objects enable students to flexibly move them, thereby exploring more various solutions compared to static pictorial representations on a paper. These findings can possibly be explained using the theory of DPL [7].

Regarding languaging, during the interventions, X-is students not only communicated with their pairs more frequently but also in more varied ways—mainly through speech or gesture and partly through a combination of both—compared to the paper-and-pencil group. The association between instructional conditions and students' peer communication types was also statistically significant. Moreover, all the teachers stated that X-is promoted students' expression of their mathematical thinking through physical and mathematical symbolic representations better than paper-and-pencil instruction. The integration of the physical and digital worlds of X-is contributed to students' various modes of meaning making. Similar to what we found in our initial research [36], we discovered that the manipulation of physical objects promoted physical and verbal representations, whereas the math sentences provided on the screen facilitated students' mathematical symbolic representation. Taken together, these observations suggest that X-is is likely to encourage students to multimodally make meaning from equation-solving concepts and procedures, thereby encouraging their languaging [30], which leads to better conceptual understanding. Additionally, students' languaging made it easier for not only the teachers but also the researchers to evaluate their thinking.

In terms of learning through social interaction, the findings indicate that X-is evidently encouraged peer interactions more than paper-and-pencil instruction. During the intervention, the paper-and-pencil students mainly worked silently and separately. In contrast, the X-is students consistently manipulated X-is together while simultaneously discussing or giving and taking advice from each other, thereby indicating social interaction in learning [31]. This study confirms prior research [43,45] that tangible technologies enhance peer interaction in mathematics classrooms. Most of the teachers believed that X-is would promote social interaction because it allowed students to share and manipulate simultaneously. The teachers' views support evidence from previous observations [84,85] that shared interfaces (i.e., in our case, physical objects and a tablet) encourage equal participation, which indicates that TUI may promote active collaboration. According to Pontual Falcão and Price [85], the active collaboration observed in our study was probably influenced by three main factors. First, multiple physical objects provided both students in the pair with the opportunity to gain access to them. Second, X-is allowed multiple inputs concurrently, thereby enabling both students to manipulate X-is parallelly. Third, as proposed by Price [16], the co-located design of X-is could resolve the concern of the two mathematics teachers about single-user constraints—one student taking control of the manipulative—which occurred during some of the X-is interventions. Because the input and output occurred at the same point of interaction, the students' attention was drawn to the tablet. Therefore, students could easily see each other's actions through the GUI on the shared screen, which contributes to their learning through social interaction.

Regarding learning through discovery, our results are in line with earlier research [45] that found that TUIs encouraged students to independently experiment and discover to-be-learnt mathematics content. According to the findings from our initial research [36] and those of others [11], both physical and digital properties of X-is could contribute to students' self-discovery. When manipulating physical objects, students tended to speak aloud what they were doing or thinking. Therefore, their languaging might not only help them to organise their own mathematical thinking but also allow their peers to listen to and reflect on their thinking [31], thereby building knowledge together. Further, GUIs provide students with guidance and real-time feedback. This scaffolding allows them to explore equation modelling and solving independently.

5.3. Usability

5.3.1. Class Intervention

Both groups of students completed all eight equations within a similar time frame (paper-and-pencil group: M = 21.1 min/pair, SD = 4.8; X-is group: M = 23.1 min/pair, SD = 2.8). Originally, the class teachers were supposed to guide the students on how to use X-is at the beginning of the intervention in the same way that they would in normal classes when working with new manipulatives. However, while the teachers were occupied with the paper-and-pencil students, the X-is students started to learn how to use X-is by themselves. After watching a 1 min introductory animation on the tablet, two pairs out of six were able to use X-is to solve all equations on their own. Four other pairs needed our minimal guidance for the first equation and after that were able to solve the rest by themselves. Excluding the time needed to model and solve the first equation, in which students had to learn how to use X-is, the task completion time of both conditions was almost the same (paper-and-pencil group: M = 2.6 min/equation, SD = 0.6; X-is group: M = 2.7 min/equation, SD = 0.3). For the X-is group, their task completion time included X-is manipulation and recording the solution on the worksheet.

5.3.2. Student Questionnaires, Interviews, and Thinking Aloud Sessions

All 12 students in the X-is group responded positively regarding the usability of X-is. They perceived X-is as easy and enjoyable to use. Moreover, all of them expressed their intention to use X-is for solving equations in the future. According to the students, the straightforward UI and guidance as well as their familiarity with tablets contributed to its ease of use. For example, one high-attaining fourth grader stated that 'It is not that difficult to use. You just have to put blocks and press an arrow'. Another student stated, 'It is easy to use because I have used a lot of computers and things like. I think this [X-is] is nicer than [solving equations on] paper'. (Medium-attaining fifth grader).

All students (eight who strongly agreed and four who agreed) expressed their enjoyment using X-is for various reasons. Some enjoyed its ease of use, while others appreciated its pedagogical benefits. Some mentioned that it was pleasant that X-is allowed them to work in pairs. A low-attaining fifth grader stated 'I am able more or less to solve equations [using X-is] and start to be interested in solving equations. That's why it is nice . . . But if it isn't nice, and I can't do it, then I usually give up'. Likewise, a medium-attaining fourth graders stated, 'It was quite fun, because you could do it with your friend'.

Moreover, some students emphasised that the differences between X-is and typical learning materials (e.g., paper and pencil) and its technological aspects made X-is enjoyable. For example, 'It is quite fun using this because you can move these blocks. Writing on the paper is quite boring' (Low-attaining fourth grader). According to a medium-attaining fourth grader, 'Nowadays, it is a must for [those of] us who are at my age to spend time using digital devices. It is nice to have a chance to work with digital devices. I enjoy using computers also at home'.

When asked whether they would like to use X-is again when solving equations, all the students—including the two fourth graders who disagreed with the statement about its effectiveness in terms of learning support—affirmed their intention for future use of X-is. Different pedagogical benefits (e.g., support for understanding and learning through discovery, scaffolding, and guidance) of X-is were the main reasons for the students' positive responses. Other reasons—such as enjoyability, ease of use, and technological aspects—were also mentioned:

> "*It would be nice if you could have this kind of app for other maths, even from the first grade when learning, for example, addition, subtraction, and multiplication. It's good to have these blocks compared to just digital [elements].*"(Medium-attaining fifth grader).

During the thinking aloud sessions, most of the students (9/12) were able to use X-is to model the given equation with physical objects and then solve it within 35–40 s without any difficulty. Only two fourth graders and one fifth grader struggled with how to use X-is. It took them about 2–5 min to finish the same task. This result may be explained by the fact that during the class interventions, these students were somewhat dominated by their pairs, thereby not having a chance to properly learn how X-is functioned. However, after receiving guidance from us or looking at instructions and prompts provided by X-is, they were eventually able to finish the task. Moreover, they could complete another similar task within one minute. During the thinking aloud sessions, some students also benefited from the guidance and scaffolding provided by X-is. For example, textual instruction guided students on how to proceed, while the status of the scale and the blinking working zone prompted students to take correct actions.

Nevertheless, problematic usability issues were identified. The thinking aloud sessions revealed a minor hurdle with the X-is navigation and operation menu. After modelling an equation, about half of the students (5/12) attempted to proceed to the equation solving part without pressing the play button. However, most of the students (8/12) expressed their satisfaction with the current usability of X-is and did not provide any suggestion for improvement. Some students made suggestions for how the usability could be developed, for example, by providing clearer, step-by-step instructions; adding an info button for detailed instructions; and offering an on–off math expression display for hiding or displaying the math expression when needed.

5.3.3. Discussion of Usability

To sum up, the findings demonstrate that the usability of X-is was clearly acceptable. Because the X-is students did not require much time or effort to learn to work with X-is during the interventions, learnability [86] (i.e., ease of learning [79]) of X-is was satisfactory. The efficiency [86] of X-is usability was rather high as students could quickly model and solve equations during the interventions and thinking aloud sessions. There was no evidence that they required substantial additional time to complete the intervention tasks (i.e., manipulating X-is and recording their work on the worksheet) compared to the paper-and-pencil condition. This observation supports the work of Martin et al. [83] but contrasts that of Uttal et al. [10]. The students also rated X-is as easy to use. Manches and O'Malley [15] proposed that one benefit of TMs appears to be that children do not need to learn how to manipulate physical objects. In our case, base-10 blocks were likely already familiar to the students, which may have also contributed to the ease of use of X-is. This finding partly supports the work of Sapounidis and Demetriadis [84], who reported that children ages 5–8 found interaction with TUI to be easier than with GUI, while children ages 11–12 felt the opposite.

Despite its high learnability, efficiency, and ease of use, some minor usability problems were found. The improvement of these problems would increase the usability of X-is. Regarding satisfaction [86], all students perceived X-is as enjoyable because of its pedagogical and social benefits as well as integration of digital technology. Additionally, all students stated that they intended to use X-is again. Similar findings were also reported in previous studies [44,84] on tangible technologies for learning.

5.4. Limitations and Future Research and Development

The limitations of the current study should be highlighted. First, only the post-test was conducted after the class intervention, and the comparison group was from the higher grade levels. Research employing pre-post test design and using the comparison and experimental groups from the same grade level would increase the precision of the students' learning achievement analysis and decrease the ambiguity of the result interpretation.

Second, the quantitative results must be interpreted with caution due to the small convenience sample size. Future research using larger random samples would contribute to external validity. Moreover, the interventions were conducted in a short period of time, which might not have been long enough to influence students' learning achievement but was still short enough to create a novelty effect. This shortcoming calls for longitudinal studies, which would demonstrate the long-term benefits of the manipulative for students and ensure that positive findings towards the manipulative are not a result of its novelty.

Third, there were some issues regarding research reliability. The internal consistency (i.e., Cronbach's alpha) of some of the questionnaire scales was relatively low. More items within the scales would increase the instrument reliability. Additionally, the qualitative analyses were conducted mainly by the first author. Future research should employ a triangulation of researchers. Despite the lack of researcher triangulation, the quality of the current mixed-methods research was promoted by the triangulation and concurrence of the research methods, data source, and data analysis as well as the integration of quantitative and qualitative findings.

Fourth, our study focused only on an overview of the potentials of tangible technologies for learning linear equations. Future research could investigate specific aspects examined in the current study in greater depth. Research focusing on students with a specific degree of achievement, other educational levels, different mathematics contents, and distance learning could contribute to our better understanding of how tangible technologies facilitate learning.

Fifth, it should be noted that off-shelf technology was employed for the proposed manipulative because our research was aimed at deployment in present classroom contexts. Thus, future studies on TMs employing purposely designed technologies might yield different results. Moreover, the limitations of the Wizard of Oz prototype, such as its delayed output, might have influenced the research results. A reliably working, fully interactive prototype with all functions and features would not only prevent this possibility but also expand the to-be-evaluated attributes of the manipulative. Evidently, our prototype was limited by the available off-shelf technology. The main challenge is of the unreliability of the object tracking using a webcam, especially when the colour contrast or brightness is too low or physical objects are too close to each other. Further software and hardware development and testing in a real classroom environment could enhance the system's reliability. It is possible that future advancements in tangible technology may enable the system to directly detect objects without requiring external devices and connections, thereby increasing product usability, practicality, and reliability. Additional features (e.g., free experiments with the balance scale, students' own equation set-up, equations generated based on students' performance, scoreboard, and analytics data) could enhance the user experience for X-is.

6. Conclusions

This study explored the potentials of tangible technologies for learning linear equations in real classrooms. Taken together, the findings add to the growing body of research indicating that there are advantages to utilising tangible technologies in mathematics classrooms. The initial TM is likely to benefit pre-primary through ninth-grade students of different attainment levels. It not only supported their conceptual understanding, languaging, and learning through discovery and social interaction but also enhanced their learning achievement. Moreover, the usability evaluation results demonstrated that the manipulative was learnable, easy to use, useful, and engaging, which would ensure its successful adoption in real educational contexts. The empirical evidences suggest that the integration of physical and virtual attributes of manipulatives is likely to contribute to these positive findings.

Our research results have important implications for mathematical classroom pedagogy and the development of TMs. Regarding classroom pedagogy, the study demonstrated that TMs should be adopted into mathematics classrooms—which are usually dominated by paper-and-pencil instruction [87,88]—to better assist diverse learners. However, it is worth noting here that TMs alone cannot contribute to mathematics learning [13,46,47]. To be beneficial, they should be used in cooperation with appropriate pedagogy, in our case, languaging. Moreover, the tangible attributes (i.e., physical–digital interactions) of the proposed manipulative, which were perceived as useful and engaging, may encourage the use of manipulatives in the upper-grade classrooms, in which use of manipulatives normally declines [61,89,90]. TM design and development could also benefit from the current work by taking advantage of the unique features of TUIs to facilitate mathematics learning. Additionally, continued efforts are needed to make tangible technologies accessible for all classrooms.

