UNIVERSITY OF TURKU

USING GAME-BASED LEARNING TO ENHANCE ADAPTIVE NUMBER KNOWLEDGE

Boglárka Brezovszky

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. B OSA - TOM. 476 | HUMANIORA | TURKU 2019

UNIVERSITY OF TURKU

USING GAME-BASED LEARNING TO ENHANCE ADAPTIVE NUMBER KNOWLEDGE

Boglárka Brezovszky

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA - SER. B OSA - TOM. 476 | HUMANIORA | TURKU 2019

University of Turku

Faculty of Education
Department of Teacher Education and Centre for Research on Learning and Instruction
Doctoral Programme on Learning, Teaching and Learning Environments Research

Supervised by

Professor Erno Lehtinen
Centre for Research on Learning and Instruction
University of Turku
Finland

Senior Researcher Koen Veermans
Centre for Research on Learning and Instruction
University of Turku
Finland

Associate Professor
Minna Hannula-Sormunen
Centre for Research on Learning and Instruction
University of Turku
Finland

Reviewed by

Professor Wouter van Joolingen
Utrecht University
The Netherlands

Professor Markus Hähkiöniemi
Adjunct Professor
University

Opponent

Professor Wouter van Joolingen
Utrecht University
The Netherlands

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-7621-8 (PRINT)
ISBN 978-951-29-7622-5 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Grano Oy - Turku, Finland 2019

TIIVISTELMÄ

TURUN YLIOPISTO
Kasvatustieteiden tiedekunta
Opettajankoulutuslaitos ja Oppimisen ja opetuksen tutkimuskeskus
Oppimisen, opetuksen ja oppimisympäristöjen tutkimuksen tohtoriohjelma

BOGLÁRKA BREZOVSZKY: ADAPTIIVISEN LUKUKÄSITTEEN TUKEMINEN OPPIMISPELIN AVULLA

Tämän väitöskirjatutkimuksen tavoitteena oli tutkia, miten oppimispeliä voisi käyttää joustavan ja adaptiivisen lukukäsitteen kehityksen tukemisessa. Väitöskirja koostuu viidestä osatutkimuksesta, jotka käsittelevät Number Navigation pelin (NNG) kehittämistä, testaamista sekä pelin vaikutuksia adaptiiviseen lukukäsitteeseen ja muihin alakoulun oppilaille tärkeisiin matemaattisiin taitoihin ja tietoihin. Tutkimuksissa käytetään erilaisia menetelmiä ja tarkastellaan erilaisia tutkimuskysymyksiä pelin kehittämisen ja testaamisen eri vaiheissa. Keskeisenä punaisena lankana on kuitenkin se, miten pelimekaniikka ja matemaattinen sisältö ovat yhteydessä toisiinsa ja miten pelin avulla toteutettu harjoittelu on yhteydessä oppimistuloksiin.

Tutkimuksessa 1 testattiin NNG:n ensimmäisen prototyypin toimivuutta käyttäjien kokemusten näkökulmasta ja tutkittiin niitä peliprosesseja, joiden voi olettaa tukevan adaptiivisen lukukäsite kehittymistä. Peliä tutkittiin kahden pelitilanteen avulla. Ensimmäisessä 11-vuotias oppilas pelasi peliä yksinään ja toisessa 9- ja 11-vuotiaan oppilaat pelasivat peliä yhdessä. Tutkimusaineisto koostui pelisessioiden videoinneista, tietokoneen ruudunkaappauksen tallennuksesta sekä pelikokemusta kartoittaneesta avoimesta haastattelusta. Tulokset osoittivat, että jo pelin varhainen prototyyppi sai aikaan aktiivista sitoutumista testaamaan erilaisia lukujen ja operaatioiden suhteita ja yhdistelmiä sekä keskustelemaan niistä. Tulokset viittasivat siihen, että NNG:n pelaaminen voi herättää sellaista matemaattista ajattelua ja ongelmanratkaisua, jonka oletettiin edistävän oppilaiden adaptiivista lukukäsitettä.

Tutkimus 2 koostui kolmesta osatutkimuksesta (5.luokkalaiset, n = 55, yliopisto-opiskelijat, n = 55 ja 6.luokkalaiset, n = 22) ja sen tavoitteena oli kehittää ja testata aritmeettisten lausekkeiden tuottamisen tehtävää, joka on kynä-paperi-tehtävä, jolla mitataan adaptiivista lukukäsitettä. Tässä tutkimuksessa tutkittiin sekä yksilöllisiä eroja aritmeettisten lausekkeiden tuottamistehtävässä että adaptiivisen lukukäsitteen ja muiden matemaattisten taitojen välistä yhteyttä. Tulokset osoittivat yksilöllisten erojen olevan samanlaisia eri otoksissa ja ikäryhmissä. Adaptiivisen lukukäsitteen taso oli yhteydessä aritmeettiseen sujuvuuteen ja käsitteelliseen aritmeettiseen tietoon. Tulosten mukaan aritmeettisten lausekkeiden tuottamisen tehtävällä pystyttiin tunnistamaan, kuinka osallistujat tunnistavat mahdollisuuksia käyttää lukujen eri ominaisuuksia aritmeettisessa ongelmanratkaisussa, mitä voidaan pitää adaptiivisen lukukäsitteen indikaattorina.

Tutkimus 3 oli ensimmäinen yritys tutkia NNG:n vaikutuksia ja pelisuorituksen ja adaptiivisen lukukäsitteen yhteyttä. Tutkimuksessa oli kvasi-kokeellinen alku- ja loppumittausasetelma. Osallistujat olivat 11 paria kuudesluokkalaisia oppilaita (11 tyttöä, ikä 11-13 vuotta) yhdestä luokasta, joka pelasi NNG-peliä 7 viikkoa. Tulokset osoittivat osallistujien adaptiivisen lukukäsitteen ja aritmeettisen sujuvuuden kehittyneen. Pelisuoriutuminen ennusti oppilaiden lukukäsitteen monioperationaalista aspektia loppumittauksessa.

Tutkimus 4 on yksityiskohtainen kuvaus NNG:n taustalla olevista pelin suunnitteluperiaatteista teorioihin ja aikaisempiin NNG:tä käyttäneisiin pilottitutkimuksiin (tutkimukset 1 ja 2) perustuen. Tämän tutkimuksen jälkeen lopullinen NNG-prototyyppi suunniteltiin tutkimukseen 5.

Pilotti- ja kehitysvaiheiden jälkeen tutkimus 5 kohdistui NNG:n vaikutusten tutkimiseen luuokkahuoneessa ja isommalla otoksella ja useilla luokka-asteilla. Tutkimus 5:ssä tutkittiin, miten NNG:llä harjoittelu vaikuttaa alakoululaisten adaptiiviseen lukukäsitteeseen, sujuvuuteen ja esi-algebraa koskevaan tietoon eri luokka-asteilla, sekä kuinka oppilaiden pelisuoritus NNG:ssä vaikuttaa matemaattisiin oppimistuloksiin. Alku- ja loppumittauksellinen satunnaistettu koe- ja kontrolliryhmäkokeilu oli tutkimuksen design. Koeryhmä pelasi NNG-peliä 10 viikon ajan matematiikan tunneilla ja kontrolliryhmä osallistui normaaliin matematiikan opetukseen. Tulokset osoittivat koeryhmän kehittyneen merkitsevästi paremmin kaikissa matemaattisissa testeissä kuin kontrolliryhmä. Pelisuoriutuminen oli yhteydessä koeryhmän loppumittaustuloksiin myös alkumittauksen pisteiden ja luokkatason kontrolloinnin jälkeen. Lisäksi tulokset osoittivat, että NNG kehittää lukukäsitteen eri aspekteja, aritmeettista sujuvuutta ja esi-algebrallisia tietoja eri tavalla eri luokka-asteilla.

Kaiken kaikkiaan tämän väitöskirjan löydökset osoittavat, että NNG voi tarjota alakouluikäisille tehokkaita ja uusia mahdollisuuksia harjoitella joustavia matemaattisia taitoja ja niihin liittyvää adaptiivista, joustavaa lukukäsitettä. Peliympäristö mahdollisti suuren harjoitusmäärän avoimilla tehtävillä ja lukuisilla erilaisilla luku-operaatio-yhdistelmillä, mikä on tavallisesti vaikeasti saavutettavissa tyypillisessä matematiikan opetuksessa luokkahuoneessa. Väitöskirjan tulokset antavat empiiristä tukea teoreettiselle mallille adaptiivisesta lukukäsitteestä joustavan aritmeettisen osaamisen osatekijänä. Lisäksi väitöskirjatutkimuksissa kehitettiin uusi, skaalautuva mittari ja uusi harjoittelumuoto adaptiiviselle lukukäsitteelle. Tulokset korostavat iteratiivisen design-prosessin merkitystä oppimispelien kehittämisessä niin, että kasvatuksellinen oppimissisältö ja pelimekaniikka integroidaan, mittaamiskysymykset ratkaistaan ja mahdollistetaan harjoittelu pelissä.

ABSTRACT

UNIVERSITY OF TURKU
Faculty of Education
Department of Teacher Education and Centre for Research on Learning and Instruction
Doctoral Programme on Learning, Teaching and Learning Environments Research

BOGLÁRKA BREZOVSZKY: USING GAME-BASED LEARNING TO ENHANCE ADAPTIVE NUMBER KNOWLEDGE

The aim of this dissertation was to explore how game-based learning can be used to develop primary school students' adaptive number knowledge. The dissertation comprises five studies that explore the Number Navigation Game's (NNG) development, testing and effects in enhancing primary school students' adaptive number knowledge and related mathematical skills and knowledge. The studies use different methodologies and ask varying questions through the process of development and testing of the NNG. However, as a common thread, each study explores the relationship between game mechanics and mathematical content, and the relationship between game-based training and learning-outcome measures.

Study I tested the NNG's first working prototype from a user experience point of view and investigated instances of gaming processes which may support the development of adaptive number knowledge. The game was tested in two sessions, with one student (age 11) playing individually and two students (ages 9 and 11) using the game collaboratively. Data were collected using video-recorded observations, gameplay screen capture and open-ended interview questions regarding the game experience. Results showed that already, the NNG prototype triggered players' active engagement with different number combinations and numerical relations as well as reflection and discussion about these numerical relations. This suggested that gameplay with the NNG could trigger the type of mathematical thinking and problem solving which was expected to help enhance students' development of adaptive number knowledge.

Study II comprised three sub-studies (fifth-graders, n = 55; university students, n = 55; and sixth-graders, n = 22) and aimed to develop and test the Arithmetic Production Task, a paper-and-pencil measure of adaptive number knowledge. Individual differences in the Arithmetic Production Task and the relation between adaptive number knowledge and other mathematical skills and knowledge were explored. Results showed similar patterns of individual differences across the different samples and age groups. Adaptive number knowledge was found to be related to arithmetic fluency and conceptual knowledge of arithmetic. Results suggest that the Arithmetic Production Task was able to capture how participants recognise opportunities to use numerical features in their arithmetic problem solving, which can be used as an indicator of their adaptive number knowledge.

Study III was the first attempt to test the NNG's effects and examine the relationship between game performance and adaptive number knowledge. The study used a quasi-experimental pre-test/post-test design. Participants were 11 pairs of sixth-grade students (11 females, age range: 11-13) from one classroom who played the NNG over a seven-week period. Results showed improvement in

participants' performance in adaptive number knowledge and math fluency. Game performance was found to be a predictor of students' post-test performance on the multi-operational aspect of adaptive number knowledge.

Study IV provides a detailed overview of game design decisions in the NNG, which are based on theory and the results from previous pilot studies using the NNG (Studies I and III). Another game-development phase followed Study IV, resulting in the final NNG prototype used in Study V.

After the initial pilot and development stages, Study V aimed to test the NNG's effects in the classroom and on a larger scale across different grade levels. Study V explored how training with the NNG affects the development of primary school students' adaptive number knowledge, arithmetic fluency and pre-algebra knowledge at different grade levels; and how students' performance on the NNG affects the development of the mathematical learning outcomes. A pre-test/post-test randomised control design was used, in which the experimental group played the NNG for 10 weeks during math class, and the control group received regular math instruction. Results showed significant overall improvement in the experimental group compared with the control group on all measured mathematical learning outcomes. Game performance was related to the experimental group's post-test scores even after controlling for pre-test scores and grade. Additionally, results showed that the NNG develops different aspects of adaptive number knowledge, math fluency and pre-algebra knowledge across grade levels.

Overall, the present dissertation's findings show that the NNG can provide efficient and novel training opportunities for developing primary school students' adaptive number knowledge and related mathematical skills and knowledge. The game-based format allowed for a large amount of open-ended practice with various number-operation combinations, which usually is difficult to achieve in regular mathematics classrooms. The dissertation's results provide empirical evidence for the theoretical model of adaptive number knowledge as a component of adaptivity with arithmetic. Additionally, a new and scalable measurement and novel training of adaptive number knowledge were developed. The results also highlight the importance of an iterative design process in educational game development, with the following guiding principles: integration of educational content and game features, measurement questions and application of training.

ACKNOWLEDGEMENTS

This PhD work was a truly collaborative process, and I am so grateful that I could learn and develop so much as part of a team creating something meaningful.

Thank you to my supervisor Professor Erno Lehtinen for taking me on board and offering opportunities during my PhD work and already during my MA studies. Working in your project and on the Number Navigation Game (NNG) was a truly rewarding journey. I appreciate the freedom, flexibility and security that allowed me to focus on my work fully. I am grateful for your trust and encouragement. Thank you for pushing me when it was needed and for always taking time to answer my questions and listen to my concerns.

To my supervisor, Associate Professor Minna Hannula-Sormunen, thank you for all your support, encouragement, honest discussions and friendly smiles. Thank you for your help at critical times during the data-gathering phase of the NNG project. You not only showed me true focus, determination and dedication, but you also helped me reflect on the importance of work-life balance.

Thank you to my supervisor, Dr. Koen Veermans, for the uncountable times you challenged my brain to work to its fullest capacity and sometimes beyond. Thank you for never giving a ready answer and for showing the importance of critical thinking in—and out—of academia. Thank you for your help and support with different methods of data analyses, and for taking me on as a teacher of the methodology course. Year after year, it was challenging but also largely rewarding to be part of the course and the LLEES (now EdLearn) multicultural community.

Thank you Professor Markus Hähkiöniemi and Professor Wouter van Joolingen for being the external reviewers of my dissertation. Additionally, Professor van Joolingen, I greatly appreciate your acceptance to be the opponent at my doctoral defence.

I always felt very lucky to be part of the larger process that developed an idea into the complex and meaningful product that is the Number Navigation Game. I am grateful for learning from the shared work and effort and for working with so many wonderful people during these years. It has been a pleasure to work with you Dr. Jake McMullen; your work on developing the concept and measures of adaptive number knowledge was indispensable to this dissertation. Thank you for providing your 2016 publication to be part of this work. I appreciate our many discussions on adaptivity and flexibility with arithmetic, the good times at conferences, your many comments and revisions, and all your help during my PhD.

Dr. Nonmanut Pongsakdi, I was lucky enough to wander those first years with you. Thank you for your wisdom, kindness and all your help during these years. Dr. Gabriela Rodríguez-Aflecht, I am thankful to have worked with you on the main stages of developing and testing of the NNG. My thanks go to Dr.

Tarja-Riitta Hurme for helping us with managing the practicalities of the large-scale NNG project. I am grateful to the developers who made the NNG a reality, the testers and coders, and all the schools and students who made this work possible.

Senior researcher Eero Laakkonen, thank you for providing help during the revisions of my last publication, which finally allowed this PhD to be completed after a much-awaited time.

I am grateful for the financial support received from the Academy of Finland CUMA project (PI: Professor Erno Lehtinen), the Academy of Finland's DECIN project (PI: Associate Professor Minna Hannula-Sormunen), and funding received from the Doctoral Programme on Learning, Teaching and Learning Environments Research (OPPI) at the University of Turku Graduate School.

Apart from the NNG community, I was fortunate to be part of a larger team of PhD candidates and staff members. I am thankful for having worked at the Centre for Research on Learning and Instruction within the Department of Teacher Education, University of Turku, Finland. I am grateful to the mathematics research group and our Wednesday seminars. Thank you to Maikki, Erkka, Phoung, Jing, Diana, Cristina, Emmanuel and so many others for the lunches, coffees and good laughs! Dear Pamela, I am convinced that words would not do justice here. You are the sunshine after rain. Thank you so much for everything!

I feel grateful and lucky to have parents who always supported me and gave me the freedom and luxury to follow my path even if that came with sacrifices from their side. I am thankful to my brother who so many times dealt with the reality of life while I was busy following this path. To my dearest bunch of people who destroy any last crumb of sentimentalism, thank you for the now-approaching 20 years of friendship, and for always being there. I know you know what I mean.

Dear Tamás and Lili, you are the best team ever! None of this would make any sense without the two of you. With all my heart, thank you!

11.03.2019, Turku

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This doctoral thesis is based on the following five studies reported in five original articles, referred to in the text by their roman numerals. In Studies I, II and V, Boglárka Brezovszky was first author and contributed to study conception and design, data collection, analyses and interpretation, and was responsible for writing the manuscripts. In Studies II and IV, Boglárka Brezovszky was second author and contributed to the studies' conception and design, data collection and interpretation, and worked on both manuscripts' writing and revision process. All co-authors provided critical feedback and contributed to all manuscript revisions.

Study I Brezovszky, B., Lehtinen, E., McMullen, J., Rodriguez, G. & Veermans, K. (2013). Training flexible and adaptive arithmetic problem-solving skills through exploration with numbers: The development of Number Navigation game. In *7th European Conference on Games Based Learning, ECGBL 2013* (Vol. 2, pp. 626–634).

Study II McMullen, J., Brezovszky, B., Rodríguez-Aflecht, G., Pongsakdi, N., Hannula-Sormunen, M. M. & Lehtinen, E. (2016). Adaptive number knowledge: Exploring the foundations of adaptivity with whole-number arithmetic. *Learning and Individual Differences*, 47, 172–181. https://doi.org/10.1016/j.lindif.2016.02.007

Study III Brezovszky, B., Rodríguez-Aflecht, G., McMullen, J., Veermans, K., Pongsakdi, N., Hannula-Sormunen, M. M. & Lehtinen, E. (2015). Developing adaptive number knowledge with the number navigation game-based learning environment. In J. Torbeyns, E. Lehtinen & J. Elen (Eds.), *Describing and Studying Domain-Specific Serious Games* (pp. 155–170). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20276-1_10

Study IV Lehtinen, E., Brezovszky, B., Rodríguez-Aflecht, G., Lehtinen, H., Hannula-Sormunen, M. M., McMullen, J., Pongsakdi, N., Veermans, K. & Jaakkola, T. (2015). Number Navigation Game (NNG): Design Principles and Game Description. In J. Torbeyns, E. Lehtinen & J. Elen (Eds.), *Describing and Studying Domain-Specific Serious Games* (pp. 45–61). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20276-1_4

Study V Brezovszky, B., McMullen, J., Veermans, K, Hannula-Sormunen, M., Rodríguez-Aflecht, G, Pongsakdi, N., Laakkonen, E. & Lehtinen, E. (2019). Effects of a Mathematics Game-Based Learning Environment on Primary School Students' Adaptive Number Knowledge. *Computers & Education*, 128, 63-74 https://doi.org/10.1016/j.compedu.2018.09.011

1. INTRODUCTION

Play and games are fundamental to our lives. We can create a story with its own set of rules and goals, and we learn to experiment with these different roles without fear. It is perhaps imagination that makes games so powerful in shaping the development of our thinking and learning as humans. The importance of play has been acknowledged over many different disciplines (for a review, see Sutton-Smith, 1997), and playful learning or learning by experimentation is considered to be fundamental in early cognitive development (Pellegrini, 2009; Piaget, 1952; Rogers & Sawyers, 1988). During play, we are free to construct rules and experiment with different outcomes in a safe environment. Immersion in the imaginary world of play can result in deeper engagement with the task at hand and, thus, a deeper form of learning (e.g., Csikszentmihalyi, 1991). These long-standing theoretical propositions have been tested and revised in many empirical studies over the years (for an overview, see Fisher, Hirsh-Pasek, Golinkoff, Singer & Berk, 2010; Ilgaz, Hassinger-Das, Hirsh-Pasek & Golinkoff, 2018), and the use of games and playful learning became a widespread recommendation in policy documents and curricula across various academic disciplines and age groups (e.g., Finnish National Agency for Education, 2014; OECD 2010, 2018).

With technological advancements, learning through experimentation and play was adopted into the digital context, and nowadays, the term *game-based learning* often is associated with the use of digital games that aim to produce some form of learning. Due to the affordances of technology and the long-standing traditions of using games and play for educational purposes, digital game-based learning became very popular in a short period of time. However, after the initial excitement, it is now clear that games do not represent a magic bullet. While most reviews find that game-based learning can be associated with greater learning gains compared with more traditional instruction methods, the difference is usually small (Boyle et al., 2016; Hainey, Connolly, Boyle, Wilson & Razak, 2016; Hays, 2005; Wouters, van Nimwegen, van Oostendorp & van der Spek, 2013; Young et al., 2012).

The explanation for this discrepancy between the potential of game-based learning and the lack of strong empirical results that would substantiate these claims is usually two-fold. First, despite the affordances of technology to provide new methods to interact with educational content, most technology-supported applications are still used only as alternative content-delivery methods (Bray & Tangney, 2017). In the specific domain of game-based learning, this means that learning content is often not delivered through gameplay, but that the gameplay is used as an incentive to make students interact with educational content (Devlin, 2011; Habgood & Ainsworth, 2011; Habgood, 2007). This can create frustration and disappointment, and can explain why game-based learning is not found to be more motivating for students than traditional instructional methods (Wouters et al., 2013), as well as why most game-based interventions are only marginally effective (Hainey et al., 2016; Hays, 2005; Wouters et al., 2013). Second, a disconnect often exists between the type of skills and knowledge practised within the context of a game-based learning environment and the means and methods of measuring expected learning outcomes (All, Nuñez Castellar & Van Looy, 2014, 2015, 2016). Measurement problems, such as using standardised tests to measure the outcomes of complex game-based training effects, relying solely on game-performance measures, overuse of quasi-experimental designs and a lack of large-scale randomised studies are often raised as problems that make the estimation of game-based learning's real effects difficult (All et al.,2014, 2016; Connolly, Boyle, Hainey & Boyle, 2012; de Freitas & Oliver, 2006; Wouters et al., 2013; Young et al., 2012).

Thus, to make use of the potential affordance of games for educational purposes, it is important to design the interaction of learning content and game features carefully and implement appropriate measures that can capture different aspects of intended learning outcomes. Achieving this balance requires an iterative process of design and testing, with game and learning interacting at the centre (All et al., 2015; Vanden Abeele et al., 2012). However, in reality, this is rarely done, and most attempts to use game-based learning for educational purposes remain very limited in scope. To achieve coherence among game features, educational content and assessment, it is important to examine how game-based learning can be used to provide a new method of interaction with the learning content that was unavailable or difficult to achieve in a different format, as well as how to measure this novel type of knowledge within and outside the game context.

For example, in the domain of mathematics, flexibility and adaptivity with arithmetic are strongly highlighted curricular objectives in many countries, but research shows that direct teaching of strategies in the classroom doesn't lead to flexible and adaptive problem solving in new contexts and tasks (Hickendorff, 2018). Therefore, it is both theoretically and practically relevant to explore how game-based learning can be used to develop flexibility and adaptivity with arithmetic that are transferable to novel tasks.

Flexibility and adaptivity with arithmetic problem solving are often operationalised in terms of a large repertoire of problem-solving strategies and the ability to select appropriate strategies for a given problem (Heinze, Star & Verschaffel, 2009; Verschaffel, Luwel, Torbeyns & Van Dooren, 2009). Developing flexibility and adaptivity with arithmetic can be challenging because: 1) even if students have a large repertoire of strategies, they often fail to notice opportunities to apply them in practice (Canobi, Reeve & Pattison, 2003), and 2) a strategy's efficiency or appropriateness is very subjective and context-dependent (Heinze et al., 2009; Threlfall, 2002, 2009). Thus, training should include attempts to strengthen students' underlying understanding of numerical relations, which are expected to help identify opportunities when using alternative solutions can make problem-solving more efficient (Baroody, 2003; Lehtinen, Hannula-Sormunen, McMullen & Gruber, 2017; Verschaffel et al., 2009).

Adaptive number knowledge is a component of adaptivity with arithmetic and is characterised by a well-connected representation of numerical characteristics and relations (McMullen et al., 2017). Strengthening students' adaptive number knowledge requires an extended amount of practising with different combinations of numbers and operations, solving arithmetic problems in multiple ways and reflecting on solutions and their underlying arithmetical relations (Baroody, 2003; Heinze et al., 2009). Regular instruction offers limited opportunities for this type of extended practice, but game-based learning's affordances can be ideal for incorporating a large amount of practice with open-ended arithmetic problems. Additionally, the game's rules and goals can provide a meaningful context to trigger students' engagement in looking for alternative solutions, as well as their reflection on these solutions.

The present dissertation aimed to explore, through a series of different methods and cycles of development and testing, how game-based learning can be used to trigger the type of open-ended mathematical practice that is considered foundational in developing adaptivity with arithmetic problem solving, but which is difficult to achieve in the regular school context. The dissertation consists of five studies that show interactions among learning content, assessment and game design during the cyclical development and testing of the Number Navigation Game (NNG).