Supplementary Materials: The following are available online at http://www.mdpi.com/2414-4088/4/4/77/s1. Video S1: Level 1 equation solving during one fourth-grade thinking aloud session. Video S2: Level 2 equation solving during one fifth-grade thinking aloud session.

Author Contributions: Conceptualisation, D.L.; software, L.M.; methodology, investigation, and formal analysis, D.L.; writing—original draft preparation, D.L. and L.M.; writing—review and editing, D.L., L.M., J.J., and P.P.; visualisation, D.L. and L.M.; supervision, J.J. and P.P.; project administration, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

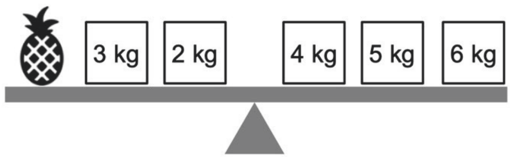
Acknowledgments: The authors would like to acknowledge the contributions of Fouzia Khan, Juho Korkala, Roni Perälä, Niko Sainio, Krishna Bagale, and Pekka Mäkiaho for software development; Tapio Lehtonen for the UI; and Kari Kouhian for the prototyping. The authors appreciate the teachers and the students who participated in the research.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The paper-based test contains six open-response items. The students were required to (1) translate the six equations presented through natural, pictorial, or mathematical symbolic language into two other representations, (2) verbally explain or show the mathematical steps they used to solve the equations, and (3) provide the value of unknowns. Figure A1 demonstrates the three types of the test items.

1. (a) Explain in your own words what the given picture means. (b) Form an equation from the given picture. (c) Show the steps you follow to determine how much one fruit weighs when the scale is balanced. (d) Provide the correct answer.



2. (a) Draw pictures on the balance scale according to the given equation. (b) Explain in your own words what the given equation means. (c) Show the steps you follow to determine the values of the unknown x that make the equation true. (d) Provide the correct answer.

   $x + 1 + 3 = 11$

3. (a) Draw pictures on the balance scale according to the given word question. (b) Form an equation from the given word question. (c) Show the steps you follow to determine the values of the unknown x that make the equation true. (d) Provide the correct answer.

   When mother weighs the cake ingredients, she notices that 5 g of butter and 12 g of flour are as heavy as one egg and 9 g of sugar. How much does one egg weigh?

Figure A1. The three types of the paper-based test items.

Appendix B

The teacher questionnaire contains the following scales:

1. In your opinion, how well did/would the X-is compared to the paper-and-pencil working method help students with understanding the following equation-solving concepts?
   - 1.1 Both sides of an equation are equal
   - 1.2 An unknown and solving for its value
   - 1.3 An equation stays equivalent when the same operation is performed on both sides
2. In your opinion, how well did/would X-is compared to the paper-and-pencil working method help students with expressing their mathematical thinking by using the following mediums?
   - 2.1 Tactile language
   - 2.2 Pictorial language
   - 2.3 Verbal natural language
   - 2.4 Written natural language

2.5 Mathematical symbolic language

3. In your opinion, how well did/would X-is compared to the paper-and-pencil working method support the following aspects for the students?

    3.1 Learning through first-hand experience and exploration
    3.2 Learning through collaboration with peers
    3.3 Active learning

References

1. Groves, S. Developing mathematical proficiency. *J. Sci. Math. Educ. Southeast Asia* **2012**, *35*, 119–145.
2. Kilpatrick, J.; Swafford, J.; Findell, B. *Adding It Up: Helping Children Learn Mathematics*; National Academy Press: Washington, DC, USA, 2001; ISBN 0-309-50524-0.
3. Brownell, W.A. The progressive nature of learning in mathematics. *Math. Teach.* **2007**, *100*, 26–34.
4. Figueira-Sampaio, A.; dos Santos, E.E.F.; Carrijo, G.A. A constructivist computational tool to assist in learning primary school mathematical equations. *Comput. Educ.* **2009**, *53*, 484–492. [CrossRef]
5. Fyfe, E.R.; McNeil, N.M.; Borjas, S. Benefits of "concreteness fading" for children's mathematics understanding. *Learn. Instr.* **2015**, *35*, 104–120. [CrossRef]
6. Bruner, J.S. *Toward a Theory of Instruction*; Harvard University Press: Cambridge, MA, USA, 1966; ISBN 9780674897014.
7. Martin, T.; Schwartz, D.L. Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cogn. Sci.* **2005**, *29*, 587–625. [CrossRef]
8. McNeil, N.; Jarvin, L. When theories don't add up: Disentangling the manipulatives debate. *Theory Pract.* **2007**, *46*, 309–316. [CrossRef]
9. Pouw, W.T.; van Gog, T.; Paas, F. An embedded and embodied cognition review of instructional manipulatives. *Educ. Psychol. Rev.* **2014**, *26*, 51–72. [CrossRef]
10. Uttal, D.H.; Amaya, M.; del Rosario Maita, M.; Hand, L.L.; Cohen, C.A.; O'Doherty, K.; DeLoache, J.S. It works both ways: Transfer difficulties between manipulatives and written subtraction solutions. *Child Dev. Res.* **2013**, *2013*, 216367:1–216367:13. [CrossRef]
11. Suh, J.; Moyer, P.S. Developing students' representational fluency using virtual and physical algebra balances. *J. Comput. Math. Sci. Teach.* **2007**, *26*, 155–173.
12. Moyer-Packenham, P.; Westenskow, A. Effects of virtual manipulatives on student achievement and mathematics learning. *Int. J. Virtual Pers. Learn. Environ.* **2013**, *4*, 35–50. [CrossRef]
13. Manches, A.; O'Malley, C.; Benford, S. The role of physical representations in solving number problems: A comparison of young children's use of physical and virtual materials. *Comput. Educ.* **2010**, *54*, 622–640. [CrossRef]
14. Pires, A.C.; González Perilli, F.; Bakała, E.; Fleisher, B.; Sansone, G.; Marichal, S. Building blocks of mathematical learning: Virtual and tangible manipulatives lead to different strategies in number composition. *Front. Educ.* **2019**, *4*, 81:1–81:11. [CrossRef]
15. Manches, A.; O'Malley, C. Tangibles for learning: A representational analysis of physical manipulation. *Pers. Ubiquitous Comput.* **2012**, *16*, 405–419. [CrossRef]
16. Price, S. Tangibles: Technologies and interaction for learning. In *The SAGE Handbook of Digital Technology Research*; Price, S., Jewitt, C., Brown, B., Eds.; SAGE: Los Angeles, CA, USA, 2013; pp. 307–325. ISBN 9781446200476.
17. Borenson and Associates, Inc. Hands-on Equations and Developing Fractions Sense Products. Available online: http://www.borenson.com/Products (accessed on 2 August 2020).
18. Borenson and Associates, Inc. Hands-on Equations Program on IOS, Androids, Kindle, & Windows Apps. Available online: https://www.borenson.com/Products/Mobile-Apps (accessed on 2 August 2020).
19. NLVM. Algebra Balance Scale. Available online: http://nlvm.usu.edu/en/nav/frames_asid_201_g_3_t_2.html?open=instructions&from=topic_t_2.html (accessed on 2 August 2020).
20. Reinschlüssel, A.; Alexandrovsky, D.; Döring, T.; Kraft, A.; Braukmüller, M.; Janßen, T.; Reid, D.; Vallejo, E.; Bikner-Ahsbahs, A.; Malaka, R. Multimodal algebra learning: From math manipulatives to tangible user interfaces. *i-com* **2018**, *17*, 201–209. [CrossRef]

21. McNeil, N.M.; Hornburg, C.B.; Devlin, B.L.; Carrazza, C.; McKeever, M.O. Consequences of individual differences in children's formal understanding of mathematical equivalence. *Child Dev.* **2019**, *90*, 940–956. [CrossRef]
22. Poon, K.; Leung, C. Pilot study on algebra learning among junior secondary students. *Int. J. Math. Educ. Sci. Technol.* **2010**, *41*, 49–62. [CrossRef]
23. Otten, M.; van den Heuvel-Panhuizen, M.; Veldhuis, M. The balance model for teaching linear equations: A systematic literature review. *Int. J. STEM Educ.* **2019**, *6*, 1–21. [CrossRef]
24. Lemke, J.L. Mathematics in the middle: Measure, picture, gesture, sign, and word. In *Educational Perspectives on Mathematics as Semiosis: From Thinking to Interpreting to Knowing*; Anderson, M., Sáenz-Ludlow, A., Zellweger, S., Cifarelli, V.V., Eds.; Legas: Brooklyn, NY, USA, 2003; pp. 215–234. ISBN 1894508394.
25. O'Halloran, K.L. The language of learning mathematics: A multimodal perspective. *J. Math. Behav.* **2015**, *40*, 63–74. [CrossRef]
26. Schleppegrell, M.J. Language in mathematics teaching and learning: A research review. In *Language and Mathematics Education: Multiple Perspectives and Directions for Research*; Moschkovich, J.N., Ed.; Information Age Pub: Charlotte, NC, USA, 2010; pp. 73–112. ISBN 1617351598.
27. O'Halloran, K.L. Mathematics as multimodal semiosis. In *Mathematics, Substance and Surmise: Views on the Meaning and Ontology of Mathematics*, 1st ed.; Davis, E., Davis, P.J., Eds.; Springer: Cham, Switzerland, 2015; pp. 287–303. ISBN 978-3-319-21473-3.
28. Morgan, C. The place of pupil writing in learning, teaching and assessing mathematics. In *Issues in Mathematics Teaching*; Gates, P., Ed.; Routledge Falmer: London, UK, 2001; pp. 232–244. ISBN 0415238641.
29. O'Halloran, K.L. *Mathematical Discourse: Language, Symbolism and Visual Images*; Continuum: London, UK, 2004; ISBN 9780826468574.
30. Joutsenlahti, J.; Kulju, P. Multimodal languaging as a pedagogical model: A case study of the concept of division in school mathematics. *Educ. Sci.* **2017**, *7*, 9. [CrossRef]
31. Joutsenlahti, J.; Rättyä, K. Kielentämisen käsite ainedidaktisissa tutkimuksissa [The concept of languaging in subject-didactic studies]. In *Rajaton Tulevaisuus. Kohti Kokonaisvaltaista Oppimista [Unlimited Future. Towards Holistic Learning]*; Kauppinen, M., Rautiainen, M., Tarnanen, M., Eds.; Finnish Research Association for Subject Didactics: Jyväskylä, Finland, 2015; pp. 45–61. ISBN 1799-960X.
32. Lesh, R.A.; Landau, M.; Hamilton, E. Conceptual models in applied mathematical problem solving research. In *Acquisition of Mathematics Concepts and Processes*; Lesh, R.A., Landau, M., Eds.; Academic Press: New York, NY, USA, 1983; pp. 263–343. ISBN 012444220X.
33. Bauersfeld, H. Language Games in the Mathematics Classroom: Their Function and Their Effects. In *The Emergence of Mathematical Meaning: Interaction in Classroom Cultures*; Cobb, P., Bauersfeld, H., Eds.; Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ, USA, 1995; pp. 271–294.
34. Goldin, G.; Shteingold, N. Systems of representations and the development of mathematical concepts. In *The Roles of Representation in School Mathematics*; Cuoco, A., Curcio, F.R., Eds.; National Council of Teachers of Mathematics: Reston, VA, USA, 2001; pp. 1–23. ISBN 0873534956.
35. Ceuppens, S.; Deprez, J.; Dehaene, W.; De Cock, M. Design and validation of a test for representational fluency of 9th grade students in physics and mathematics: The case of linear functions. *Phys. Rev. Phys. Educ. Res.* **2018**, *14*, 020105:1–020105:19. [CrossRef]
36. Lehtonen, D.; Joutsenlahti, J. The benefits of using manipulatives in classrooms for equation concepts understanding. In *Changing Subjects, Changing Pedagogies: Diversities in School and Education*; Pyyry, N., Tainio, L., Juuti, K., Vasquez, R., Paananen, M., Eds.; Finnish Research Association for Subject Didactics: Helsinki, Finland, 2017; pp. 164–185. ISBN 978-952-5993-23-3.
37. Ishii, H.; Ullmer, B. Tangible bits: Towards seamless interfaces between people, bits and atoms. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 22–27 March 1997; pp. 234–241.
38. Ishii, H.; Ullmer, B. Tangible user interfaces. In *The Human-Computer Interaction Handbook Fundamentals, Evolving Technologies, and Emerging Applications*, 3rd ed.; Jacko, J.A., Ed.; Taylor & Francis: Boca Raton, FL, USA, 2012; pp. 465–490. ISBN 9781439829431.
39. Zhou, Y.; Wang, M. Tangible user interfaces in learning and education. In *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed.; Wright, J.D., Ed.; Elsevier: Oxford, UK, 2015; pp. 20–25. ISBN 978-0-08-097087-5.