1.1 Game-based learning: definitions and terminology

There is a large variety of games or game-based learning environments that use different genres and design principles, and target a wide variety of learning outcomes, such as skills, conceptual understanding, motivation, behaviour, etc. (for a review, see Boyle et al., 2016; Connolly et al., 2012; Hays, 2005). This variation makes it challenging to provide one definition of *game-based learning*. Usually, it is agreed that rules, goals and some quantifiable outcomes are the basic, necessary components of game-based learning environments (Salen & Zimmerman, 2003), but many other features – e.g., competition, challenge, exploration and fantasy – arguably are necessary to call a game a game (Whitton, 2010), especially when it comes to digital games. The present dissertation does not intend to go into detailed discussions about the definition of *games* in general, but will assume that the following are necessary features to consider when referring to a digital game: well-defined goals and rules, interactivity, challenge and feedback. This criteria set is common in many definitions that different authors have provided in this domain (e.g., Hays, 2005; Salen & Zimmerman, 2003; Wouters et al., 2013)

Apart from issues related to the definition of *game-based learning*, extensive debate has been sparked regarding terminology usage. Common terms in extant literature include: *serious games, instructional games, learning games* and *computer games* (Hays, 2005; Plass, Homer & Kinzer, 2015; Squire, 2008; Tobias & Fletcher, 2012). Sometimes these terms are used to differentiate constructs, sometimes overlap exists and sometimes different terms are used to describe the same construct (Boyle et al., 2016; Breuer & Bente, 2010; Crookall, 2010; Wouters et al., 2013). In the common practice of educational research, the terms *educational digital games, game-based learning environments* and *serious games* are often used interchangeably (Boyle et al., 2016), and they mostly refer to the same type of learning environment, which uses technology in a game-based format to support learning.

As the present dissertation describes a process of developing and testing the NNG game-based learning environment, it was important that the terminology used be inclusive enough and able to describe the NNG at all stages of development, including simple prototypes. Thus, the term *game-based learning environment* (for specific training) and *game-based learning* (for the domain in general) were selected. In the present dissertation, the term *game-based learning environment* refers to a digital game developed specifically with intended educational purposes in mind (Boyle et al., 2016; Plass et al., 2015; Torbeyns, Lehtinen & Elen, 2015; Wouters et al., 2013). The term *digital* (as in *digital game-based learning*) is assumed to be understood and has been omitted throughout the dissertation for reasons of parsimony.

1.2 Development and assessment in game-based learning

Overall, reviews and meta-analyses suggest that game-based learning can be more effective than traditional instruction methods (Cheung & Slavin, 2013; Clark, Tanner-Smith & Killingsworth, 2016;

Connolly et al., 2012; Hainey et al., 2016; Hays, 2005; Kebritchi & Hirumi, 2008; Wouters & van Oostendorp, 2013; Wouters et al., 2013). However, these effects are usually small and questionable for the following reasons: large variations in study designs and control types used; lack of stable terminology, which elicits wide variation in the game-based interventions used; and a lack of large-scale, randomised studies, eliciting difficulties in conducting proper meta-analyses. Recommendations from different reviews and meta-analyses on how to overcome these problems include: developing terminology consistency and systematic criteria for categorising game-based learning environments; more longitudinal, large-scale, randomised, controlled studies; and alignment of learning outcomes, game mechanics and outcome measures (e.g., Connolly et al., 2012; Hainey et al., 2016; Hays, 2005; Kebritchi & Hirumi, 2008; Wouters et al., 2013). For an overview of the main findings and conclusions from some recently published major reviews and meta-analyses in the field of game-based learning, see Appendix 1.

On the other hand, the domain of game-based learning is large, and variation is to be expected. Thus, raising more specific questions regarding the relationship between specific game features and learning content, rather than questions regarding overall efficiency, can be more meaningful (Wouters et al., 2013; Young et al., 2012). This shift places design decisions at the centre of research and explores how efficiency can be conceptualised and measured in different contexts. For example, in their review of technology usage in mathematics education, Bray and Tangney (2017) propose that besides cognitive, motivational or behavioural gains, the efficiency of using technology can also be defined by the extent to which technology usage transforms how students interact with educational content or provides new methods of training that were impossible in other contexts. Adopting this perspective in the specific domain of game-based learning places emphasis on the design process and the extent to which gameplay can offer players a new method of interacting with the intended educational content.

This can be achieved by focusing on the interaction between learning content and game features, and by aiming to integrate the two in the most seamless way possible. *Integrated game design* means that learning content is delivered through the gameplay's basic structure, so that the largest percentage of students' interactions with the game equals time spent on tasks with the intended learning content (Habgood & Ainsworth, 2011; Habgood, 2007; Salen & Zimmerman, 2003; Vanden Abeele et al., 2012). If educational content and game mechanics are well-integrated, game progress happens through interaction with relevant learning content, which is expected to produce more engagement with the task at hand and stronger learning outcomes (Garris, Ahlers & Driskell, 2002; Gunter, Kenny & Vick, 2008; Habgood, 2007; Ke & Abras, 2013; Wilson et al., 2009). On the other hand, highly integrated game design can also incur costs related to the transfer of learning outcomes, as it is possible for students to become proficient at solving problems within the context of a specific game-based learning environment, but cannot connect this learned material with material in the context of the regular classroom (e.g., Leemkuil & de Jong, 2012; ter Vrugte & de Jong, 2012).

Integration is not a solution in itself, but is an important aspect to consider during the process of designing game-based learning environments because it can increase educationally relevant time-on-task during interactions with a game. In addition, game-based learning environments ideally should also provide opportunities for interactions with learning content that is difficult to include in normal classroom settings (Bray & Tangney, 2017; Devlin, 2011). This implies that measures should be selected carefully so that they can capture these novel and often more complex learning outcomes, both within and outside the game context (All et al., 2014). Finding balance in the relationship among game design, learning content and learning outcomes requires continuous iterations of testing and development. The game design framework that Vanden Abeele and colleagues (2012) suggest integrates these different aspects and provides a general framework for an iterative educational game design. The proposed model includes three phases: 1) the concept-design phase, which includes a detailed definition of *learning outcomes* and how they can be translated into the game's context; 2) the game design phase, which includes creating and testing basic prototypes of the game; and 3) the game-development phase, which includes the testing of high- and low-fidelity game versions and further development. These three phases have smaller units, but all parts are related to and rely on each other. Information is used for continuous development, and a final prototype is created and tested after phase three. Integrating game features and learning content is at the centre of the process, with testing and development done by an interdisciplinary team that focuses on the player in all stages.

1.3 Use of game-based learning in mathematics instruction

Interest has grown in the use of game-based learning environments to enhance the teaching of science, technology, engineering and math (STEM) subjects (Boyle et al., 2016), and this trend is especially strong in the domain of mathematics (Hainey et al., 2016). Generally, the use of technology in mathematics education is emphasised strongly in many curricula (NCTM 2014; Finnish National Agency for Education 2014), as it provides opportunities that can transform mathematics teaching and learning (Bray & Tangney, 2017). Numerous reviews and meta-analyses have explored the use of technology, including different forms of game-based learning, in the field of mathematics instruction (e.g., Bray & Tangney, 2017; Cheung & Slavin, 2013; Li & Ma, 2010; Slavin & Lake, 2008; Young, 2017). Most of these field-specific reviews mirror the findings from reviews published in the general domain of game-based learning and show small, but positive, effects of technology-supported instruction over more traditional instructional methods (Cheung & Slavin, 2013; Li & Ma, 2010; Slavin & Lake, 2008; Young, 2017); a lack of large-scale randomised studies (Cheung & Slavin, 2013; Li & Ma, 2010; Slavin & Lake, 2008); effect sizes shaped by the type of technology used and the study's methodological quality (Young, 2017); and issues with aligning learning outcomes' measures with complex learning objectives (Slavin & Lake, 2008).

In a recently published review in the specific domain of game-based learning and math instruction, Byun and Joung (2018) explored trends in digital game-based learning in K-12 mathematics education, including 33 studies between 2005 and 2014, out of which 17 were included in a meta-analysis. Like general reviews' findings in game-based learning, the results showed moderate effects (Cohen's d = 0.37) for game-based learning over more traditional methods of mathematics instruction. Findings regarding general trends are in line with broader reviews' conclusions on technology-enhanced mathematics instruction (e.g. Bray & Tangney, 2017; Young, 2017), showing that most game-based learning environments used were limited in scope and that the game-based format served only as an alternative content-delivery medium. Most of the game-based learning environments examined used *drill-and-practise*, which the authors defined as a '*gamified version of paper-based worksheets*' (Byun & Joung, 2018, p. 122). This format limited attainable learning outcomes to the practice of procedural fluency in regular school math tasks (i.e., numbers and operations, algebra and geometry). Very few examples exist of game-based learning environments that aim to promote more complex mathematical skills and knowledge that go beyond routine classroom math practice. Despite technological affordances, flexible, higher-order arithmetic expertise and mathematical thinking still are rarely achieved (Byun & Joung, 2018; Devlin, 2011; Ke, 2008, 2009; Seo & Bryant, 2009). Most game-based learning environments function in an augmentative (i.e., content delivery), rather than transformative manner, so that they cannot provide an alternative method of interaction with the mathematical learning content (Bray & Tagney, 2017; Young, 2017).

A few examples of previously published studies using game-based learning environments that aimed to promote more complex and deeper-level mathematical skills and knowledge (e.g., Bakker, van den Heuvel-Panhuizen & Robitzsch, 2015; Habgood & Ainsworth, 2011; Pope & Mangram, 2015; van den Heuvel-Panhuizen, Kolovou & Robitzsch, 2013) include learning objectives such as number sense (flexibility with numbers and operations, and an understanding of properties of numbers), pre-algebra skills (reasoning about relations between quantities) and multiplicative reasoning (understanding of multiplicative relations and the relationship between multiplication and division). A common feature in these game-based learning environments is that they aimed to promote students' reflection on alternative ways to solve mathematical problems and promote students' understanding and use of numerical relations and properties of numbers in their problem solving. Not surprisingly, the type of game-based learning environments used in the aforementioned studies are very different. In some, game features are limited and serve rather as a motivational factor (i.e., Bakker et al., 2015; van den Heuvel-Panhuizen et al., 2013). Conversely, some of the studies focused mainly on design aspects and the relationship between game features and mathematical content, but not so much on the relationship between these to outside-game measures of mathematical learning outcomes (Habgood & Ainsworth, 2011; Pope & Mangram, 2015).

In summary, a need seems to exist for tools that develop higher-level mathematical knowledge, such as flexibility and adaptivity with arithmetic, which entails understanding numerical relations and the use of these relations in novel arithmetic contexts (Baroody, 2003; Lehtinen et al., 2017; Nunes, Dorneles, Lin & Rathgeb-Schnierer, 2016). While the affordances of games offer many opportunities to develop complex and flexible mathematical learning environments, research shows that a disconnect exists between theoretical claims and research practice (Bray & Tangney, 2017; Byun & Joung, 2018; Young, 2017). Most game-based learning environments have limited mathematical learning objectives, the relationship between game features and mathematical learning content is problematic and measures of mathematical learning outcomes often are disconnected from the skills and knowledge practised within game contexts (Bray & Tangney, 2017; Byun & Joung, 2018; Cheung & Slavin, 2013; Li & Ma, 2010; Slavin & Lake, 2008; Young, 2017). Thus, the present dissertation aimed to develop and test a game-based learning environment that can be used flexibly to develop both basic arithmetic fluency, as well as more complex mathematical skills and knowledge, such as adaptive number knowledge and pre-algebra knowledge. Furthermore, the various studies' objectives included in this dissertation were aimed at providing a comprehensive overview of the relationships among game design, mathematical learning objectives and the measurement aspect of these new types of complex mathematical learning outcomes.

1.4 Enhancing adaptivity with arithmetic: theoretical considerations in NNG development

Developing adaptive expertise with arithmetic (Hatano & Oura, 2003) represents one of the main objectives of primary school mathematics instruction (Nunes et al., 2016; NCTM 2014). Within this framework, researchers of learning and mathematics education have highlighted flexibility and adaptivity with arithmetic problem solving (Baroody, 2003; Hatano & Oura, 2003). *Flexibility* can be described as the ability to switch between different strategies, and *adaptivity* refers to the ability to select the optimal strategy for a given problem (Baroody, 2003; Schneider, Rittle-Johnson & Star, 2011; Verschaffel et al., 2009).

Within adaptivity with arithmetic, the definition of *optimal* or *efficient problem solving* has always been a central question, as the meanings of *optimal* and *efficient* can be very different for various individuals and can depend largely on the context's norms or the particular problem's characteristics (Heinze et al., 2009; Threlfall, 2009; Verschaffel et al., 2009). Therefore, the focus instead should be placed on describing the *in situ* nature of one's ability to retrieve and develop different solutions for a given problem while considering a problem's characteristics and other relevant features (Baroody, 2003; Threlfall, 2002, 2009; Verschaffel et al., 2009). In the present dissertation, this type of well-connected knowledge of numerical characteristics and relations between numbers, which underlies adaptivity with arithmetic, is referred to as *adaptive number knowledge*. Adaptive number knowledge and procedural flexibility are both necessary components of adaptivity with arithmetic (McMullen et al., 2017; Schneider, Rittle-Johnson & Star, 2011; Verschaffel et al., 2009). In other words, having a wide repertoire of different strategies (procedural flexibility), is needed in order to be optimal and efficient in arithmetic problem solving, but these strategies will only be actively used if there is also a strong mental representation of numerical characteristics and relations (adaptive number knowledge) which makes them available.

To develop this adaptivity with arithmetic problem solving, it is important to provide students with opportunities to practise solving open-ended mathematical problems in novel and unknown arithmetical contexts. However, pedagogical recommendations and models for enhancing flexibility and adaptivity are rare and mostly limited to teaching the use of a few strategies (Heinze et al., 2009; Verschaffel et al., 2009). This approach is insufficient, as the definition of flexibility and adaptivity with arithmetic problem solving is open-ended. As a result, 'efficient strategies' cannot be directly instructed, but instead need to be constructed and discovered individually (Threlfall, 2002, 2009) by practising with numbers and reflecting on what it means to have multiple solutions to the same problem (Baroody, 2003; Star & Rittle-Johnson, 2008; Verschaffel et al., 2009). It is challenging to provide practical guidelines on how to achieve this type of mathematical thinking and problem solving in the larger context of a classroom where little time for individualised instruction is available, and the practice often supports the development of beliefs and misconceptions regarding the exact nature of mathematics with one algorithm to find 'the right solution' to a problem (Pepin & Rösken-Winter, 2015).

As a general guideline, it is suggested that students should be provided with opportunities that trigger the exploration of underlying relations between numbers and operations, as practising with different number-operation combinations can help develop stronger networks of numerical relations and enable students to notice and use these relations during arithmetic problem solving (Baroody, 2003; Threlfall, 2002, 2009; Verschaffel et al., 2009). Solving arithmetic problems in multiple ways and practising with various combinations of numbers and operations may help strengthen the mental representation of the system of natural numbers (Blöte, Klein & Beishuizen, 2000; Rittle-Johnson & Star, 2009; Star & Seifert, 2006). In turn, a well-connected mental representation of the natural number system is expected to make different numerical characteristics and relations more readily available, enable the recognition and use of key numbers in arithmetic problem solving and strengthen the tendency to notice and make use of different numerical relations during arithmetic problem solving (Sowder, 1992). A strong mental representation of the system of natural numbers enables, for example, recognising numbers with many factors or multiples, recognising numbers in equations that are close to other numbers that are easy to work with, or understanding the base-10 structure and basic arithmetic principles like associativity or commutativity (Berch, 2005; McMullen et al., 2017).

Previous attempts that aimed to enhance flexibility and adaptivity with arithmetic problem solving by addressing the underlying understanding of numerical relations used external representations of the number system to promote students' discovery of their own strategies in mental arithmetic (Heirdsfield & Cooper, 2004; Heirdsfield, 2011), or prompted reflection on using alternative solutions for the same problem in solving equations (Blöte et al., 2000; Rittle-Johnson & Star, 2009; Star & Seifert, 2006). However, these attempts were not related directly to developing adaptive number knowledge. The main challenge when aiming to develop adaptive number knowledge is that the training should be open-ended and consider individual differences in perceiving different methods' efficiency to solve arithmetic problems (Heinze et al., 2009; Threlfall, 2009; Verschaffel et al., 2009). At the same time, the training also should be structured in a way that promotes reflection on different solutions, as well as a tendency to look for alternative solutions when solving arithmetic problems (Rittle-Johnson & Star, 2007, 2009; Star & Rittle-Johnson, 2008). In the classroom context, opportunities for this type of practice are limited, as it necessitates individualised instruction methods and teachers' scaffolding. The affordances of game-based learning can help to overcome this limitation by providing students with more opportunities to train extensively with various number-operation combinations, and the open nature of tasks can naturally support differentiation that accounts for individual differences (Lehtinen et al., 2017).

It is expected, that as a result of extensive practice with various number-operation combinations using the NNG, primary school students will develop their mental representations of the system of natural numbers. This in turn will enable them to notice and use more of these numerical relations during arithmetic problem solving. In the long run, this type of practice is expected to help students retrieve efficient solutions because they have a well-connected mental representation of various numerical relations available that enables them to consider and evaluate different solution strategies based on problem characteristics or context.

1.5 Role of different NNG game features in enhancing adaptive number knowledge

To utilise game-based learning's affordances fully, any training using a game-based learning environment should aim to transform how students interact with educational content in a way that was difficult, if not impossible, to achieve by other means (Bray & Tangney, 2017; Young, 2017). The NNG is a game-based learning environment that aims to improve primary school students' adaptive number knowledge by providing opportunities to engage in strategic work with various combinations of numbers and operations. The NNG uses the 100 square as the external representation of the natural number system (1-100), and the game narrative uses the metaphor of navigation with a ship on different landscapes of land and sea (see Figure 1), in which a map represents a basic game unit. The players' goal is to collect different raw materials to build a settlement. Navigation is done by strategically selecting and executing different calculations to move the ship from one number to the next. Challenge is provided by the following main features of the game: changing landscapes (islands), harbour and target locations, different game modes (moves or energy) and each map's difficulty level.

As the mathematical content is integrated into the game's core mechanics (Habgood, 2007; Salen & Zimmerman, 2003), game features serve to trigger mathematically relevant actions. To make progress with the game, students need to think about various alternative solutions for the same problem and reflect on their solutions' efficacy, considering different criteria. They are encouraged to use a wide variety of different number-operation combinations in their calculations, as well as look for different numerical relations and key numbers (i.e., numbers with several factors) within the natural number system. An important aspect of the NNG design is that it makes reflection about alternative arithmetic solution methods a means of making progress within the game. Using the screenshot of an NNG map in Figure 1 as an example, the relationship between the basic NNG features and mathematical learning content is discussed in the following section.

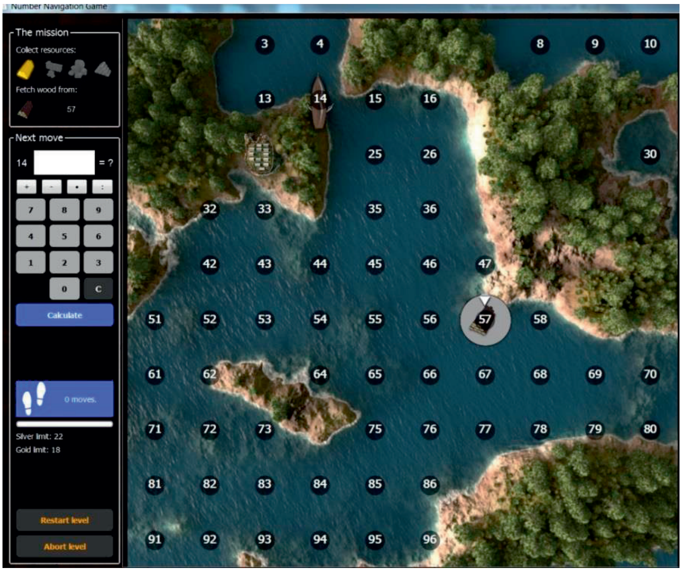


Figure 1. Screenshot of a NNG map.

Use of the 100 square. Using external representations of natural numbers can support the development of number sense and numerical understanding of children (Beishuizen, 1993; Klein, Beishuizen, & Treffers, 1998; Laski & Siegler, 2014; Siegler & Ramani, 2009; Wilson et al., 2006). While most studies in this domain explore the use of number line, the affordances of only linear representation might be limited when the aim is to work with numerical relations. For the NNG the 10 by 10 square was selected as the representation of the natural number system because it allows for both horizontal (number line) and vertical (base-10 system) representation and movement within those representations. Additionally, combining these two into diagonal movements can support decomposing and recombining numbers in various different ways. The square used in the NNG version tested in the present dissertation consisted of numbers between 1-100 as this seems most suitable for the target population: primary school students with a basic understanding of the base-10 system and skills to perform multi-digit mental arithmetic. However, the grid structure could in principle also accommodate other base structures (e.g., base-12) or larger numbers (1-10.000) for older or more skilled players.

Landscapes and harbour-target location. Each map's layout is different, including islands' sizes and shapes. Numbers that islands cover become unavailable to players, so each map provides a different set of numbers for players to work with, increasing variability in using different combinations of numbers and operations. In Figure 1, the starting number (harbour) is No. 14, and the goal number (target) is No. 57. Depending on the island layout within the map, where harbours and targets are placed becomes an important design decision, as players will use a combination of different numbers and operations between these two values.

Different game modes. Two main game modes are provided that aim to add more variability in players' use of numbers and operations, as they require adopting different 'optimal strategies' to make progress. If the game is in the 'moves mode', the player needs to reach the target using the least number of moves (operations) possible. In Figure 1, for example, **14** + 43 = **57** would be a good solution in the 'moves mode', as it only requires one move. If the game is in the 'energy mode', the player needs to reduce the magnitude of the sum of numbers used in calculations. In Figure 1, for example, **14** * 4 = 56; 56 +1 = **57** (4 + 1 = 5 energy points) could be a good option. As the example shows, the moves mode mainly aims to trigger the use of addition and subtraction with larger two-digit numbers, while the energy mode aims for more complex numerical combinations that use all four arithmetic operations. This adds further variation in students' use of different numbers and operations, prompting reflection on the efficiency of applying different strategies while considering the different game modes' rules.

Additional features. Movement in the game is always two-ways between two numbers: one indicating the harbour and one where the material is picked up before returning to the harbour. At the start this can be useful to generate awareness of the reverse nature of operations but can also facilitate players to look for alternative trajectories between two numbers. In order to prevent that players develop a habit of mechanical repetition of the reverse operations on the way back, the game introduces pirates after the first few maps. The pirates appear on specific numbers that were used in calculations on the way to the target material making it impossible to take the same route on the way back. This feature makes players redirect their trajectories thus promoting further variation in the use of numbers and operations.

In developing students' flexibility and adaptivity with arithmetic problem solving, it is challenging that these skills cannot be instructed or practised directly (Threlfall, 2009; Verschaffel et al., 2009). Research shows that even if students have a large repertoire of solution strategies, they often fail to notice instances when switching between methods would be useful, and they fail to apply these in practice (Blöte et al., 2000). The design of the NNG training aims to guide students into considering alternative solutions to the same problem and reflect on what an efficient problem-solving method would look like, given the characteristics of a context (i.e., game mode, map layout, etc.). From a mathematical perspective, an important aspect of the overall NNG design is that making progress in the game and executing equations require mentally executing, reflecting on, comparing and rejecting several other mental calculations. While the NNG overall is a flexible and open-ended environment, specific game rules and goals provide a meaningful context in which students have to look for and make use of different solutions during gameplay.

Overall, the tested game versions had 64 maps in which players needed to collect materials and return them to the harbour four times within each map. This, in combination with the aforementioned game features, offered ample opportunities for students to practise arithmetic problem solving in a flexible and adaptive manner. Such practising is expected to trigger recognition and use of numerical characteristics and relations, as well as enable efficient problem-solving strategies in mental arithmetic within and outside of the game context.

2. AIMS

Practical guidelines on training flexibility and adaptivity with arithmetic problem solving are rare, mostly due to the subjective and context-dependent nature of the complex mathematical skills and knowledge required (Threlfall, 2009; Verschaffel et el., 2009). Accordingly, any training in this domain should be open-ended and flexible enough to account for this intra- and inter-individual variability, but also be applicable to classroom practice. Game-based learning provides affordances that could be used to create ample opportunities for training with numbers and operations in an open-ended and flexible environment that can be applicable in classroom practice. However, using this type of media also elicits some inherent questions regarding design decisions, measurement, the integration of game features and learning content, and training applicability. Therefore, the present dissertation aimed to:

1) Operationalise adaptive number knowledge within a game-based learning environment and design a game for primary school students targeted primarily at developing adaptive number knowledge.
2) Operationalise adaptive number knowledge outside the game-based learning environment and show effects outside the game's context.
3) Examine how different features of the game can contribute to the development of students' adaptive number knowledge and related arithmetical skills and knowledge.
4) Provide empirical evidence for the effects of the NNG training in developing primary school students' adaptive number knowledge and related mathematical skills and knowledge, first in controlled settings and later when included in classroom practice.