40. O'Malley, C.; Fraser, D.S. *Literature Review in Learning with Tangible Technologies*; Report 12; Futurelab: Bristol, UK, 2004; pp. 1–48. ISBN 0-9548594-2-1.
41. Alibali, M.W.; Nathan, M.J. Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *J. Learn. Sci.* **2012**, *21*, 247–286. [CrossRef]
42. Pontual Falcão, T.; Meira, L.; Gomes, A.S. Designing tangible interfaces for mathematics learning in elementary school. In Proceedings of the 3rd Latin America Conference on Human–Computer Interaction (CLIHC 2007), Rio de Janeiro, Brazil, 10–11 September 2007.
43. Starcic, A.I.; Cotic, M.; Zajc, M. Design-based research on the use of a tangible user interface for geometry teaching in an inclusive classroom. *Br. J. Educ. Technol.* **2013**, *44*, 729–744. [CrossRef]
44. Salvador, G.; Pérez, D.; Ortega, M.; Soto, E.; Alcañiz, M.; Contero, M. Evaluation of an augmented reality enhanced tabletop system as a collaborative learning tool: A case study on mathematics at the primary school. In Proceedings of the 33rd Annual Conference of the European Association for Computer Graphics (Eurographics 2012), Cagliari, Italy, 13–18 May 2012; pp. 9–16.
45. Zamorano Urrutia, F.J.; Cortés Loyola, C.; Herrera Marín, M. A tangible user interface to facilitate learning of trigonometry. *Int. J. Emerg. Technol. Learn.* **2019**, *14*, 152–164. [CrossRef]
46. Ball, D.L. Magical hopes: Manipulatives and the reform of math education. *Am. Educ.* **1992**, *16*, 14–18.
47. Baroody, A. Manipulatives don't come with guarantees. *Arith. Teach.* **1989**, *37*, 4–5.
48. Bruner, J.S. The act of discovery. *Harv. Educ. Rev.* **1961**, *31*, 21–32.
49. Neber, H. Discovery learning. In *Encyclopedia of the Sciences of Learning*, 1st ed.; Seel, N., Ed.; Springer: Berlin, Germany, 2012; pp. 1009–1012. [CrossRef]
50. Vygotsky, L.S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press: London, UK, 1978; ISBN 0674576292.
51. Tam, M. Constructivism, instructional design, and technology: Implications for transforming distance learning. *Educ. Technol. Soc.* **2000**, *3*, 50–60.
52. Brooks, J.G.; Brooks, M.G. In Search of Understanding: The Case for Constructivist Classrooms, with A New Introduction by the Authors. Association for Supervision and Curriculum Development: Alexandria, VA, USA, 1999; ISBN 1-4166-0283-6.
53. Slavin, R.E. Cooperative learning. In *International Encyclopedia of Education*, 3rd ed.; Peterson, P., Baker, E., McGaw, B., Eds.; Elsevier: Oxford, UK, 2010; pp. 177–183. ISBN 978-0-08-044894-7.
54. Devin-Sheehan, L.; Feldman, R.; Allen, V. Research on children tutoring children: A critical review. *Rev. Educ. Res.* **1976**, *46*, 355–385. [CrossRef]
55. Gogus, A. Constructivist learning. In *Encyclopedia of the Sciences of Learning*, 1st ed.; Seel, N., Ed.; Springer: New York, NY, USA, 2012; pp. 783–786. [CrossRef]
56. Finnish National Agency for Education. *National Core Curriculum for Basic Education 2014*; Finnish National Agency for Education: Helsinki, Finland, 2016; ISBN 9789521362590.
57. Lehtonen, D.; Joutsenlahti, J. The use of manipulatives in mathematics classrooms from the pedagogical and practical perspective. In Proceedings of the 17th Biennial EARLI Conference for Research on Learning and Instruction, Tampere, Finland, 27 August–2 September 2017.
58. Lehtonen, D.; Joutsenlahti, J. Promoting the use of manipulatives for mathematical concepts understanding in comprehensive school classrooms: Taking pedagogy and practicality into account. In Proceedings of the 6th NoFa-Conference 2017, Odense, Denmark, 29–31 May 2017; p. 17.
59. Hoda, R.; Henderson, A.; Lee, S.; Beh, B.; Greenwood, J. Aligning technological and pedagogical considerations: Harnessing touch-technology to enhance opportunities for collaborative gameplay and reciprocal teaching in NZ early education. *Int. J. Child–Comput. Interact.* **2014**, *2*, 48–59. [CrossRef]
60. Hatfield, M.M. Use of manipulative devices: Elementary school cooperating teachers self-report. *Sch. Sci. Math.* **1994**, *94*, 303–309. [CrossRef]
61. Marshall, L.; Paul, S. Exploring the use of mathematics manipulative materials: Is it what we think it is? In Proceedings of the EDU-COM 2008 International Conference, Perth, Australia, 19–21 November 2008; pp. 338–350.
62. Bedir, G.; Özbek, Ö.Y. Primary school teachers' views on the preparation and usage of authentic material. *High. Educ. Stud.* **2016**, *6*, 138–146. [CrossRef]
63. Spacedesk Home Page. Available online: https://spacedesk.net (accessed on 1 July 2020).

64. Unity Asset Store. OpenCV for Unity. Available online: https://assetstore.unity.com/packages/tools/integration/opencv-for-unity-21088 (accessed on 1 July 2020).
65. Akhtar, H.; Kemao, Q.; Kakarala, R. A review of sensing technologies for small and large-scale touch panels. In Proceedings of the 5th International Conference on Optical and Photonics Engineering (icOPEN 2017), Singapore, 13 June 2017; pp. 1044918:1–1044918:13.
66. Walker, G. A review of technologies for sensing contact location on the surface of a display. *J. Soc. Inf. Disp.* **2012**, *20*, 413–440. [CrossRef]
67. Liang, R.; Cheng, K.; Su, C.; Weng, C.; Chen, B.; Yang, D. GaussSense: Attachable stylus sensing using magnetic sensor grid. In Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST'12), Cambridge, MA, USA, 7–10 October 2012; pp. 319–326.
68. Nintendo. Amiibo. Available online: https://www.nintendo.com/amiibo/ (accessed on 1 July 2020).
69. Reinschluessel, A.V.; Döringring, T.; Alexandrovsky, D.; Thieme, D.; Malaka, R. Beyond math manipulatives: Smart tangible objects for algebra learning. In Proceedings of the 6th Workshop on Interacting with Smart Objects Co-Located with ACM Conference on Human Factors in Computing Systems (CHI '18), Montreal, QC, Canada, 21–22 April 2018; pp. 9–16.
70. Bottino, A.; Martina, A.; Strada, F.; Toosi, A. GAINE–A portable framework for the development of edutainment applications based on multitouch and tangible interaction. *Entertain. Comput.* **2016**, *16*, 53–65. [CrossRef]
71. Tevella. Kymmenjärjestelmä Pieni [Small Base 10 System]. Available online: https://www.tevella.fi/kuluttajat/kymmenjarjestelma-pieni (accessed on 14 July 2020).
72. Beaudouin-Lafon, M.; Mackay, W.E. Prototyping tools and techniques. In *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, 3rd ed.; Jacko, J.A., Ed.; CRC Press: New York, NY, USA, 2012; pp. 1081–1104. ISBN 1410606724.
73. Organisation for Economic Cooperation and Development (OECD). *Education Policy Outlook 2015: Making Reforms Happen*; OECD Publishing: Paris, France, 2015.
74. All European Academies. *The European Code of Conduct for Research Integrity*, Revised ed.; All European Academies: Berlin, Germany, 2017; ISBN 978-3-00-055767-5.
75. Creswell, J.W.; Plano Clark, V.L. *Designing and Conducting Mixed Methods Research*, 3rd ed.; SAGE: Thousand Oaks, CA, USA, 2018; ISBN 9781483344379.
76. Slavin, R.E. Research on cooperative learning and achievement: What we know, what we need to know. *Contemp. Educ. Psychol.* **1996**, *21*, 43–69. [CrossRef]
77. Elo, S.; Kyngäs, H. The qualitative content analysis process. *J. Adv. Nurs.* **2008**, *62*, 107–115. [CrossRef] [PubMed]
78. Charters, E. The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Educ.* **2003**, *12*, 68–82. [CrossRef]
79. Lund, A.M. Measuring usability with the USE questionnaire. *Usability Interface* **2001**, *8*, 3–6.
80. Hinton, P.R.; McMurray, I.; Brownlow, C. *SPSS Explained*, 2nd ed.; Routledge: East Sussex, UK, 2014.
81. Nikiforidou, Z. Probabilities and preschoolers: Do tangible versus virtual manipulatives, sample space, and repetition matter? *Early Child. Educ. J.* **2019**, *47*, 769–777. [CrossRef]
82. Clements, D.H. 'Concrete' manipulatives, concrete ideas. *Contemp. Issues Early Child.* **2000**, *1*, 45–60. [CrossRef]
83. Martin, T.; Svihla, V.; Smith, C.P. The role of physical action in fraction learning. *J. Educ. Hum. Dev.* **2012**, *5*, 1–17.
84. Sapounidis, T.; Demetriadis, S. Tangible versus graphical user interfaces for robot programming: Exploring cross-age children's preferences. *Pers. Ubiquitous Comput.* **2013**, *17*, 1775–1786. [CrossRef]
85. Pontual Falcão, T.; Price, S. Interfering and resolving: How tabletop interaction facilitates co-construction of argumentative knowledge. *Int. J. Comput. Supported Collab. Learn.* **2011**, *6*, 539–559. [CrossRef]
86. Nielsen, J. Usability 101: Introduction to Usability. Available online: www.nngroup.com/articles/quantitative-studies-how-many-users (accessed on 31 July 2020).
87. Perkkilä, P.; Joutsenlahti, J.; Sarenius, V. Peruskoulun matematiikan oppikirjat osana oppimateriaalitutkimusta [Elementary school mathematics textbooks as part of research on learning materials]. In *Matematiikan Opetus Ja Oppiminen [Mathematics Teaching and Learning]*; Joutsenlahti, J., Silfverberg, H., Räsänen, P., Eds.; Niilo Mäki Institute: Jyväskylä, Finland, 2018; pp. 344–367. ISBN 978-951-39-7584-5.

88. Van den Heuvel-Panhuizen, M.; Wijers, M. Mathematics standards and curricula in the Netherlands. *ZDM Math. Educ.* **2005**, *37*, 287–307. [CrossRef]
89. Moyer-Packenham, P.S.; Salkind, G.M.; Bolyard, J.; Suh, J.M. Effective choices and practices: Knowledgeable and experienced teachers' uses of manipulatives to teach mathematics. *Online J. Educ. Res.* **2013**, *2*, 18–33.
90. Uribe-Flórez, L.; Wilkins, J. Elementary school teachers' manipulative use. *Sch. Sci. Math.* **2010**, *110*, 363–371. [CrossRef]

PUBLICATION
IV

Constructing a Design Framework and Design Methodology from Educational Design Research on Real-World Educational Technology Development

Daranee Lehtonen

Educational Design Research, 5(2), Article 38
https://doi.org/10.15460/eder.5.2.1680

Educational Design Research

Volume 5 | Issue 2 | 2021 | Article 38

Contribution Academic Article

Title **Constructing a design framework and design methodology from educational design research on real-world educational technology development**

Author **Daranee Lehtonen**
Tampere University
Finland

Acknowledgements The initial research was supported by the Finnish Cultural Foundation, Häme Regional Fund 2015. The working prototype was developed in cooperation with Fouzia Khan, Juho Korkala, Roni Perälä, Niko Sainio, Lucas Machado, and Krishna Bagale under Pekka Mäkiaho's supervision.