To accomplish these aims the present work comprises five sub-studies that represent a continuous circle of testing and development. Figure 2 is based on the guidelines of Vanden Abeele and colleagues (2012) and provides an overview of the NNG iterative design process in light of the five publications. The main aim throughout these studies always was to base design decisions on both available empirical evidence and theoretical background and to find methods within the NNG context that support students' mathematically relevant engagement with the game in the most organic way.

Study I aimed to explore an early NNG prototype from a usability perspective focusing on the relationship of game features and the mathematical learning content. Study II comprised a series of pilot studies that aimed to develop and validate measures of adaptive number knowledge outside the game context. Study III was the first small-scale pilot study that used a second NNG prototype that aimed to focus on the relationship between students' game performance and the development of their mathematical learning outcomes. Findings from Studies I, II and III were used to develop the NNG and in-game and paper-and-pencil measurements of adaptive number knowledge. Study IV is placed at the centre of Figure 2 because it describes in detail the different stages of game development, both from practical and theoretical perspectives. Finally, Study V was a large-scale, randomised, controlled study using a final NNG prototype on three grade levels, aiming to test the effects of the NNG in developing primary school students' adaptive number knowledge and related mathematical skills and knowledge such as arithmetic fluency and pre-algebra knowledge. Additionally, Study V aimed to explore the relation between game performance and these math-learning outcomes when the NNG is used in a naturalistic classroom setting.

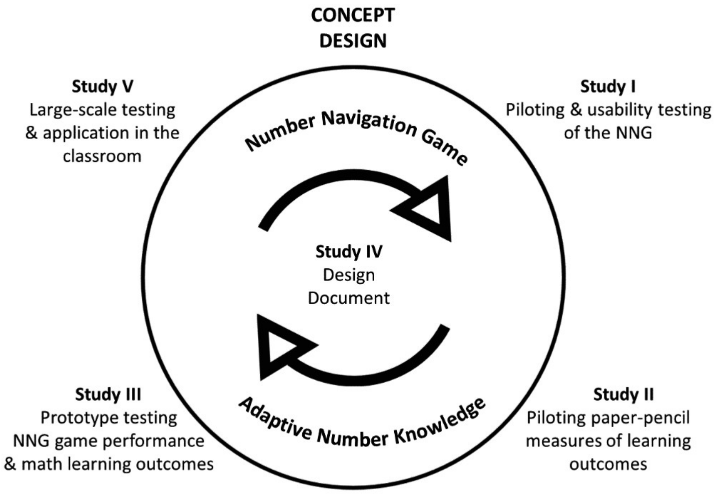


Figure 2. Iterative development and design of the NNG. Adapted from the P-III framework (Vanden Abeele et al., 2012, *p*. 83).

3. METHODS

The present dissertation consists of five studies related to the process of testing and developing the NNG. More qualitative and explorative methods were used at the beginning of the work for developmental purposes, then more quantitative methods by the end of the work for testing purposes. Study I used multiple sources of qualitative data collection (video-recording, screen capture and interviews) to test a basic NNG version and different intended game features. This was followed by a set of pilot studies in which paper-and-pencil measures of adaptive number knowledge were refined and tested across different samples and age groups (Study II), with the first NNG prototype's effects tested using a quasi-experimental design (Study III). Study IV provides an overview of the NNG development process and includes a detailed practical description of the final NNG prototype tested in Study V. Study V used a large-scale, randomised, experimental design and tested the final NNG prototype's effects in developing primary school students' adaptive number knowledge and related mathematical skills through a 10-week-long intervention in a naturalistic classroom setting. The following section provides an overview of the different methods, samples and procedures used in these studies. Table 1 presents a complete overview of methods used.

3.1 Research context

All five studies were conducted at the Centre for Research on Learning and Instruction within the Department of Teacher Education at the University of Turku in Finland. The studies were conducted as part of a larger project related to the Number Navigation Game that included game development, motivational aspects and mathematical learning outcomes. The present dissertation focused mainly on mathematical learning outcomes related to NNG. Motivational aspects have been presented in the dissertation of Rodríguez-Aflecht (2018). Studies within the present dissertation were conducted as part of the Academy of Finland CUMA project (PI: Professor Erno Lehtinen) and as part of the Academy of Finland's DECIN project (PI: Associate Professor Minna Hannula-Sormunen). In addition, the present doctoral work was funded (January 2014 to November 2017) through the Doctoral Programme on Learning, Teaching and Learning Environments Research (OPPI) doctoral programme in the University of Turku Graduate School.

3.2 Ethical considerations

Participation was voluntary in all studies, with participants and their parents' consent obtained from all participants whose data were used in the present set of studies. The University of Turku's ethical guidelines were followed in all studies. In Study I, due to the video data's sensitive nature, special care was taken to protect subjects' anonymity. In Studies II and III, parents' and students' consent were obtained before starting the sub-studies. In sub-study IIb, the university students were asked to fill out a consent form. In Study V, special attention was given to informing all parties (students, parents, school authorities and teachers) on the study's details. The teachers distributed consent forms to parents via their children, who returned them. If students or parents did not agree to participate, students were free to do some other activity during testing or participate in the study's activities as part of their regular schoolwork. All the data from students who did not provide parental consent were destroyed before any analyses were conducted. Gameplay and measurement data, and any information revealing the students' identities, were never stored in the same place. Data analyses were conducted by different people than data coding, and only after the data were anonymised. For ethical reasons, after the post-test, the control group in Study V also received the NNG and used the game as part of its math classes for six weeks so that these students were not denied any potentially beneficial effects from playing the NNG.

3.3 Participants

All participants came from the southwest region of Finland and were attending primary school or higher education in the Finnish language. A detailed description of participants can be found in Table 1. Study I was a case study using pairs of players (9 and 11 years old) and one individual player (11 years old) to test the NNG. Study II comprised three sub-studies and used different participants. Study IIa used students from third to fifth grades, Study IIb used university students, and Study IIc used sixth-grade students from one classroom. Gender distribution was equal in the samples of primary school students, but females were overrepresented in the sample of university students. Study III used the same sample as Study IIc. Study V used a sample of 1,168 primary school students from fourth to sixth grades, with an equal gender distribution.

3.4 Instruments

For better reliability and validity, it is suggested that learning outcomes related to game-based training be measured using a combination of different methods and measurement types, such as game-based and paper-and-pencil measures, near- and far-transfer tasks, and standardised tests (All et al., 2014, 2016; Cheung & Slavin, 2013; Slavin & Lake, 2008). In the present dissertation, the main learning outcome, adaptive number knowledge, was measured with the Arithmetic Production Task, developed for these studies' purposes. As it was expected that students' adaptive number knowledge would be related to their general arithmetic fluency and pre-algebra knowledge (McMullen et al., 2017), these additional learning outcomes were measured with a combination of standardised and self-constructed tests. Finally, in addition to the paper-and-pencil measures, game-log data were used to extract measures for players' game performance.

3.4.1 Adaptive number knowledge

Different versions of the Arithmetic Production Task were used to measure participants' adaptive number knowledge. The task was developed continuously and slightly modified throughout Studies II, III and V to achieve better reliability and validity. However, the main task in each version was the same: to calculate a given target number by combining a set of four to five given numbers and four arithmetic operations in as many ways as possible within a limited amount of time.

The task's objective was to provide an estimate of participants' adaptive number knowledge by measuring their ability to notice and use different numerical relations in their solutions. The method of scoring solutions in answers was developed and validated mainly in Study II. Participants' solutions were scored on two criteria: 1) total number of correct solutions that matched the instructions, and 2) total number of multi-operational solutions (that use multiple steps and different arithmetic operations in the equations). These criteria were considered to be the quantitative and qualitative aspects of adaptive number knowledge and were treated as separate variables in all studies that used the Arithmetic Production Task.

In the small-scale pilot studies (Studies II and III), participants' answers were coded directly on these two criteria. For reliability reasons, in the large-scale intervention (Study V), participants' answers were transcribed in spreadsheets, and Excel macros were written to extract participants' scores on each item based on these two aspects.

The number of items and the numbers used in each item, as well as the time given to solve each item, were modified slightly throughout the studies to achieve higher reliability. Usually, four to five items were used, except for the Study V post-test, which used eight items to achieve better reliability. For Study V, during both the pre- and post-tests, a practice item was used with the same timing, but easy combinations and the opportunity to ask clarification questions after completing it. The time allowed to solve an item was between 60 and 90 seconds. Appendix 2 shows an example item from the Arithmetic Production Task. Appendix 3 provides an overview of the items used in Studies II, III and V, as well as the items' timing and the Arithmetic Production Task's reliability (correct solutions and multi-operational solutions) in these studies.

3.4.2 Related mathematical measures

Arithmetic fluency. As basic arithmetic fluency is foundational to developing adaptive and flexible arithmetic problem-solving skills. The Woodcock-Johnson Math Fluency sub-test was used to measure students' basic math fluency in Study III (IIc) and Study V (McGrew, LaForte & Schrank, 2014). The Woodcock-Johnson IV Test of Achievement is a widely used, standardised and reliable measure with high-to-excellent reliability scores reported (Villarreal, 2015). The test comprises two pages with a total of 160 items in which students complete as many arithmetic problems (simple addition, subtraction and multiplication) as they can in three minutes, with increasing difficulty as the test progresses. Final scores comprised the total number of correct answers (each correct answer was worth one point).

Pre-algebra knowledge. In Study V, a measure of pre-algebra knowledge was used as a far-transfer measure of the type of relational thinking and complex, multi-step mental work with arithmetic relations that can be foundational to algebraic thinking. The task comprised short-answer and multiple-choice questions on solving equations, such as 12 + ___ = 11 + 15, with six multiple-choice questions in the pre-test and six more-difficult multiple-choice questions and six fill-in questions in the post-test. The list of items for the two time points is provided in Appendix 4. On both time points, students had eight minutes to answer all questions. Each correct item was worth one point, so the maximum pre-test score was six, while the maximum post-test score was twelve. Cronbach’s alpha reliability values for the pre-algebra knowledge test were .73 for the pre-test and .88 for the post-test.

Knowledge of arithmetic concepts. A multiple-choice test of students' conceptual knowledge of arithmetic operations and properties was administered (Study II). The test contained eight items measuring students' knowledge of order of operations, commutativity, associativity and conceptual understanding of multiplication. Cronbach's alpha was .70.

3.4.3 Game performance

Like adaptive number knowledge, the measuring method for participants’ game performance was continuously refined and developed throughout the studies. The NNG version in Study I did not include a game log data-save option yet, but the transcripts of players’ equations served as guidance in developing one. The NNG version in Study II had the first version of the game log save feature, which could produce time-stamped event logs of each equation that players made, providing information regarding the map played and signalling when material was picked up. A map was considered the basic game unit both in Studies III and V. There were 64 maps in total and players needed to pick up and return four raw materials to complete a map.

Game log data in Study V was already able to provide rich and reliable information on players’ progress in the game. As the learning content seemed to be supported by the game mechanics, it was hypothesised that the number of maps completed can be a good indicator of game performance, as players need to gather four raw materials to complete a map, so they needed to be engaged with the game’s mathematically relevant aspects during their time spent on task. Thus, the total number of maps completed was considered to be a measure of game performance in Studies III and V.

Students ran the NNG in Studies III and V on pen drives and the log data was also saved and stored on these pen drives. The data were saved in .txt format and transported to Excel, where macros were developed to summarise players’ information and the summarised information was then used for regular statistical analyses together with data from the paper-and-pencil measures of mathematical learning outcomes.

Table 1. *Overview of the present dissertation's methods*

Study	Aims	Participants	Instruments	Procedures	Analyses
Study I	Explore the relationship between game features and in-game markers of adaptive number knowledge	N = 3 Session A: 11-year-old Session B: 9- and 11-year-olds (collaborative play) all male	▪ Video-recorded observation ▪ Screen capture ▪ Open-ended semi-structured interviews	Two play sessions (1.5 hours long each) with NNG prototype and paper mock-ups of different game features	▪ Video-data analyses to identify examples of adaptivity with arithmetic ▪ Frequency analyses based on transcripts of players' equations to identify patterns ▪ Qualitative analyses of interviews
Study II	Develop measures of adaptive number knowledge and explore individual differences in adaptive number knowledge (Sub-studies IIa, IIb and IIc)	IIa) N = 55, third-through fifth-graders, 25 females IIb) N = 55, university students, 44 females IIc) N = 22, sixth-graders, 11 females	▪ Arithmetic Production Task (IIa, IIb and IIc) ▪ W-J math fluency task (IIc) ▪ Knowledge of arithmetic concepts test (IIc) ▪ Math grades (IIa and IIc)	Sub-studies IIa and IIb: Participants completed paper-and-pencil tasks once in the classroom or during a university lecture IIc: pre-test data from Study III	▪ Descriptive statistics ▪ K-Means cluster analyses to show individual differences ▪ ANOVAs to test groups differences ▪ IIc: correlations and linear regression to explore the relation between mathematical skills
Study III	Explore the effects of the NNG and the relationship between game performance and adaptive number knowledge	N = 22 (same participants as Study IIc)	▪ Arithmetic Production Task ▪ W-J math-fluency task ▪ Game performance (total maps completed)	Pre-test/post-test quasi-experimental design (no control group) Seven weeks of NNG play in pairs; 45 minutes per session, once a week	▪ Descriptive statistics ▪ Paired samples t-test to test pre- and post-test differences ▪ Correlations and stepwise linear regression to explore the relation between game performance

Study	Aims	Participants	Instruments	Procedures	Analyses
					and mathematical learning outcomes
Study IV	Provide theoretical and practical background on the NNG game design	Not available	▪ Design documents ▪ Play and bug testing ▪ Video-recorder observations and field notes from Studies I and III	Not available	Description of the NNG design process, including the following aspects: theoretical background, use of 100 square, game mechanics, feedback, game structure, scoring modes, customisation, log data, future plans, examples of game development
Study V	Explore the effects of the NNG on primary school students' adaptive number knowledge and related arithmetic skills and knowledge Explore the relationship between game performance and mathematical learning outcomes	N = 1,168 Grade 4 (n = 135) Grade 5 (n = 606) Grade 6 (n = 427) M_{age} = 11.41 (SD = .77), 546 females	▪ Arithmetic Production Task ▪ W-J math-fluency task ▪ Pre-algebra task ▪ Game performance (total maps completed)	Pre-test/post-test experimental design Classrooms randomised to experimental (n = 642) and control (n = 526) groups. Experimental group: 10 weeks of NNG play during math classes (30-45 minutes per session, four hours average time on task) Control group: regular math teaching	▪ Descriptive statistics ▪ ICC analyses ▪ Likelihood ratio test to compare model fits ▪ Linear mixed model analyses to test group differences ▪ Hierarchical linear regression to explore the relationship between game performance and mathematical learning outcomes

4. OVERVIEW OF STUDIES

4.1 Study I

Brezovszky, B., Lehtinen, E., McMullen, J., Rodriguez, G. & Veermans, K. (2013). Training flexible and adaptive arithmetic problem-solving skills through exploration with numbers: The development of Number Navigation game. In *7th European Conference on Game-Based Learning, ECGBL 2013* (Vol. 2, pp. 626–634).

The aim of Study I was to explore the first working prototype game version of the NNG from a user-experience perspective and conceptualise adaptive number knowledge within the game's context. More precisely, the study aimed to capture the type of game action that can be associated with students' active engagement in working with different number combinations and numerical relations, which was expected to be foundational in the development of their adaptive number knowledge. Study I focused on the following questions: 1) what kind of game patterns can be noticed during gameplay, and 2) what is the relationship between different game features and mathematical problem-solving patterns?

A trial game version was tested in two sessions, with one student playing individually (age 11) and two students (ages 9 and 11) playing the game collaboratively. Data were collected using video-recorded observations, screen capture of gameplay and open-ended interview questions regarding the game experience. The prototype game version consisted of three maps and had no built-in rule or reward system. Instead, the play sessions served as a mock-up test for three different scoring modes and other game features. The transcript of players' equations, synchronised with their video-recorded observations during gameplay, comprised the principal source of data for analyses and was extended with post-game interviews. Both play sessions lasted around 1.5 hours each, including instructions and interview questions.

Results showed that this prototype NNG version already was able to trigger players' active engagement with various different number combinations and numerical relations as players executed roughly 200 mental calculations, from which around 35% were unique (in which calculations were not repeated). Data showed examples of players' own discoveries related to numerical relations within the game's context, such as recognising the usefulness of creating equations using the numbers 9 and 11 to move diagonally in the 100 square, substituting addition with multiplication to arrive at faster solutions and recognising the reverse nature of operations. The use of multiplication and division was less frequent in players' equations, but introducing certain game rules (i.e., minimum energy scoring mode) successfully could trigger the use of a larger variety of numbers and operations. The layout of a map (i.e., placements of starting and finishing numbers combined with the placement of islands) proved to be important in defining the amount of reflection and planning during gameplay and the amount of variation in players' use of numbers. Data showed many instances of complex planning using multiple steps in equations. During pair play, players used discussions and tended to evaluate various alternative solutions' efficiency out loud before making moves.

These results suggested that this basic game version already was able to promote the desired mathematically relevant game action in the NNG. Additionally, results showed that taking extra care in planning the layout of each map and using different scoring modes in the game were essential to triggering players' exploration and reflection within the game. The results of this exploratory pilot study were used to develop a more complex prototype game version that was tested in Study III.

4.2 Study II

McMullen, J., Brezovszky, B., Rodríguez-Aflecht, G., Pongsakdi, N., Hannula-Sormunen, M. M. & Lehtinen, E. (2016). Adaptive number knowledge: Exploring the foundations of adaptivity with whole-number arithmetic. *Learning and Individual Differences*, 47, 172–181. https://doi.org/10.1016/j.lindif.2016.02.007

Two common issues in game-based learning are related to: 1) matching the learning content within the game context (Devlin, 2011; Young et al., 2012), and 2) finding relevant measures for learning outcomes (All et al., 2016). It is often the case that complex learning outcomes are measured with standardised test packages or regular school tasks that are very far from the game's intended learning objectives. Therefore, while Study I aimed to focus on the first issue, Study II aimed to conceptualise how adaptive number knowledge can be measured in a paper-and-pencil format that can be used flexibly with different age groups and applied on a large scale. Study II focused on the following questions: What type of individual differences can be observed in primary school students and adults' adaptive number knowledge, and how is adaptive number knowledge related to other mathematical skills and knowledge?

Thus, three sub-studies were conducted that examined elementary school students and adults' responses on slightly different versions of the Arithmetic Production Task (measuring adaptive number knowledge). The task's objective was to calculate a target number by combining four or five given numbers and the four arithmetic operations in as many ways as possible in a limited amount of time. Participants in the three sub-studies were third-, fourth- and fifth-graders (Study IIa; n = 55), university students (Study IIb; n= 55) and sixth-graders (Study IIc; n= 22 – the same participants used in Study III). Participants' responses were coded based on the number of their correct and multi-operational solutions, and these responses were explored across the dense (items with many numerical relations) and sparse (items with fewer and more complex relations) items on the Arithmetic Production Task.

Results from both Study IIa (elementary school students) and IIb (university students) showed the following four clusters based on students' correct and multi-operational responses on the dense and sparse items: 1) the low group provided just a few correct and a few multi-operational solutions; 2) the simple group provided quite many correct solutions, but few multi-operational solutions; 3) the complex group had a more equal proportion of correct and multi-operational solutions, especially on the sparse items; and 4) the high group had both many correct and many multi-operational solutions. The same clusters could be distinguished in the sample of third- to fifth-graders and among university students. In Study IIc (sixth-graders), adaptive number knowledge was found to be related to arithmetic fluency and conceptual knowledge in arithmetic.

Together, this set of studies represented the first attempt to develop an instrument for measuring adaptive number knowledge that does not rely on the strategy-choice model and which can be applied on a large scale. The individual differences found in participants' responses across the different age groups and samples suggest that the Arithmetic Production Task was able to capture how participants recognise opportunities to use numerical features in their arithmetic problem solving, which can be used as an indicator of their adaptive number knowledge. Based on these results, the Arithmetic Production Task was refined and developed further, and administered on a larger scale in Study V as a main outcome measure of the NNG training.

4.3 Study III

Brezovszky, B., Rodríguez-Aflecht, G., McMullen, J., Veermans, K., Pongsakdi, N., Hannula-Sormunen, M. M. & Lehtinen, E. (2015). Developing adaptive number knowledge with the number navigation game-based learning environment. In J. Torbeyns, E. Lehtinen & J. Elen (Eds.), *Describing and Studying Domain-Specific Serious Games* (pp. 155–170). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20276-1_10

Study III represented the first attempt to test the effects of the NNG on the development of primary school students' adaptive number knowledge after the initial stage of piloting and development (Studies I and II). The NNG version used in this study included 64 maps, the two basic scoring modes integrated into the game mechanics, and logged players' game data. Study III focused on the following question: How is game performance in NNG related to the development of primary school students' adaptive number knowledge?

Participants were sixth-grade students broken down into 11 pairs (11 females; age range: 11-13) from one classroom who played the NNG for seven weeks. One play session lasted an entire class period once a week. Each play session took place during students' regular math classes. Students' adaptive number knowledge was measured during pre- and post-tests, and students' solutions were scored based on the two aspects described in Study II (correct and multi-operational solutions). Math fluency was used as a control measure. Game performance was measured as the total number of maps completed during the seven weeks training with the NNG.

Results showed an improvement in participants' performance in their correct solutions (d = 1.50), multi-operational solutions (d = .55) and math fluency (d = .46). Game performance was found to be a predictor of students' post-test performance on the multi-operational aspect of adaptive number knowledge. This relation was not found for the number of correct solutions or for math fluency. The unique relationship found between students' game performance and their development in noticing complex numerical relations in their arithmetic problem solving suggests that training with the NNG promoted aspects of the type of mathematical thinking that is necessary in the development of adaptive number knowledge.

Players' log data from Study III, as well as data from field observations and video-recorded observations of gameplay that were collected, but not reported, in Study III, were used to finalise the challenge-reward system and other specific changes in the NNG. This process of game development and a comprehensive overview of the theoretical foundations and practical design decisions of the NNG are described in Study IV. Piloting from Studies I-III and game development described in Study IV were used to create a final prototype of the NNG, which was tested on a larger scale in Study V.

4.4 Study IV

Lehtinen, E., Brezovszky, B., Rodríguez-Aflecht, G., Lehtinen, H., Hannula-Sormunen, M. M., McMullen, J., Pongsakdi, N., Veermans, K. & Jaakkola, T. (2015). Number Navigation Game (NNG):

Design Principles and Game Description. In J. Torbeyns, E. Lehtinen & J. Elen (Eds.), *Describing and Studying Domain-Specific Serious Games* (pp. 45–61). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20276-1_4

Study IV describes the theoretical background and specific design decisions implemented in the NNG, which aims to trigger students' active engagement in working with different combinations of numbers and operations. This practice was intended to develop students' recognition and use of numerical characteristics and relations that, in turn, were expected to lead to increased adaptivity with arithmetic problem solving.

Study IV presents in detail the overview of the following aspects of the NNG: NNG structure (64 maps and four challenge levels), players' progress, practical examples of the two main scoring modes, game story (building settlements and gathering raw materials), reward system, customisation options and log data. Specific examples of these design decisions described in Study IV are: 1) using the idea of navigation to translate the type of relational thinking needed for the development of adaptive number knowledge; 2) using maps as the main unit of progress in the game; 3) using the 100 square as an external representation of the natural number system, which supports navigation in different directions (horizontal, vertical and diagonal) when performing mental calculations; 4) open-ended game design with no single correct solution to support exploration and reflection; 5) progressive level design to create a game that can benefit players from different age groups and skill levels; 6) adding game rules and challenges to trigger more variability in players' use of number-operation combinations (i.e., use of islands to make certain numbers unavailable, minimum moves and minimum energy scoring mode, pirates redirecting players' routes, etc.); 7) an encouraging reward system that supports reflection (i.e., players need to reach certain benchmarks for progress, but they are not punished for insufficient performance, as maps are always replayable). The Appendix from Study IV provides a hands-on example of the different issues that came up during Studies I and III and the solutions that were implemented in the final prototype used in Study V.

Overall, Study IV provides a detailed overview of game design decisions in the NNG that are based on theory and the results from previous pilot studies using the NNG (Studies I and III). Additionally, Study IV provided a practical description of the structure and development of the NNG. Due to strict publishing standards and space limitations in peer-reviewed journals, longer, more practical descriptions of educational game design processes and decisions rarely are published. Study IV is intended to be used as a guide and point of reference for anyone planning to use NNG in research or practice.

4.5 Study V

Brezovszky, B., McMullen, J., Veermans, K, Hannula-Sormunen, M., Rodríguez-Aflecht, G, Pongsakdi, N., Laakkonen, E. & Lehtinen, E. (2019). Effects of a Mathematics Game-Based Learning Environment on Primary School Students' Adaptive Number Knowledge. *Computers & Education*.