Abstract Educational design research (EDR) seeks to contribute to both practice and theory by developing solutions that improve educational practice and generating usable and generalisable knowledge. Most EDR researchers tend to focus on reporting their research contributions to educational practice. Therefore, there is a need for disseminating research that pays more attention to the theoretical contributions of EDR so that those outside a particular EDR project can benefit. This paper focuses on the theoretical contributions, particularly the design framework and design methodological knowledge, of a 6-year EDR enquiry that aimed to develop educational technologies that promote primary school mathematics learning and classroom practice. Informed by the literature and direct experiences of working in collaboration with teachers and various disciplines during this iterative study, a design framework for developing real-world educational technologies and guidelines for conducting EDR are proposed. The design framework highlights four essential aspects – content, pedagogy, practice, and technology – that should be considered when developing educational technologies to ensure their educational benefits, feasibility, and successful real-world utilisation and adoption. The proposed guidelines for conducting EDR, such as exploring design alternatives and employing appropriate design construction and evaluation methods, can assist other researchers, including a single doctoral student, in embracing opportunities and overcoming the challenges that may emerge.

Keywords educational design research, educational technology, technology-enhanced learning, design framework, design methodology

DOI dx.doi.org/10.15460/eder.5.2.1680

Citation Lehtonen, D. (2021). Constructing a design framework and design methodology from educational design research on real-world educational technology development. *EDeR – Educational Design Research,* 5(2), 1-29.

dx.doi.org/10.15460/eder.5.2.1680

Constructing a design framework and design methodology from educational design research on real-world educational technology development

Daranee Lehtonen

1.0 Introduction

Educational design research (EDR) strives to bridge the gap between theory and practice in educational research by contributing to both practice and theory (e.g., McKenney & Reeves, 2019; Plomp, 2013). As a contribution to educational practice, EDR aims to craft research-informed solutions, such as educational products, processes, programmes, and policies, through iterative development in real-world educational settings, where teaching and learning actually take place. At the same time, EDR seeks to contribute to the research community by advancing usable and generalisable knowledge constructed during the iteration of empirical investigation.

According to Edelson (2002), EDR can assist in developing three types of theories: domain theories, design frameworks, and design methodologies that can inform the work of others. *Domain theories* describe the teaching and learning challenges and opportunities in a real educational context (i.e., *context theories*) and explain how the design solution works in that educational setting (i.e., *outcomes theories*). A *design framework* (i.e., a generalised design solution) describes the important characteristics of a design solution to a particular educational problem. A *design methodology* provides guidelines (e.g., processes, required expertise, and roles of the individual participants) for conducting EDR to achieve the research aims. To date, Edelson's (2002) three types of theories are still used to inform EDR, such as that of Kerslake (2019).

Despite manifold possible contributions of EDR, our systematic review (Lehtonen et al., 2019) and the work of others (Anderson & Shattuck, 2012; McKenny & Reeves, 2019; Zheng, 2015) have indicated that most EDR researchers tend to focus on reporting their research contributions to educational practice and domain theories. A few researchers have reported how their EDR contributes to knowledge of design frameworks (e.g., Bergdahl et al., 2018; Lambert & Jacobsen, 2019) and design methodologies (e.g., Cowling & Birt, 2018; Di Biase, 2020). Therefore, there is a need for the dissemination of research that pays more attention to design frameworks and design methodologies so that those outside a particular EDR project can benefit. Design frameworks can inform other educational researchers and designers on how to develop a solution to a similar educational challenge in another context, while design methodologies can help other researchers overcome challenges in conducting their EDR.

A growing body of EDR has responded to the increasing utilisation of technologies in educational environments. Recently, various technological solutions – for example, a digital video game (Lambert & Jacobsen, 2019), a mixed reality simulation (Cowling & Birt, 2018), and a virtual learning environment (Bergdahl et al., 2018) – have been developed and implemented in different educational contexts. Hence, it is beneficial to share usable and generalisable knowledge gained from previous EDR on technology-enhanced learning with others so that they can accomplish their future work in the area by building on what has already been learnt. As such, this paper reports the theoretical contributions of a 6-year EDR enquiry that aimed to develop educational technologies that promote primary school mathematics learning and classroom practice. Informed by the literature and my direct experiences of working in collaboration with teachers and various disciplines during this EDR on technology-enhanced learning, this paper seeks to further knowledge of:

1. Design frameworks: key aspects to be taken into account when developing educational technologies to ensure their educational benefits, feasibility, and successful real-world utilisation and adoption.
2. Design methodologies: guidelines for successfully conducting EDR.

2.0 EDR on educational technologies for learning linear equations

2.1 Educational problem

Linear equation solving, an important area in algebra, is often challenging for students at different educational levels to master (e.g., McNeil et al., 2019; Poon & Leung, 2010), particularly when they lack sufficient understanding of key concepts for solving equations, such as equations and equivalence (e.g., Knuth et al., 2006; McNeil et al., 2019). *Conceptual understanding* (i.e., understandings of mathematical concepts, operations, and relations) is one of the most important mathematical proficiencies (Kilpatrick et al., 2001). While strong conceptual understanding has several benefits for mathematics learning, insufficient conceptual understanding can hinder students' learning and performance in mathematics (e.g., Andamon & Tan, 2018; Kilpatrick et al., 2001).

Manipulative materials, such as beads and base-10 blocks, have long been used, particularly in preschools and primary schools, as hands-on learning tools that allow students to concretely explore abstract mathematical concepts through different senses. There is evidence that when manipulatives are used meaningfully, they can promote students' understanding of mathematical concepts. For example, when they are used for developing conceptual understanding (instead of attaining procedural fluency) and making links between various representations constructed through the manipulatives and mathematical symbols of the concept to be learnt (e.g., Kilpatrick et al., 2001; McNeil & Jarvin, 2007; Uttal et al., 2013). Nevertheless, disagreement exists between manipulatives' pedagogical benefits and classroom practice. Despite primary and lower secondary school teachers regarding manipulatives as useful learning tools, for instructional activities in their classrooms, they usually favour traditional teacher-centred and paper-and-pencil instruction over manipulatives (e.g., Marshall & Swan, 2008; Toptaș et al., 2012). This finding implies that pedagogically sound manipulatives may not be adopted in the classroom, possibly because of classroom-practice-related reasons. Thus, there is a call for a study on the application of the theoretical knowledge of manipulatives for the promotion of manipulative use in the classroom to enhance students' mathematical concept understanding.

2.2 Study overview

This study employed an EDR approach to bridge the gap between research on manipulatives and its direct practical contributions to real-world educational challenges (i.e., the disagreement between manipulatives' pedagogical benefits and classroom practice). It aimed to investigate the use of manipulatives in real educational contexts and then to develop a research-informed manipulative that not only enhances primary school students' understanding of equation-solving concepts, but also promotes its utilisation and adoption in the classroom. This study was conducted from 2015–2020 and was my doctoral research. It was self-initiated and not part of any research project; thus, I conducted this 6-year EDR enquiry independently.

Drawing on the widely used EDR process proposed by McKenney and Reeves (2019, pp. 83–84), Figure 1 shows the overall process of this study. The study involved multiple iterations of investigation, design and construction, and evaluation and reflection. The research comprised three main phases (i.e., initial research, concept development, and design development), which were divided into six iterative sub-cycles. Although the process flow depicted in Figure 1 moves from left to right, the actual process was not linear, but rather iterative (i.e., results from one element repeatedly fed into others) and flexible (i.e., some sub-cycles were revisited).

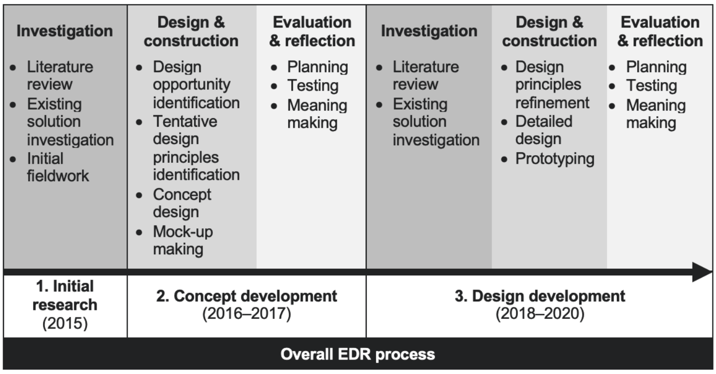


Figure 1: Overall EDR process of the study

The empirical study took place in primary and lower secondary schools in Finland. The uniform quality of the Finnish education system and its teachers, coupled with the students' homogeneous mathematics performance regardless of their socioeconomic background (Organisation for Economic Cooperation and Development, 2017, 2020), enabled the study to be conducted in any school. Mixed methods research, which combines qualitative and quantitative research (see e.g., Creswell & Plano Clark, 2017), was employed for data collection, analysis, and interpretation to better understand the real-world complexity (e.g., Anderson & Shuttuck, 2012; McKenney & Reeves, 2019). Research ethics and integrity were assured; the study was conducted according to the guidelines of the Finnish National Board on Research Integrity (2009, 2012) and All European Academies (2017). Altogether, 18 teachers (teaching experience 3–27 years), 98 primary school students (aged 9–12), and 65 lower secondary school students (aged 13–16) participated in different phases of the study. Table 1 summarises the mixed methods research design of the empirical sub-cycles (i.e., initial fieldwork, concept evaluation, and design evaluation).

Table 1: The mixed methods research design of the study's empirical sub-cycles

Empirical sub-cycles	Data	Analysis
Phase 1: Initial fieldwork	• Teacher interviews (N = 4)	• Inductive content analysis
	• Class intervention observations (teachers N = 4; students in paper-and-pencil group n = 25, in physical manipulative group n = 25, in virtual manipulative group n = 24)	• Inductive content analysis
	• Student paper-based tests (paper-and-pencil group n = 25, physical manipulative group n = 25, virtual manipulative group n = 24)	• Descriptive statistical analysis • Inferential statistical analysis (95% confidence intervals)
	• Student self-evaluations (paper-and-pencil group n = 25, physical manipulative group n = 25, virtual manipulative group n = 24)	• Descriptive statistical analysis
Phase 2: Concept evaluation	• Teacher questionnaires (N = 12)	• Descriptive statistical analysis
	• Teacher interviews (N = 12)	• Inductive content analysis
Phase 3: Design evaluation	• Class intervention observations (teachers n = 2; students in paper-and-pencil group n = 12, in developed manipulative group n = 12)	• Inductive content analysis • Deductive content analysis • Descriptive statistical analysis • Inferential statistical analysis (Pearson's chi-squared test)
	• Student paper-based tests (paper-and-pencil group n = 12, developed manipulative group n = 12, comparison group without participating in class intervention n = 65)	• Descriptive statistical analysis • Inferential statistical analysis (Mann–Whitney U test)
	• Thinking aloud sessions of students in developed manipulative group (n = 12)	• Inductive content analysis
	• Questionnaires of students in developed manipulative group (n = 12)	• Descriptive statistical analysis
	• Interviews of students in developed manipulative group (n = 12)	• Inductive content analysis
	• Teacher questionnaires (N = 6, two participated in the class interventions)	• Descriptive statistical analysis • Inferential statistical analysis (Wilcoxon matched-pairs signed-rank test)
	• Teacher interviews (N = 6, two participated in the class interventions)	• Inductive content analysis

The research design and results of the initial fieldwork have been reported in detail in Lehtonen and Joutsenlahti (2017). The concept evaluation has not been published elsewhere; its research design is elaborated in Section 2.3.2 and results in Section 3.3. The design and development of the design solution (i.e., design principles refinement, technological development and implementation, features and interactions, prototyping, and future development) and its evaluation (i.e., research design and results) have been thoroughly reported in Lehtonen et al. (2020).

2.3 Research phases

2.3.1 Phase 1: Initial research

Initial research was undertaken to gain knowledge of the following two key areas to construct a context theory: (1) the to-be-solved educational problem and the target educational context (i.e., challenges, opportunities, and needs regarding teaching and learning equation solving in primary school classrooms) and (2) existing solutions (see McKenney & Reeves, 2019). First, I conducted a literature review to investigate the state-of-the-art relevant to the research, including learning theories/models, equation solving, and manipulative use. Then, I analysed various existing physical and virtual manipulatives and educational games for equation solving in terms of their key benefits and limitations to mathematics classrooms. Based on the literature review and existing manipulative analysis findings, fieldwork was conducted in real classrooms to define the problem in practice, understand the real context, and investigate how existing manipulatives support or hinder classroom activities (see Lehtonen & Joutsenlahti, 2017).

Class interventions were implemented with four primary school teachers and their students (n = 74) with no/low prior knowledge of equation solving. Students with different attainment levels were equally divided into either learning with paper-and-pencil (n = 25) or learning with manipulatives (physical manipulative: n = 25, virtual manipulative: n = 24). The teachers were interviewed before the class interventions about their prior experiences and needs in using manipulatives, and after the interventions about their experiences and opinions of the interventions. All the interventions were observed and video recorded. Afterwards, all the students completed the same paper-based test with no access to the manipulatives and evaluated their learning experiences and achievements.

2.3.2 Phase 2: Concept development

This phase aimed to generate alternative design concepts informed by Phase 1 results, evaluate the generated concepts, and select promising one(s) for further development (see Ulrich & Eppinger, 2016). I used the Phase 1 findings to identify design opportunities and tentative design principles that provided initial ideas about how a manipulative can promote students' equation-solving concept understanding and its use in the classroom. Based on that, I explored different potential solutions by first generating ideas and then reviewing each from not only rational and analytical viewpoints, but also instinctive and intuitive viewpoints, as recommended by McKenney and Reeves (2019) and Ulrich and Eppinger (2016). Four generated ideas were selected for development as potential design concepts. Then, I built a nonfunctional mock-up to describe each concept in terms of its key functional features and initial visual appearance.

I conducted a concept evaluation with 12 primary school teachers (four participated in the initial fieldwork) via questionnaire and interview to determine the potential viability of each envisioned concept in the target setting and to collect teachers' feedback for further design decisions. First, I introduced all the concepts to the teachers and asked them to assess how well each concept was likely to benefit students' learning and conform to classroom and school practice. A scale of 1 (*not at all*) to 4 (*very well*) was used. They were also asked to provide an explanation of their rating responses. Teachers were selected to evaluate the concepts because they were sufficiently knowledgeable about educational contexts to envision concept implementation and adoption possibilities. Furthermore, according to McKenney and

Reeves (2019), it would not be socially responsible to excessively interrupt normal classrooms for evaluation of underdeveloped solutions.

2.3.3 Phase 3: Design development

This phase aimed to develop the selected concept(s) according to the design principles, evaluate the design, and reflect on the evaluation results (see McKenney & Reeves, 2019; Ulrich & Eppinger, 2016). To inform the development of the design solution, I conducted another literature review (e.g., technology-enhanced learning and tangible technologies) and investigated relevant educational products (e.g., textbooks and educational technologies). Subsequently, I incorporated key knowledge underlined by the literature, the existing solution investigation, and Phase 2 concept evaluation results to refine the design principles. Based on the concept evaluation, the most promising concept was selected for further development, guided by the refined design principles. When developing the selected concept, I also took into account the teachers' positive feedback from Phase 2 about other concepts to incorporate their strengths into the developed solution. Consequently, physical artifacts and class activities were developed as the solution to the educational problem.

The next step was prototyping. As I have a strong background in design (i.e., educational, graphic, user interface, and product design), I built most parts of the working prototypes myself and designed the graphics and interface of a tablet application that was part of the developed manipulative. I sought and received expert information technology and communication assistance for prototyping the application's technological aspects. Finally, the manipulative (i.e., the tablet application and related physical objects) was developed and prototyped in collaboration with a team of six university students and their supervisor at the Faculty of Information Technology and Communication Sciences as part of their coursework.

The development team and I took part in the development and formative evaluation of the manipulative prototype according to our own expertise and discipline. The development team was responsible for technology implementation and the iterative process of programming and use-case scenario testing, while I was responsible for educational, graphics, interface, and product design. Continual discussion was required for issues that required the expertise of both parties. The collaboration involved daily communication (e.g., email, and Slack: a real-time communication platform for teamwork) and monthly face-to-face official meetings. Unfortunately, at the end of the 4-month collaboration, only some components of the working prototype fully functioned as envisioned. Moreover, the technical stability of the prototype was heavily dependent on external factors, particularly classroom lighting conditions. Consequently, Microsoft PowerPoint was used to build a Wizard of Oz rapid prototype (see Beaudouin-Lafon & Mackay, 2012) of the application for class intervention evaluation.

I conducted a field test to evaluate the developed solution in real classrooms in terms of its pedagogical benefits, usability, and compatibility with classroom and school practice (see Lehtonen et al., 2020). The class interventions were implemented with two primary school teachers and their students (*n* = 24) with no/low prior knowledge of equations solving. Students with different attainment levels were divided equally into two groups: learning with the developed manipulative (*n* = 12) or learning without it (*n* = 12). Apart from that, both groups used the same components of the developed solution (i.e., a teacher guide, a worksheet, and class activities) during the class interventions. The processes for the intervention and paper-based posttest were similar to those of the Phase 1 initial fieldwork. Additionally, the students who used the developed manipulative individually participated in a thinking aloud session, in which they were asked to solve equations with the manipulative and simultaneously explain about their actions. They also completed a questionnaire and were interviewed regarding their perceptions of usefulness and usability of the manipulative. Lower secondary school students (*n* = 65), who had learnt equation solving as a part of their normal school curricula (representing traditional classrooms), took the same paper-based test, but without participating in the class intervention. To investigate the extent to which the developed solution enhanced students' learning achievement compared to traditional instruction, the paper-based test performance of the students participating in the class interventions was compared to that of the students who did not take part. After the interventions, both teachers evaluated the developed manipulative using a similar process to those of the Phase 2 concept evaluation. The manipulative was also evaluated by one special education teacher and three lower secondary school teachers to examine the possibility of its implementation and adoption in educational settings outside the studied context.

Based on the results of the formative evaluation during the working prototype development, some physical components that interacted with the application were revised to improve the technical stability of the manipulative. The field test results and reflection informed minor refinement of the manipulative to enhance its pedagogical value, usability, and practicality.

3.0 Developing a design solution

This section describes how the design solution was developed to meet the research objectives (i.e., to enhance students' equation-solving concept understanding and conform to classroom and school practice), what informed its development, and whether it accomplished the desired objectives.

3.1 Design principle construction

The initial research results from Phase 1 revealed important pedagogical factors that tend to influence the successful manipulative utilisation and adoption in the classroom (Lehtonen & Joutsenlahti, 2017). Based on this, tentative design principles were identified and used to guide potential design concept exploration. It should be noted that different practical factors that could pose challenges to manipulative utilisation and adoption were also discovered. They were not the main consideration when generating design concepts, but were used later to evaluate the generated concepts and to further develop the design solution.

In terms of pedagogy, discovery learning was identified in the literature (e.g., Bruner, 1961; Neber, 2012) as important to students' meaningful learning with manipulatives. Thus, the manipulative was designed to help students learn equation solving through firsthand experience with guidance and scaffolding. To ensure the benefits of the manipulative to students' conceptual understanding, it was also designed based on the literature (e.g., Joutsenlahti & Kulju, 2017; Lesh et al., 1987) and Phase 1 initial fieldwork to help students link various representations of equation-solving concepts (i.e., physical actions, pictures, mathematical symbols, and natural language) and to express their mathematical thinking through multiple modes. Additionally, informed by social constructivism, the manipulative was designed to encourage students to work in pairs/small groups, thereby co-constructing their knowledge through peer interaction (e.g., Slavin, 2010; Vygotsky, 1978).

Regarding to-be-learnt content, the manipulative was designed to help students learn equation solving by concretising fundamental concepts for solving equations: mathematical equivalence (i.e., both sides of an equation are equal), different terms in an equation (i.e., constants and unknowns), and equation solving (i.e., finding the values of the unknowns so that the equation is true or showing that there is no real number-value solution to the equation). These were emphasised in the literature (e.g., McNeil et al., 2019; Otten et al., 2019; Poon & Leung, 2010). To achieve the subject-matter objective, the manipulative employed the balance model (see e.g., Otten et al., 2019) to emphasise the relational (rather than operational) meaning of the equal sign and used different physical objects to represent constants (e.g., 3) and unknowns (e.g., 3x).

Empirical findings from the initial research indicated that the physical manipulative provided more benefits (e.g., concrete representation, tactile-kinesthetic interaction, flexibility, and ease of use) than the virtual manipulative. Nevertheless, the virtual manipulative had several advantages over the physical one, for example, visual and symbolic links, real-time guidance and scaffolding, and interactivity. The discovered strengths and limitations of both manipulative types were taken into account when exploring different possibilities. Consequently, four potential design concepts were generated (Figure 2).

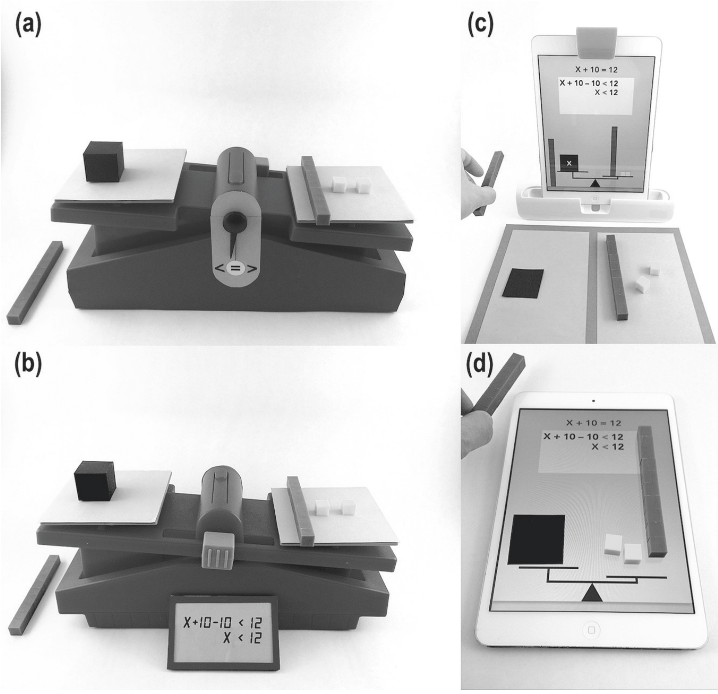


Figure 2: How to solve an equation (x + 10 = 12) using each of the four generated design concepts. (a) Concept A; (b) Concept B; (c) Concept C; (d) Concept D.

3.2 Concept design

All four concepts share similar core ideas informed by the Phase 1 findings, as described in the previous section. The concepts differ in how they utilise existing technologies to meet the design objectives. Hence, the generated design concepts were not driven by technologies, but rather by what technologies could offer to solve the educational problem. This design approach prevented the *technology before pedagogy* effect, which occurs when a technology is chosen prior to the identification of an educational problem (Watson, 2001; cf. technology-driven products; Ulrich & Eppinger, 2016).

Concept A (Figure 2a) consists of a physical balance scale and physical objects: black boxes as unknowns and base-10 blocks as constants (e.g., yellow units representing ones and green rods representing tens). Each side of the scale represents each side of the equation, while the balance or tilting of the scale represents the relationship (i.e., equality or inequality, respectively) between the mathematical expressions on each side of the equation. Concept B (Figure 2b) consists of a physical balance scale with a digital display, which shows math sentences of the current equation-solving process, and physical objects. Concept C (Figure 2c) consists of a tablet application, a mirror placed in front of a tablet camera for physical object detection, physical objects, and a mat divided into two parts representing both sides of the scale. Math sentences of the current equation-solving process and graphical images of the physical objects currently on and off the scale are displayed on a tablet screen. Concept D (Figure 2d) consists of a tablet application and physical objects. The tablet touchscreen directly detects physical objects on the screen and provides outputs (i.e., graphical images and mathematical symbols) for seamless interaction.

3.3 Concept evaluation

Each generated concept was evaluated by teachers in terms of its potential pedagogical benefits (i.e., enhancing students' understanding of equation-solving concepts, discovery learning, social interaction, and multimodal expression of mathematical thinking) and compatibility with classroom and school practice (e.g., acquisition budget, storage space, organisation and preparation, and class management). The evaluation criteria were informed by previous studies (e.g., Bedir & Özbek, 2016; Marshall & Swan, 2008) and Phase 1 fieldwork. To take policymakers' needs into account, the pedagogical criteria were also guided by the current Finnish National Core Curriculum (NCC) for Basic Education (Finnish National Agency for Education [EDUFI], 2016).