After the initial piloting phase, Study V aimed to test the NNG's effectiveness on a large scale in the naturalistic classroom setting. As no previous training or measure of adaptive number knowledge existed that could be applied without additional support and on a large scale, it was an important objective in Study V to test the effects of the NNG in a context that is very close to regular classroom teaching. Study II showed that adaptive number knowledge was related to primary school students' arithmetic fluency and pre-algebra knowledge (in grade six), and Study III showed a relationship between sixth graders' NNG game performance and their achievement on the multi-operational aspect of the Arithmetic Production Task. Thus, Study V aimed to explore these findings on a large scale and across three different grade levels. Study V focused on the following question: how does training with the NNG affect the development of primary school students' adaptive number knowledge, arithmetic fluency and pre-algebra knowledge in different grade levels; and does students' performance on the NNG affect the development of mathematical learning outcomes?

Study V used a pre-test/post-test randomised control design in which classrooms were randomly assigned either experimental or control conditions. The experimental group played the NNG for 10 weeks during math class, two to three times a week, with an average of four hours of time spent on task altogether. Teachers were provided with detailed guidelines and support regarding the game and the intervention and in order to maintain ecological validity, they were free to design the specifics of the play sessions. The control group participated in regular math classes during these 10 weeks. Measures of adaptive number knowledge, math fluency and pre-algebra knowledge were administered during the pre- and post-tests. Game performance was measured as the total number of maps completed.

Results showed significant overall improvement in the experimental group compared with the control group on all measured mathematical learning outcomes. More practice with the game was associated with better mathematical learning outcomes, and game performance was related to the experimental groups' post-test scores, even after controlling for pre-test scores and grade. Grade-level analyses revealed differences between the mathematical learning outcomes in different grades. In grade four, results showed an improvement on the number of correct solutions of adaptive number knowledge task and on math fluency; in grade five there was an improvement in both the number of correct and number of multi-operational solution; and in grade six there was an improvement in students' pre-algebra knowledge. These results suggest that the NNG game design was flexible enough to support the development of different types of mathematical skills and knowledge across the three grades.

Based on the results of Study V it seems that the core game design and added features can support players' extended practice and reflection with different combinations of numbers and operations in an organic way and possible learning outcomes can be noticed even in a far-transfer task like pre-algebra knowledge. Additionally, these results provide evidence regarding the ecological validity of the NNG and the Arithmetic Production Task, both of which were used successfully in the context of small-scale, controlled pilot studies (Studies II and III) and as part of everyday classroom practice on a large scale (Study V).

5. MAIN FINDINGS AND DISCUSSION

The main aim of this dissertation was to develop a novel, game-based method for training primary school students' adaptivity with arithmetic problem solving, which is difficult to achieve in the context of regular mathematics instruction (Hickendorff, 2018). For this reason, the NNG game-based learning environment was developed around a game logic that integrates gameplay and learning content. The aim was to provide students with a large amount of arithmetic practice while solving open-ended problems that also trigger reflection on (the efficiency of) their solutions. The development and testing of the game also required deepening the understanding of the nature of knowledge underlying flexible and adaptive arithmetic skills and the development of novel methods for measuring adaptive number knowledge outside the game context. Overall, results show that the game-based training provided a novel type of practice that strengthens the underlying numerical connections and number relations, with an emphasis on different outcomes in different grade levels. The NNG can be used successfully to enhance relatively complex mathematical learning outcomes, such as adaptive number knowledge, but also other mathematical skills and knowledge that are part of traditional classroom practice (i.e., basic arithmetic fluency and pre-algebra knowledge).

Results from Study I showed that the core game mechanics and basic representation of the number system used in the NNG were efficient in promoting mathematical thinking and problem solving. Even with the basic structure of the game used in Study I, students performed many arithmetic calculations. This basic game structure prompted students to look for alternative solutions to problems and triggered reflection on solutions. In the context of regular mathematics instruction, it is difficult to provide students with many opportunities with this type of reflective strategic work of solving arithmetic problems in multiple ways. Thus, strengthening students' tendencies to look for patterns and alternative solutions is not only important in the development of their adaptive number knowledge, but also may support breaking students' frequent belief regarding the nature of classroom mathematics that tend to seek one correct answer to a problem (Pepin & Rösken-Winter, 2015). Results indicated that mathematical content was an integral part of the gameplay as players progressed in the game by interacting with mathematically relevant problems (Devlin, 2011; Habgood & Ainsworth, 2011).

As one of the aims of this dissertation was to show the effects of the NNG on adaptive number knowledge, it was important to develop and validate measures of adaptive number knowledge. Previous measures of adaptivity and flexibility required either individual qualitative assessment (Blöte et al., 2000; Blöte, van der Burg & Klein, 2001; Reed, Stevenson, Broens-Paffen, Kirschner & Jolles, 2015), which affects their usability, or relied extensively on the use of specific problem-solving strategies (Torbeyns, de Smedt, Stassens, Ghesquière & Verschaffel, 2009), which affect their validity. The Arithmetic Production Task, developed and validated mainly in Study II, operationalises adaptive number knowledge by the amount and quality of students' recognition and use of numerical characteristics and relations in their arithmetic problem solving, representing the first attempt to measure adaptive number knowledge on a larger scale. Results grounded adaptive number knowledge as a necessary component of adaptivity with arithmetic, and showed that the Arithmetic Production Task could capture both quantitative and qualitative aspects of adaptive number knowledge across various samples and age groups. Study II's findings provided a framework to develop the game-based training and the measurements of expected learning outcomes further.

Study III built on the findings from Studies I and II and combined them into a first test of NNG's effects on developing sixth-grade students' adaptive number knowledge and math fluency. Results showed substantial improvement based on pre-test and post-test scores in measured mathematical learning outcomes and a relation between game performance and students' performance on the multi-operational aspect of the Arithmetic Production Task. These results suggested that in sixth grade, when procedural fluency with arithmetic is relatively well-established, gameplay with NNG was more likely to strengthen students' ability to look for and use more complex numerical relations in their arithmetic problem solving. These findings are in line with those of Study II regarding the relationship found between students' adaptive number knowledge and knowledge of arithmetic concepts in sixth grade and the unique effects from gameplay with NNG on sixth-grade students' pre-algebra knowledge that was found in Study V.

Study V concentrated on questions similar to those in Study III, using a further-developed game version (presented in Study IV), but extended both the scope (fourth to sixth grades) and the size of the study to naturalistic classroom settings using a pre-test/post-test randomised control design that allowed for more generalizable and reliable conclusions regarding NNG's effects on adaptive number knowledge and related arithmetical skills in the classroom. In line with Studies II and III, Study V showed performance improvement between pre- and post-test scores on all measured mathematical learning outcomes (i.e., adaptive number knowledge, math fluency and pre-algebra knowledge) after the 10-week training period with the NNG. Additionally, the results showed that the NNG develops different aspects of adaptive number knowledge, math fluency and pre-algebra knowledge across grade levels. These findings also provide valuable information regarding possible developmental trajectories in adaptive number knowledge, suggesting that teachers can use the NNG to support these developmental trajectories in everyday classroom practice. The relationship between game performance and adaptive number knowledge found in Study III was also found in Study V. Although this relationship in Study III was found only in the case of multi-operational aspects of adaptive number knowledge, Study V showed a relationship between game performance and all measured mathematical learning outcomes. This suggests that the sample in Study III probably was too small to be able to detect possible relations.

5.1 Theoretical and methodological implications

This dissertation's theoretical implications are related to game-based learning environments' design principles on one hand and to theoretical explanations of adaptivity with arithmetic on the other hand.

Many suggestions and guidelines have been proposed regarding the development of game-based learning environments, but different training methods require different considerations, depending on the learning objective and the technology type used. Therefore, it is difficult to provide standard guidelines. However, investing resources toward integrating educational content into game mechanics can be considered a general guiding principle in educational game design (All et al., 2016; Devlin, 2011; Habgood, 2007; Vanden Abeele et al., 2012; Young et al., 2012). How well the content is integrated into the game, which game features would trigger expected behaviour or cognitive work, and how players interact with the relevant features are foundational questions. It is not the game itself, but its design that determines the efficiency of game-based learning (All et al., 2016; Bray & Tangney, 2017). The design and development process should be iterative and multidisciplinary, and decisions should be based on theory and empirical results (Vanden Abeele et al., 2012). This dissertation's findings provide evidence of the value of this iterative design process. The five studies presented in this dissertation followed the aforementioned general guidelines. While game design decisions were only in focus in Studies I and IV, each study considered questions related to the relationship between core game features and how they promote the type of mathematical thinking and problem solving necessary to develop adaptive number knowledge. Progress in the NNG happens by interacting with the mathematical content. Players have the freedom to choose different solutions, but are also provided with structure (game modes and rules) that facilitates reflection on their solutions. Such thinking and problem solving describe the fundamental type of mathematically relevant behaviour that happens while using the NNG and that was expected to develop students' adaptive number knowledge.

As results of reviews show, interventions using drill-and-practice type of game-based learning environments in mathematics produce limited results (Byun & Joung, 2018; Ke, 2009). In the present dissertation, gameplay with the NNG was associated with higher mathematical learning gains across different studies and across different samples. Integrating learning content into game mechanics in the case of the NNG was successful in terms of transferring learning gains not only to novel tasks to measure advanced adaptive number knowledge, but also to more regular school tasks. This can be explained by the theoretical proposition that integrating the mathematical content and gameplay can lead to higher chances for players to interact with the mathematically relevant learning content which practice leads to better learning outcomes (Habgood, 2007; Habgood & Ainsworth, 2011; Ke, 2008). The concept of adaptive number knowledge was operationalized not only outside the game context (outcome measures) but also within the NNG as mathematically relevant player behaviour (i.e., variation in number-operation use, reflection and comparison of alternative solutions to a problem). Through the cycles of game development the definition of mathematically relevant player behaviour was constantly redefined which also allowed the fine-tuning of game features (goals, rules, story, rewards) intended to support the desired behaviour. Additionally, as gameplay was not used and an incentive, players were not forced to stop the game flow to 'do the math'. This might have helped attain stronger learning outcomes by reducing frustration and increasing engagement with the game.

The lack of transfer in extant studies of game-based learning often is associated with students' difficulties in connecting skills and knowledge practised within the game context to everyday (school) tasks (ter Vrugte & de Jong, 2012; Wouters & van Oostendorp, 2013), but the reason also could be connected to the mismatch between measures of learning outcomes and the desired learning aims (All et al., 2016). To solve this problem, using a combination of different alternative measures is often suggested (All et al., 2014, 2016; Cheung & Slavin, 2013). The present set of studies combined multiple measures and methods of measuring expected mathematical learning outcomes. More qualitative study designs were used in the piloting and development phase, followed by more quantitative approaches in the testing phase. Study designs were selected by considering the specific research questions and stages of development of the NNG. Learning outcomes were measured both directly by game-log data and also with multiple paper-and-pencil measures. A variety of paper-pencil measures were selected to cover different aspects of the mathematical learning outcomes. Thus, paper-and-pencil tasks included close (Arithmetic Production Task) and far (pre-algebra knowledge test) transfer tasks and standardised measures related to general mathematical school practice (arithmetic fluency task). This dissertation's results strongly support the use of multiple measures in studies on the effects of game-based learning (e.g., the differential effects across grades in Study V).

From the perspective of research on adaptivity and flexibility with arithmetic problem solving, the present dissertation provides the first attempt to validate adaptive number knowledge as a component of adaptivity with arithmetic problem solving. Results of both empirical studies (Hickendorff, 2018; Torbeyns, Hickendorff & Verschaffel, 2017; Blöte et al., 2001) and theoretical works (Heinze et al., 2009; Threlfall, 2002, 2009; Verschaffel et al., 2009) in the domain of adaptivity and flexibility with arithmetic show that in spite of the emphasis of developing this type of mathematical proficiency, in reality students mostly use standard, single algorithms even when explicitly instructed to look for alternative solutions (Hickendorff, 2018). Additionally, using 'clever' or 'alternative' solutions is often not associated with higher problem-solving accuracy (Hickendorff, 2018). Therefore, apart from practicing different problem solving strategies there is also a need for the training of additional skills that would allow transfer, and especially to increase students' tendency to consider alternative solution methods to a problem. Under the umbrella of adaptive expertise, the dissertation proposed that adaptivity with arithmetic has procedural flexibility and adaptive number knowledge as two necessary components.

By introducing adaptive number knowledge as a component of adaptivity with arithmetic the present dissertation aimed to broaden the focus of adaptivity and flexibility with arithmetic and explore the underlying mechanism that might enhance students' tendencies to consider and use different solutions when solving arithmetic problems. It was hypothesized that having a higher number of complex numerical relations readily available would enable students to look for alternative solutions for a problem. In turn, this could lead to more flexibility in switching between these solutions and adaptivity in selecting efficient solutions considering problem characteristics and the given context, without relying on a limited set of strategies that are problem-specific. The similar structure of individual differences found in different age groups of elementary school students and adults' adaptive number knowledge suggest that the Arithmetic Production Task can reliably capture participants' adaptive number knowledge. Results of the dissertation showed consistently that aspects of adaptive number knowledge can be enhanced (Study III and V) as students were able to retrieve a larger number (correct solutions) and also more complex (multi-operational solutions) solutions for a given problem as measured by the Arithmetic Production Task.