The teachers (N = 12) rated each concept according to the evaluation criteria, and the average scores for each concept are shown in Figure 3. All the concepts were rated relatively high (3–3.5 on the 4-point scale of 1 [*not at all*] to 4 [*very well*]) in terms of their potential benefits for students' learning. However, only Concept D was highly rated (M = 3.5, SD = 0.78) for its compatibility with classroom and school practice compared to the others, which were rated below 3 (*well*). At the end of the concept evaluation, the teachers were asked to make an acquisition decision by taking both pedagogical and practical factors into account. Most teachers (n = 9) would acquire Concept D for their own class, followed by Concept B (n = 6), Concept A (n = 5), and Concept C (n = 2). No teachers mentioned that they would not acquire Concept D for their class, but six would not acquire Concept C, and four would not acquire Concept A or Concept B. According to the teachers' explanations, there were no major differences between the concepts regarding their pedagogical benefits, so compatibility with classroom and school practice was the decisive factor in the teachers' acquisition decisions. Concept D was the most desirable because of its highly perceived pedagogical benefits, as well as its simplicity, usability, compactness, portability, durability, compatibility with existing school tablets, attractiveness to diverse learners (in terms of age and attaining levels), and enduring utility. Hence, Concept D was selected for further development.

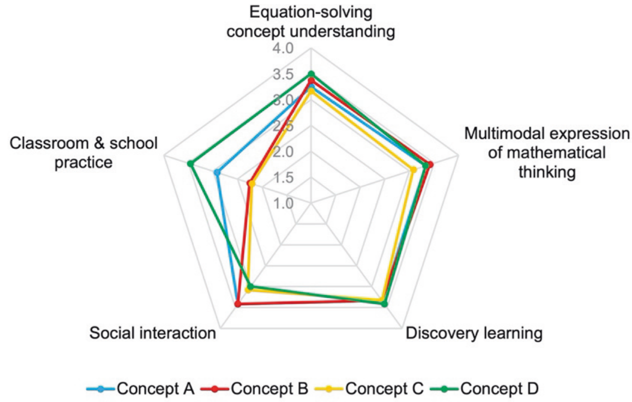


Figure 3: Average scores of each concept rated by 12 teachers regarding how well they potentially meet each criterion. Scores ranged from 1 (not at all) to 4 (very well).

3.4 Solution design and development

Based on the refined design principles (e.g., concretising key equation-solving concepts; being in agreement with curriculum; supporting multimodality, discovery learning, and social interaction; and being feasible for classroom and school practice; Lehtonen et al., 2020), physical artifacts (i.e., a manipulative, student worksheets, and teacher guides) and class activities (how the manipulative and worksheet should be used in the classroom) were proposed as the solution to the identified educational problem. The manipulative was developed from Concept D as a student learning tool to be used for recommended class activities to ensure meaningful learning. Relying heavily on the literature mentioned in Section 3.1, the class activities were designed to promote students' understanding of equation-solving concepts through discovery learning, social interaction, and multimodal mathematical thinking expression. It is recommended that, under the teacher's supervision, students use the manipulative in pairs/small groups to model and solve an equation shown on the worksheet and then individually write the equation-solving processes and a solution(s) on their own worksheet.

The instructional materials include two student worksheets (the first for lower graders to learn to solve equations by substituting values for the unknown and the second for upper graders to learn to solve equations algebraically [Figure 4]) and two corresponding teacher guides. The instructional materials were developed based on the NCC (EDUFI, 2016), mathematics textbooks currently used in Finland, and the materials used during Phase 1 class interventions. The materials were evaluated and critiqued by the doctoral research supervisors acting as experts in subject matter and mathematics education and then refined according to the experts' advice.

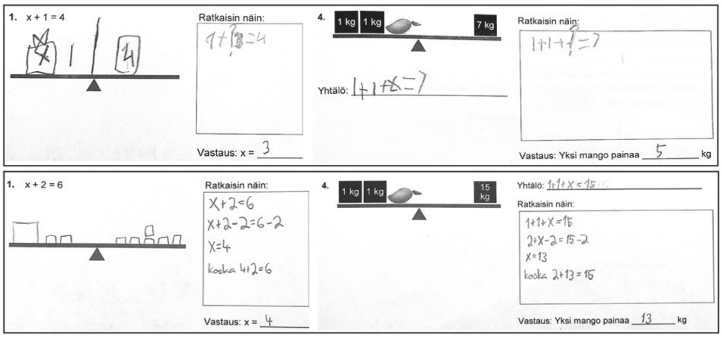


Figure 4: Extracts from the student worksheets of a fourth grader (above) and a fifth grader (below).

Each worksheet was designed to be used for one 45-minute lesson to facilitate students' conceptual understanding by working with multiple representations of equations. The worksheets show a summary of the contents and then eight equations presented through three different representations: mathematical symbols, pictures, and word problems. On the worksheets, students are required to first translate each equation into two other representations, then symbolically, pictorially, or verbally demonstrate how they have solved the equation, and finally provide the answer. The teacher guides were designed to assist teachers in planning and implementing a lesson for learning to solve equations using the manipulative and the worksheet. Each teacher guide consists of five parts: a story that introduces the equation solving, lesson objectives, suggestions for the lesson procedure, advice on how to explain the content to students, and additional information about the theoretical framework behind the class activities.

During the collaboration with the development team, the manipulative was developed based on the refined design principles and technological feasibility. Figure 5a shows the developed manipulative, including the tablet application and two types of physical objects: (1) specially designed X-Boxes representing unknowns and (2) existing and widely used base-10 blocks representing constants (see Lehtonen et al., 2020). The use of base-10 blocks was positively supported by the teachers during the concept evaluation, because teachers and students are usually familiar with them and they are commonly available in schools. The application has two levels containing the same exercises as in the worksheets for students at different grade levels to learn to solve equations. This expands the manipulative utility across primary grades to increase the likelihood of manipulative acquisition. Students manipulate physical objects on a tablet screen to model and solve equations, while the application provides corresponding guidance and scaffolding in textual, pictorial, mathematical symbolic, or audio form. The graphics and user interface design of the application was kept simple, allowing students to easily navigate the application and concentrate on the content.

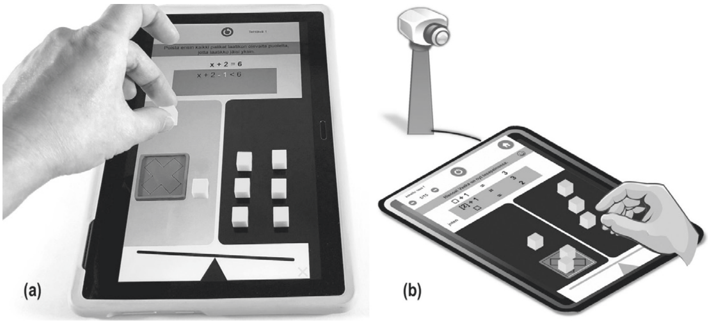


Figure 5: (a) The developed manipulative consisting of a tablet application, an X-Box(es), and base-10 blocks; (b) the object tracking via an external USB web camera. From 'The Potentials of Tangible Technologies for Learning Linear Equations,' by D. Lehtonen et al., 2020, Multimodal Technologies and Interaction, 4(4), Article 77 (https://doi.org/10.3390/mti4040077). CC BY 4.0.

While the design of the manipulative was much the same as the original concept, some trade-off decisions were made in the prototype to balance pedagogical, practical, and feasible factors, all of which were not achievable concurrently. For instance, at the beginning of the prototyping, it became clear that the original object-tracking idea (i.e., conductive technology) used in Concept D for the interaction between the physical objects and the application was not feasible due to the limitations of current tablet touchscreen technology. This led to a change from the original object tracking to image recognition via an external USB web camera (Figure 5b). A more complicated installation of the manipulative was required, which resulted in reduced practicality. However, at that time, it was technologically feasible and affordable. Importantly, it allowed for a single point of interaction (i.e., for students to manipulate physical objects on a tablet screen and look at the outputs on the screen without losing concentration); this was a critical design principle, which was informed by Concept D that should not be traded off. Another decision was to simplify a graphical representation of the balance scale on the tablet screen to overcome the limited screen space (see the original scale in Figure 2d vs. the simplified one in Figure 5a). Despite the fact that the simplified representation was not completely in line with physics, based on teachers' responses during the Phase 3 design evaluation, the simplification was acceptable for primary and lower secondary school mathematics.

Overall, the proposed solution gained highly favourable results from the design evaluation (Lehtonen et al., 2020), thereby suggesting its successful classroom utilisation and adoption. The paper-based test results, class intervention observations, and students' and teachers' responses revealed that the design solution improved students' equation-solving learning and achievement. Moreover, based on the class intervention observations and the students' responses, the developed manipulative was easy and enjoyable to use. Most students found it helpful for their learning and would like to use it in the future. Teachers rated the manipulative highly (M = 3.4 on the 4-point scale, SD = 0.48) for its compatibility with classroom and school practice. All teachers would acquire the manipulative for their classrooms, and the number of acquired manipulatives would depend on their utilisation purpose (e.g., for the whole class or remedial teaching) and school budget. Based on the design evaluation and reflection, minor revisions (e.g., redesign of the X-Box to overcome image recognition challenges) were made; new features (e.g., free experiments with the balance scale for enhancing pedagogical value) were added to the forthcoming design.

4.0 Conducting EDR

This section highlights the lessons I learnt from conducting the EDR. Based on my research journals, field notes, and communication records with the development team and teachers, here I reflect on the benefits of EDR and the challenges that emerged during my study.

4.1 Iterations

As iteration is a key characteristic of EDR (e.g., Anderson & Shattuck, 2012), the study comprised six sub-cycles. The multiple iterations of investigation, development, assessment, and refinement helped me gradually develop my theoretical understanding of the real educational context (i.e., challenges and opportunities regarding utilisation of manipulatives for teaching and learning equation-solving concepts in primary schools) and how my proposed solution could enhance students' learning and classroom practice. Moreover, through this iterative process, I was able to design, test, and refine the solution to ensure its feasibility and successful implementation and adoption in the classroom. Despite the benefits, the iterations were intensive and required considerable resources. As a single researcher, it took me 6 years to complete the study.

4.2 Data triangulation

I collected empirical data from various sources (i.e., teachers and students of different grade levels) using a range of methods (i.e., multiple instruments and various data formats), as recommended in the EDR literature (e.g., McKenney & Reeves, 2019). While the data triangulation helped me better understand the complex and dynamic real-world educational phenomena and promoted the research reliability and validity (e.g., McKenney & Reeves, 2019), it was unavoidably resource-intensive to collect and analyse such a large and varied dataset. With limited resources, I had to focus on the data directly related to the research question. As a result, a large amount of collected data was left unused.

4.3 Various participants

In line with the EDR literature (e.g., Ørngreen, 2015), the involvement of teachers and students, who were potential implementors and learners of the proposed solution, largely informed the solution development towards the desired outcomes. The concept and design evaluation results indicate that the developed solution is likely to promote students' conceptual understanding of equation solving and classroom practice. Therefore, it has the potential to be implemented and adopted into real classrooms. The design evaluation also revealed students' adaptations of the manipulative use to meet their needs in ways that would not have been discovered without their involvement.

Collaborating with teachers and students not only contributed to the advancement of theoretical knowledge and the improvement of educational practice, but also benefitted teachers and students. Most teachers stated that participation helped them realise how manipulatives can enhance students' conceptual understanding. One even changed her strong prior perception that manipulatives only benefit young students. Consequently, all teachers mentioned their intention to regularly incorporate manipulatives into their classrooms. After the class interventions, one class actually adopted a discussion about their own mathematical thinking with peers, which was implemented during the class interventions as part of their mathematics class. Some teachers expressed their appreciation for the opportunity to participate, stating that participation enabled them to experience how different instructional methods influenced their students' mathematics learning and to understand how to best support their students.

Despite these benefits, I encountered challenges in involving teachers and students. Participant recruitment required careful planning and determined actions for practical and ethical reasons. It was difficult to access teachers who were willing to participate in class interventions that required intensive organisation and resources. Approval for the students' participation had to be obtained well in advance from the city's children and youth service director, school principals, and students' guardians. Additionally, unanticipated rearrangement or cancellation of participation occurred several times.

4.4 Multidisciplinary collaboration

Collaboration among different disciplines is important for successful EDR (e.g., McKenney & Reeves, 2019). Working with experts from the fields of mathematics and mathematics education (i.e., my supervisors) and information technology and communication sciences (i.e., the development team) helped me ensure that the solution was viable in terms of subject matter, pedagogy, and technology. Although I have a multidisciplinary background, I would not have been able to accomplish my research objectives without this multidisciplinary collaboration. We collaborated successfully due to good communication, mutual respect, and shared understanding. Regular face-to-face and online communication enabled us to interact and share ideas and work progress. In short, our complementary expertise enabled us to overcome the research and design challenges.

Nevertheless, I encountered some difficulties in the multidisciplinary collaboration. It was challenging for other collaborators to gain insight into the complex educational problems, and it took me almost a year to establish the collaboration with the development team. Another challenge resulted from the experience level of the development team members: undergraduate and graduate students. Most had no experience with the technologies used for the developed manipulative, thereby requiring training and additional study. Moreover, because the students participated in the study as part of their coursework without other incentives, it was difficult to secure their full-time commitment to our collaborative project.

4.5 Technological innovations

Technological possibilities play an important role in product design and development (e.g., Ulrich & Eppinger, 2016). While the developed manipulative that employs innovative technologies provided satisfactory results, the development of such educational technology resulted in several challenges. For example, only at the beginning of the manipulative prototyping did it become clear that the envisioned object-tracking idea was not feasible with off-the-shelf technology. Consequently, a considerable amount of the development team's time was devoted to finding other object-tracking alternatives, leaving less time for actual prototyping. Moreover, there were other technical difficulties in getting the prototype functioning properly. Due to limited time and the demanding task, the development team could not develop and construct a fully working prototype that was stable and contained all the needed features for the class interventions, despite the extended project deadline. Thus, I had to build the Wizard of Oz prototype for class intervention evaluation instead.

4.6 Alternative designs

Given my design experience, I acknowledge that working with alternative designs reduces the possibility of discovering later in the development process that the developed design might not be the best solution (e.g., Ulrich & Eppinger, 2016). Thus, I thoroughly explored alternative solutions at an early stage of concept development (Phase 2). The concept evaluation with teachers not only helped me better understand the educational context but also identify the issues to be addressed prior to further development and class interventions. This resulted in the efficient utilisation of resources. The teachers' feedback also allowed me to confidently select a concept for further development.

Additionally, the evaluation of different concepts was a good method for collecting in-depth information from teachers. During the initial research, when teachers were asked to recall their experiences or provide their perceptions of manipulatives, their responses were rather superficial. However, during the concept evaluation, the concepts worked as concrete examples of manipulatives and enabled teachers to provide more detailed and relevant responses to similar questions.

4.7 Solitary researcher

Without an accompanying research team, I was involved in every process of the EDR and had multiple roles. On a positive note, I had a deep understanding of the whole process. However, inevitably, this was also challenging due to the complex nature of EDR and the scope of the research, which required considerable time, patience, and multidisciplinary expertise from a single researcher. With a multidisciplinary background and without a restricted timetable, I conducted the study mainly independently over a 6-year period. However, the study still suffered from limited human resources. It required a multidisciplinary theoretical foundation, which I did my best to master. As the only researcher, I could operate only one Wizard of Oz rapid prototype at a time during the class interventions, thereby resulting in a small sample size of students working with the developed manipulative. Moreover, I was only able to carry out the concept evaluation and one implementation of the developed solution in classrooms. It is unlikely that a single implementation of the proposed solution in a real educational setting is sufficient to collect evidence indicating the success of the solution. Thus, I have planned postdoctoral research that could contribute to the research validity by constructing a number of working prototypes of the refined design solution that can be implemented in other educational contexts with larger sample sizes.

Working alone also had a negative impact on the research reliability. It was not possible to achieve researcher triangulation during the data collection and analysis. As cautioned in the EDR literature (e.g., Plomp, 2013), taking on multiple roles (i.e., researcher, designer, and evaluator) also challenged my maintenance of objectivity. It should be noted that I did not take on an implementor's role in any class interventions; teachers were the designated implementors. Thus, there was no researcher or designer influence on any class interventions. An early developed design is rarely flawless, and there is always room for improvement. Therefore, I saw the design evaluation as a means of gathering feedback to improve the proposed solution rather than demonstrating its perfection. Furthermore, I wanted to ensure that the research was EDR, not just a design and development project. So, I consciously acted according to Edelson's (2002) recommendation as an EDR researcher (instead of only a designer) to develop a novel solution to improve educational practice and use design implementation as a strategy to construct theoretical knowledge that makes EDR different from design practice (Easterday et al., 2017).

5.0 Looking back and moving forward

Informed by the literature and my experiences of developing the design solution (Section 3) and conducting this EDR (Section 4), I have constructed usable and generalisable knowledge of design frameworks and design methodologies. I am sharing this knowledge so that others outside this EDR project may benefit from it.

5.1 Design framework for developing real-world educational technologies

Brown (1992) argued that researchers need to acquire various types of knowledge to successfully develop solutions to improve real-world educational practice. This was the case for my study. To develop a promising solution that met my research goals, I needed to know about equation solving, meaningful ways to use manipulatives, relevant classroom and school practice, and possible technologies. Consequently, I propose a *real-world educational technology design framework* (i.e., a generalisation of the study-specific design principles) to inform other researchers and designers about what should be considered when developing educational technologies for real-world educational contexts. Figure 6 shows that the framework takes into account four essential aspects – content, pedagogy, practice, and technology – that contribute to educational benefits, feasibility, and real-world utilisation and adoption of educational technologies.

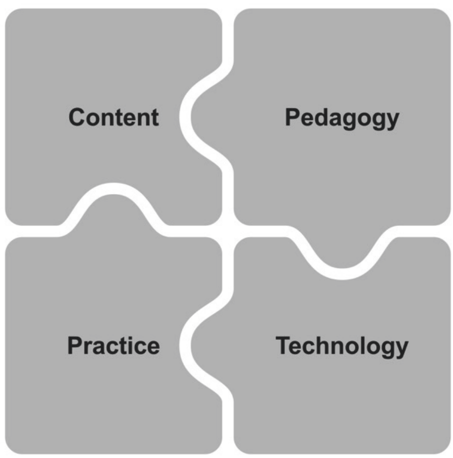


Figure 6: The real-world educational technology design framework takes into consideration content, pedagogy, practice, and technology. Illustration created by T. Lehtonen.

Content

Undeniably, an understanding of to-be-learnt subject matter content is compulsory for the development of any educational solutions. For example, to develop the manipulative and the student worksheets, which covered what was to be learnt, I needed knowledge of equation solving, important concepts required for understanding equation solving, different models for teaching equations, and equation-solving content stated in the NCC (EDUFI, 2016) and used in textbooks.

Pedagogy

An understanding of pedagogy (i.e., how to teach and learn the particular subject matter content in the target educational context) is required to ensure the meaningful use of the proposed design solution. For example, I needed to know how to use manipulatives to enhance students' understanding of equation-solving concepts. Social constructivism (e.g., discovery learning and social interaction), multimodal expression of mathematical thinking, and teaching and learning approaches recommended by the NCC (EDUFI, 2016) were used to guide my design solution.

Practice

Knowledge of practice in the target educational context largely contributes to the successful real-world implementation and adoption of educational solutions. In my case, the conformity of the manipulative to classroom and school practice (e.g., acquisition budget and class management) proved to be an important factor in teachers' acquisition decisions.

Technology

To design feasible educational technologies, it is necessary to know about technological possibilities, including what technologies are available, what they can offer, and how they work. For example, to develop the manipulative in collaboration with the development team, to some extent, I needed to understand digital technologies, tangible technologies (see Ishii & Ullmer, 2012), and different object tracking alternatives that could be used to solve the target educational problem.

It is worth noting that the importance of each aspect in the framework is usually not equal and largely depends on the nature of the educational problem, target setting, and possible technologies. In line with McKenney and Reeves (2019), during my study, design decision-making typically involved simultaneous consideration of multiple aspects. Often, I had to make trade-off decisions to fine-tune the manipulative that would best achieve the research objectives (Ulrich & Eppinger, 2016). For example: Concept D had the most pedagogical value but was not technologically feasible; Concept C was technologically feasible but not very practical or pedagogical; and Concept B was pedagogical and technologically feasible but not affordable. My theoretical, practical, and technological understandings gained from the literature, fieldwork, and multidisciplinary collaboration assisted me in making trade-off decisions.

Unintentionally, there are some similarities between the real-world educational technology design framework and the Technological Pedagogical Content Knowledge (TPACK) framework (see Koehler & Mishra, 2009). The TPACK framework builds on Shulman's (1986) construct of pedagogical content knowledge and includes technological knowledge to address essential teacher knowledge for integrating technology into teaching and learning, whereas my proposed design framework is rooted in design and development practice to inform important aspects to be considered when designing educational technologies for the real world.

5.2 Guidelines for conducting EDR

The nature of EDR (e.g., being situated in real educational settings and involving multiple iterations) results in a multifaceted and intensive enquiry requiring substantial resources (Kelly, 2013) and long-term endeavour (Collins et al., 2004). Consequently, doctoral students are often hesitant about employing EDR (Goff & Getenet, 2017; Herrington et al., 2007). Based on my experience of conducting my doctoral EDR, I agree with scholars (e.g., Herrington et al., 2007; Kennedy-Clark, 2013; McKenney & Reeves, 2019) that EDR can be undertaken even by a single doctoral student. I certainly encourage others to engage in EDR for its numerous benefits, as evidenced in my study.

Various EDR models, such as those by Easterday et al. (2017) and McKenney and Reeves (2019), have been employed; however, due to the uniqueness of each piece of EDR, researchers are typically required to adapt these models appropriately. Therefore, instead of proposing another model, I provide general guidelines (Figure 7) for conducting EDR to assist other researchers in embracing opportunities and overcoming the challenges that may emerge from their EDR.

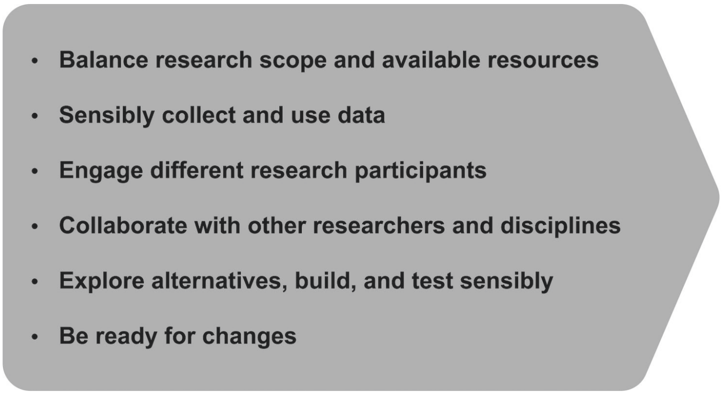


Figure 7: General guidelines for conducting EDR. Illustration created by T. Lehtonen.

Balance research scope and available resources

Because EDR is resource-intensive, it is important to estimate the resources available and then plan the EDR accordingly. This will ensure that the research is achievable and preserves the real-world iterative enquiry. I agree with Kennedy-Clark (2013) that a single researcher or a few researchers should aim for a less intensive small-scale enquiry, while a larger research team could strive for a more intensive large-scale one. In addition to the research team size, the research team type (e.g., monodisciplinary vs. multidisciplinary) and other resources (e.g., time and budget) should be taken into account when planning EDR. For example, financial support can play a crucial role in prototyping technological solutions, as in my study. Complex and long-term research can also be broken down into feasible components, for example, doctoral and postdoctoral research as in my and Goff's (2016, as cited in Goff & Getenet, 2017) cases or several small-scale studies for doctoral students to individually conduct, as recommended by Anderson and Shattuck (2012). Alternatively, results from a single iteration can be used to inform further research, as in Di Biase's (2020) case.

Sensibly collect and use data

The triangulation of data, as recommended by EDR scholars (e.g., McKenney & Reeves, 2019) and evidenced in my study, can contribute to a better understanding of multifaceted real-world phenomena and the trustworthiness of EDR. However, an endeavour to triangulate data often requires intensive resources to collect a significant amount of data (Collins et al., 2004). Due to limited resources, researchers, particularly doctoral students like Goff (2016, as cited in Goff & Getenet, 2017) and me, often select only the data that is clearly relevant to their research questions for analysis. It would be wise if the data collection and analysis were thoroughly planned and implemented by taking into account resource use and data triangulation. Moreover, aligning with the recommendation of Collins et al. (2004), sharing collected data (i.e., open data) can undoubtedly advance the research community.

Engage different research participants

The success of the design solution in the real world depends largely on various people who are directly and indirectly involved in its implementation and adoption. The engagement of different stakeholders in EDR has proven indispensable for my study and those of others (e.g., Bergdahl et al., 2018; Cowling & Birt, 2018). For instance, it can be vital for understanding the complex problems in real-world settings, developing a solution to meet the needs of those involved, and achieving respondent triangulation. Thus, I agree with McKenney and Reeves (2019) that different types of participants, including direct users (i.e., teachers and students) and other relevant stakeholders (e.g., schools, parents, and policymakers), should be appropriately engaged in different phases of EDR. In line with Herrington et al. (2007), stakeholder participation should aim to benefit both the EDR (i.e., scientific and practical outputs) and the participants (i.e., societal outputs). Moreover, participation often requires intensive and long-term collaboration, thereby yielding possible difficulties in recruitment and execution for practical and ethical reasons. Thus, the collaboration should be carefully planned and implemented by taking into account relevant issues, such as research permission, social responsibility (e.g., interruptions to the participants' normal activities), and suitable timing for all involved.

Collaborate with other researchers and disciplines

It is possible for a researcher to undertake EDR alone, as in Di Biase's (2020) case. However, solving multifaceted problems or developing complex solutions, such as technological innovations, often requires a larger number of people from different disciplines. I support the recommendation of other scholars (e.g., Kennedy-Clark, 2013; McKenney & Reeves, 2019) that a multidisciplinary research team can improve the feasibility, rigour, robustness, and trustworthiness of EDR. For example, my research certainly benefited from the skills and expertise of my supervisors (in mathematics and pedagogy) and the development team (in computer science). Good teamwork (e.g., a shared vision and understanding, strong group cohesion, and respect for others) and communication can promote successful collaboration (McKenney & Reeves, 2019).

Explore alternatives, build, and test sensibly

It is essential in the design field to explore alternative ideas before selecting promising ones for further development (e.g., Ulrich & Eppinger, 2016). However, only a few EDR researchers have explicitly reported working with alternative solutions (Lehtonen et al., 2019; Ørngreen, 2015). Exploring alternative solutions in the early stages, as evidenced in my study, can help to ensure that the solution developed with considerable resources is the best answer to the educational problem (McKenney & Reeves, 2019; Ørngreen, 2015). Moreover, I agree with Easterday et al. (2017) that researchers should employ construction and evaluation methods appropriate for their theoretical knowledge level and design stage to use resources efficiently. For example, early in my study, the mock-up of each concept was simply built and evaluated only via teacher interviews and questionnaires to identify the issues to be solved and to quickly exclude unsuccessful concepts. Later on, the prototypes of the selected concept were carefully built and evaluated by teachers and students through various methods to validate the efficacy of the developed theory and solution.

Be ready for changes

EDR is flexible and adaptive in nature. It is conducted in complex real-world contexts full of variables, unlike laboratory settings. Moreover, unpredictable changes consistently occur during the iterative process of investigation, design, testing, and refining. Thus, the research design of later cycles usually needs to be adjusted based on the results from previous cycles (e.g., Herrington et al., 2007; Plomp, 2013). Often, there are other situations in which adjustments or changes are required. In my case, for example, a teacher withdrew from the field research on short notice, and the working prototype developed in collaboration with the development team did not function reliably for class interventions. Thus, I agree with Kennedy-Clark (2013) that EDR researchers should be prepared to react promptly to adjustments and changes.

While this paper appears to contribute to the design framework and methodological knowledge of EDR in technology-enhanced learning, the presented framework and guidelines were built on only one researcher's experience from a single EDR. Therefore, more similar studies by other researchers could help validate the results of this paper and advance the design framework and methodological knowledge of EDR.

6.0 References

All European Academies. (2017). *European code of conduct for research integrity* (Revised ed.). https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

Andamon, J. C., & Tan, D. A. (2018). Conceptual understanding, attitude and performance in mathematics of grade 7 students. *International Journal of Scientific & Technology Research, 7*(8), 96–105.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Beaudouin-Lafon, M., & Mackay, W. E. (2012). Prototyping tools and techniques. In J. A. Jacko (Ed.), *Human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (3rd ed., pp. 1081–1104). CRC Press.

Bedir, G., & Özbek, Ö. Y. (2016). Primary school teachers' views on the preparation and usage of authentic material. *Higher Education Studies, 6*(3), 138–146. https://doi.org/10.5539/hes.v6n3p138

Bergdahl, N., Knutsson, O., & Fors, U. (2018). Designing for engagement in TEL – a teacher-researcher collaboration. *Designs for Learning, 10*(1), 100–111. https://doi.org/10.16993/dfl.113

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31,* 21–32.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Science, 13*(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2

Cowling, M., & Birt, J. (2018). Pedagogy before technology: A design-based research approach to enhancing skills development in paramedic science using mixed reality. *Information, 9*(2), Article 29. https://doi.org/10.3390/info9020029

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Di Biase, R. (2020). Using design-based research to explore the influence of context in promoting pedagogical reform. *Educational Design Research, 4*(2), Article 21. https://doi.org/10.15460/eder.4.2.1427

Easterday, M. W., Lewis, D. G. R., & Gerber, E. M. (2017). The logic of design research. *Learning: Research and Practice, 4*(2), 131–160. https://doi.org/10.1080/23735082.2017.1286367

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences, 11*(1), 105–121. https://doi.org/10.1207/S15327809JLS1101_4

Finnish National Agency for Education. (2016). *National core curriculum for basic education 2014* (Publication 2016:5). https://www.ellibslibrary.com/book/9789521362590

Finnish National Board on Research Integrity. (2009). *Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review*. https://tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf

Finnish National Board on Research Integrity. (2012). *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*. https://tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf

Goff, W. M., & Getenet, S. (2017). Design based research in doctoral studies: Adding a new dimension to doctoral research. *International Journal of Doctoral Studies, 12*, 107–121. https://doi.org/10.28945/3761

Herrington, J., McKenney, S., Reeves, T., & Oliver, R. (2007). Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. In C. Montgomerie & J. Seale (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007* (pp. 4089–4097). AACE.

Ishii, H., & Ullmer, B. (2012). Tangible user interfaces. In J. A. Jacko (Ed.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (3rd ed., pp. 465–490). Taylor & Francis.

Joutsenlahti, J., & Kulju, P. (2017). Multimodal languaging as a pedagogical model: A case study of the concept of division in school mathematics. *Education Sciences, 7*(1), Article 9. https://doi.org/10.3390/educsci7010009

Kelly, A. E. (2013). When is design research appropriate? In T. Plomp & N. Nieveen (Eds.), *Educational design research, Part A: An introduction* (pp. 135–150). Netherlands Institute for Curriculum Development.

Kennedy-Clark, S. (2013). Research by design: Design-based research and the higher degree research student. *Journal of Learning Design, 6*(2), 26–32. https://doi.org/10.5204/jld.v6i2.128

Kerslake, L. S. (2019). *The playground of ideas: A design-based research investigation into dialogic thinking with six- and seven-year-old children in England* [Doctoral dissertation, University of Cambridge]. Apollo.

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

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70.

Lambert, D. G., & Jacobsen, M. D. (2019). Implementing an intervention into a grade six learning environment: A design-based research framework. *Educational Design Research, 3*(1), Article 20. https://doi.org/10.15460/eder.3.1.1388

Lehtonen, D., & Joutsenlahti, J. (2017). Using manipulatives for teaching equation concepts in languaging-based classrooms. In N. Pyyry, L. Tainio, K. Juuti, R. Vasquez & M. Paananen (Eds.), *Changing subjects, changing pedagogies: Diversities in school and education* (pp. 164–185). Finnish Research Association for Subject Didactics.

Lehtonen, D., Jyrkiäinen, A., & Joutsenlahti, J. (2019). A systematic review of educational design research in Finnish doctoral dissertations on mathematics, science, and technology education. *International Journal on Math, Science and Technology Education, 7*(3), 140–165. https://doi.org/10.31129/LUMAT.7.3.399

Lehtonen, D., Machado, L., Joutsenlahti, J., & Perkkilä, P. (2020). The potentials of tangible technologies for learning linear equations. *Multimodal Technologies and Interaction, 4*(4), Article 77. https://doi.org/10.3390/mti4040077

Lesh, R., Post, T., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of representations in the teaching and learning of mathematics* (pp. 33–40). Lawrence Erlbaum.

Marshall, L., & Swan, P. (2008). Exploring the use of mathematics manipulative materials: Is it what we think it is? In J. Renner, J. Cross & L. McCormack (Eds.), *Proceedings of EDU-COM 2008 International Conference – Sustainability in higher education: Directions for change* (pp. 338–350). Edith Cowan University.

McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research* (2nd ed.). Routledge.

McNeil, N. M., Hornburg, C. B., Devlin, B. L., Carrazza, C., & McKeever, M. O. (2019). Consequences of individual differences in children's formal understanding of mathematical equivalence. *Child Development, 90*(3), 940–956. https://doi.org/10.1111/cdev.12948

McNeil, N. M., & Jarvin, L. (2007). When theories don't add up: Disentangling the manipulatives debate. *Theory into Practice, 46*(4), 309–316.

Neber, H. (2012). Discovery learning. In N. Seel (Ed.), *Encyclopedia of the sciences of learning* (1st ed., pp. 1009–1012). Springer.

Organisation for Economic Cooperation and Development. (2017). *Education at a glance 2017: OECD indicators.* OECD Publishing. http://dx.doi.org/10.1787/eag-2017-en

Organisation for Economic Cooperation and Development. (2020). *Education policy outlook: Finland*. www.oecd.org/education/policy-outlook/country-profile-Finland-2020.pdf

Otten, M., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2019). The balance model for teaching linear equations: A systematic literature review. *International Journal of STEM Education, 6*, Article 30. https://doi.org/10.1186/s40594-019-0183-2

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research, Part A: An introduction* (pp. 11–50). Institute for Curriculum Development.

Poon, K., & Leung, C. (2010). Pilot study on algebra learning among junior secondary students. *International Journal of Mathematical Education in Science and Technology, 41*(1), 49–62. https://doi.org/10.1080/00207390903236434

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Slavin, R. E. (2010). Cooperative learning. In P. Peterson, E. Baker & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 177–183). Elsevier.

Toptaş, V., Çelik, S., E., & Karaca, T. (2012). Pedagogical materials use of primary grade teachers in mathematics education. *Elementary Education Online, 11*(4), 1121–1130.

Ulrich, K. T., & Eppinger, S. D. (2016). *Product design and development* (6th ed.). McGraw-Hill Education.

Uttal, D. H., Amaya, M., del Rosario Maita, M., Hand, L. L., Cohen, C. A., O'Doherty, K., & DeLoache, J. S. (2013). It works both ways: Transfer difficulties between manipulatives and written subtraction solutions. *Child Development Research, 2013,* 1–13. https://doi.org/10.1155/2013/216367

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Watson, D. M. (2001). Pedagogy before technology: Re-thinking the relationship between ICT and teaching. *Education and Information Technologies, 6*(4), 251–266. https://doi.org/10.1023/A:1012976702296

Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. *Journal of Computers in Education, 2*(4), 399–420. https://doi.org/10.1007/s40692-015-0036-z

Ørngreen, R. (2015). Reflections on design-based research: In online educational and competence development projects. In J. Abdelnour Nocera, B. R. Barricelli, A. Lopes, P. Campos & T. Clemmensen (Eds.), *Human work interaction design: Analysis and interaction design methods for pervasive and smart workplaces* (pp. 20–38). Springer.

Author Profile

Daranee Lehtonen is a doctoral researcher and instructor at the Faculty of Education and Culture, Tampere University, Finland. She has a background in design, particularly educational design, and primary school education. Her research interests include educational technology, technology-enhanced learning, digitalisation, and multimodality in different primary school subjects.

Author Details

Daranee Lehtonen
Tampere University
Kalevantie 4
33100 Tampere
Finland
+358504377415
daranee.lehtonen@tuni.fi

Editor Details

Prof. Dr. Tobias Jenert
Chair of Higher education and Educational Development
University of Paderborn
Warburger Straße 100
Germany
+49 5251 60-2372
Tobias.Jenert@upb.de

Journal Details

EDeR – Educational Design Research
An International Journal for Design-Based Research in Education
ISSN: 2511-0667
uhh.de/EDeR
#EDeRJournal (our hashtag on social media services)

Published by

Hamburg Center for University Teaching and Learning (HUL)
University of Hamburg
Schlüterstraße 51
20146 Hamburg
Germany
+49 40 42838-9640
+49 40 42838-9650 (fax)
EDeR.HUL@uni-hamburg.de
hul.uni-hamburg.de

In collaboration with

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek Hamburg – Landesbetrieb
Von-Melle-Park 3
20146 Hamburg
Germany
+49 40 42838 7146
info.hup@sub.uni-hamburg.de
hup.sub.uni-hamburg.de