However, in the context of the present dissertation the relationship of adaptive number knowledge and of general flexibility and adaptivity with arithmetic problem solving remains on a theoretical level. The Arithmetic Production Task was not compared to previously used tests of adaptivity and flexibility with arithmetic as most of these tests focus on the use or lack of certain pre-define 'optimal solutions' for a problem which did not align with the open-ended, in-situ definition of adaptive number knowledge used in the present dissertation. Conversely, open-ended tests which explore solution methods in detail would not have allowed scalability, which was needed in the context of this dissertation. Thus, finding suitable measures, and validating the concept of adaptive number knowledge and exploring the exact mechanism in developing flexibility and adaptivity with arithmetic trough strengthening adaptive number knowledge remains a task for future research.

The iterative cycles of developing and testing the NNG and measures of adaptive number knowledge elicited a refined theoretical description of the notion of adaptive number knowledge in the present dissertation. The theoretical description hypothesised adaptive number knowledge as a construct represented by the relations between the mental networks of natural numbers and arithmetical operations, which allow for the use of flexible and adaptive arithmetical strategies. The present dissertation's empirical results support this hypothesised theoretical model. Extensive practice with various combinations of numbers and operations resulted in improvement on paper-and-pencil tasks that measured adaptive number knowledge. Additionally, the effects from gameplay also translated into improvements in mechanical arithmetic fluency and far-transfer tasks, such as pre-algebra knowledge.

5.2 Practical implications

The efficiency of using technology for educational purposes can be defined by the extent to which the affordances of technology can transform the ways in which students interact with the desired educational content (Bray & Tangney, 2017). Traditionally, training that aims to strengthen students' understanding and representation of numerical characteristics and relations mainly has used skilled teacher support, one-on-one guidance and scaffolding (e.g., Heirdsfield, 2011). While this method can be very effective, it also requires considerable effort and time from teachers, which are not always available. The NNG represents the first attempt to strengthen adaptive number knowledge, using a game-based format that can be applied in classroom settings alongside regular instructional methods. It seems that the NNG can be applied flexibly at different grade levels to primary school mathematics instruction as it can have positive effects on basic arithmetic fluency in lower grades and progressively on more complex mathematical skills and knowledge such as adaptive number knowledge and pre-algebra knowledge. This can provide valuable practical help for teachers when they aim to extend their regular classroom practice.

The Arithmetic Production Task developed within the present dissertation's larger context also provides practical solutions for measuring adaptive number knowledge on a large scale and over different age groups. The Arithmetic Production Task is a novel and flexible method of measuring aspects of adaptivity with arithmetic that was not available before. This can have practical effects both in research practice and classroom assessment.

The results from Studies I and II were used to develop the NNG version used in Study III, and based on Study III's results, the final prototype used in Study V was developed. While an iterative process of game development often is highlighted as a desirable standard, due to a lack of funding or other practical issues, it is rarely done in research practice (Mayo, 2009). The current dissertation is an important addition to the field, as it describes these iterative cycles of NNG testing and development, and shows not only this iterative approach's benefits, but also how design and evaluation can mutually inform each other during these cycles. Future research and developmental work in the game-based learning domain could make use of the iterative design model proposed, for example, by Vanden Abeele and collegues (2012) and the practical implementation of this method presented in the current dissertation.

5.3 Limitations and future studies

One of the present dissertation's general limitations is its lack of recognition the value of qualitative data. While rich qualitative data were collected in Studies I and III (i.e., multiple sources of video data, interview and field notes) and were used to inform the game's development, such data were used only on an anecdotal level in Study I and were not reported at all in Study III. Much of the qualitative data collected in this early phase were used to analyse and understand patterns found in quantitative data and for game-development purposes. However, including more detailed descriptions of the role of the qualitative data would have shown the triangulation of data in pilot studies more clearly and could have strengthened the manuscripts. Unfortunately, current standards and practices in science make it difficult to include this kind of often-valuable information in publications.

Regarding the development of the concept of adaptive number knowledge and measures of adaptive number knowledge, main shortcomings are related to comparing different task versions. As shown in Appendix 3, the Arithmetic Production Task details varied across the different studies and sub-studies both in terms of items and each item's timing. While this is inherent to developing and testing new instruments and was done to improve reliability and validity (which was achieved toward the last few implementations of the measurement; see Appendix 3), this variation also raises questions regarding the direct comparison of results from the different studies and sub-studies. Furthermore, scoring of the Arithmetic Production Task is limited to the numbers of correct and number of multi-operational solutions in the present dissertation. However, related studies show (see McMullen et al., 2017) that different considerations can also be used when scoring the items. Specifically, using more aspects in scoring the Arithmetic Production Task could allow for drawing more complex conclusions regarding the nature of adaptive number knowledge that were impossible in the present dissertation's context.

Measurement methods could also be developed further in relation to players' game performance, which was measured through the total number of maps completed. While this is a reliable measure of students' actual time spent on task, it is also rather crude and does not reveal qualitative differences in students' gameplay. Rich data are available both in terms of log data and Arithmetic Production Task transcripts. Thus, future studies could make use of data-mining techniques to explore different and more-detailed indicators of game performance in combination with additional aspects of adaptive number knowledge. This should help in understanding various player behaviours and strategies, and developing the conceptualisation and measurement of adaptive number knowledge. Additionally, grade level is a crude separator for examining the effects of NNG training on different mathematical learning outcomes. For example, future studies could look at prior knowledge in combination with more detailed game-performance analysis as determinants for mathematical development.

Regarding study design, certain limitations are connected to using a quasi-experimental design for testing the first NNG prototype, as well as to the implementation of NNG training in the classroom and on a large scale. While using a quasi-experimental design was an efficient method for pilot testing, the lack of a control group and the use of a convenience sample strictly limited Study III results' generalisability. Additionally, students used the NNG in pairs in Study III, but the small sample size did not allow for the use of statistical methods that could account for the nested data structure.

In the last NNG testing phase (Study V), it was important to implement the training on a large scale and in the context of regular classroom teaching. Thus, the Finnish teachers, who are accustomed to a large level of autonomy, did not receive strict guidelines for implementation, but merely general guidelines regarding the amount, place and frequency of play sessions using the NNG. While this decision created a more ecologically valid setting, a lack of information about the game's implementation in different classrooms limits opportunities to draw conclusions relevant for practical use of games in mathematics classrooms. A large sample size and the randomised design aimed to control this unknown variation in the implementation. Log files show that in terms of time spent on task, on average, no large differences existed between classes in the experimental group. Comparisons of individual vs. pair gameplay also did not reveal major differences, which shows that the game is rather robust against implementation variation. More data could have been collected (i.e., journals, questionnaires, interviews and observations) from teachers regarding the implementation of the NNG in their classroom practice and teachers' roles during the play sessions (i.e., amount of scaffolding, making explicit connections with the everyday math curriculum). Collecting these additional sources of data could have provided richer information regarding implementation fidelity and, more importantly, information regarding the actual application process of the NNG in the regular classroom practice. Future studies should focus more on the teacher's role and on the method of NNG implementation in curricula and classroom practice, as studies show that game-based learning yields significantly higher efficiency when the added benefits of teachers' support are considered (Hays, 2005; Wouters & van Oostendorp, 2013).

Other limitations are connected to the design of the different NNG versions used in the dissertation. While the game versions tested became more and more complex in terms of game-like elements, such as the story and the relationship between the reward system and the game's progress, or visual design, these developments still were limited. The game versions used in the present dissertation are all prototypes and not final products intended for everyday application by the larger public. This could have affected the development of the measured mathematical learning outcomes in the present dissertation. Evidence indicates that changes in game design over the different versions of the NNG affect players' game experience (Bui et al., 2018). Thus, an objective for future studies could be to investigate in more depth how different game design decisions (e.g., adding more game-like elements) affect gameplay and the game experience, and how these aspects affect mathematical learning outcomes.

5.4 Conclusions

Findings from the present dissertation show that the NNG can provide efficient and novel training opportunities for developing primary school students' adaptive number knowledge and related mathematical skills and knowledge. Results suggest that choosing the game-based format for this training was meaningful, as it allowed for a large amount of open-ended practice with various number-operation combinations, which usually is difficult to achieve in a regular mathematics classroom. Results of this dissertation provide empirical evidence for the theoretical model of adaptive number knowledge as a component of adaptivity with arithmetic. Additionally, a new and scalable measurement tool and a novel training on adaptive number knowledge were developed. From the perspective of educational game design, the results highlight the importance of iterations with the following central aspects to consider during the process: the relationship between game features and educational content, in-game and outside-game measures of learning outcomes, and the application of game-based training in everyday classroom practice.

REFERENCES

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games*, *1*(2), 3–21. https://doi.org/10.17083/ijsg.v1i2.18

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2015). Towards a conceptual framework for assessing the effectiveness of digital game-based learning. *Computers & Education*, *88*, 29–37. https://doi.org/10.1016/j.compedu.2015.04.012

All, A., Nuñez Castellar, E. P., & Van Looy, J. (2016). Assessing the effectiveness of digital game-based learning: Best practices. *Computers & Education*, *92–93*, 90–103. https://doi.org/10.1016/j.compedu.2015.10.007

Bakker, M., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2015). Effects of playing mathematics computer games on primary school students' multiplicative reasoning ability. *Contemporary Educational Psychology*, *40*, 55–71. https://doi.org/10.1016/j.cedpsych.2014.09.001

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In *The Development of Arithmetic Concepts and Skills: Constructive Adaptive Expertise* (pp. 1–33). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410607218

Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in dutch second grades. *Journal for Research in Mathematics Education*, *24*, 294-323. doi:10.2307/749464

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

Blöte, A. W., Klein, A. S., & Beishuizen, M. (2000). Mental computation and conceptual understanding. *Learning and Instruction*, *10*(3), 221–247. https://doi.org/10.1016/S0959-4752(99)00028-6

Blöte, A. W., van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology*, *93*(3), 627–638. https://doi.org/10.1037/0022-0663.93.3.627

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers &Education*, *94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Bray, A., & Tangney, B. (2017). Technology usage in mathematics education research – A systematic review of recent trends. *Computers & Education*, *114*, 255–273. https://doi.org/10.1016/j.compedu.2017.07.004

Breuer, J. J., & Bente, G. (2010). Why so serious? On the relation of serious games and learning. *Eludamos. Journal for Computer Game Culture*, *4*(1), 7–24. https://doi.org/10.1177/1461444812450426

Bui, P.; Rodríguez-Aflecht, G.; Brezovszky, B.; Hannula-Sormunen, M. M.; Laato, S.; Lehtinen, E. (2018). *Understanding Students' Game Experiences throughout the Developmental Process of the Number Navigation Game.* Manuscript submitted for publication.

Byun, J., & Joung, E. (2018). Digital game-based learning for K-12 mathematics education: A meta-analysis. *School Science and Mathematics*, *118*(3–4), 113–126. https://doi.org/10.1111/ssm.12271

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2003). Patterns of knowledge in children's addition. *Developmental Psychology*, *39*(3), 521–534. https://doi.org/10.1037/0012-1649.39.3.521

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

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Connolly, T. M., Boyle, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers &Education*, *59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation and Gaming*, *41*(6), 898–920. https://doi.org/10.1177/1046878110390784

Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience.* Harper Perennial, New York.

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, *46*(3), 249–264. https://doi.org/10.1016/j.compedu.2005.11.007

Devlin, K. (2011). *Mathematics education for a new era: Video games as a medium for learning*. A K Peters.
Finnish National Agency for Education (2014). Core curriculum for basic Education 2014. http://www.oph.fi/english/curricula_and_qualifications/basic_education/curricula_2014.
Fisher, K., Hirsh-Pasek, K., Golinkoff, R. M., Singer, D. G., & Berk, L. (2010). *Playing Around in School: Implications for Learning and Educational Policy*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195393002.013.0025
Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming*, *33*(4), 441–467. https://doi.org/10.1177/1046878102238607
Gunter, G. A., Kenny, R. F., & Vick, E. H. (2008). Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games. *Educational Technology Research and Development*, *56*(5–6), 511–537. https://doi.org/10.1007/s11423-007-9073-2
Habgood, M. P. J. (2007). *The Effective Integration of Digital Games and Learning Content.* (Unpublished doctoral dissertation). University of Nottingham, Nottingham, UK. Retrieved from: *https://eprints.nottingham.ac.uk/10385/1/Habgood_2007_Final.pdf*
Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. *Journal of the Learning Sciences*, *20*(2), 169–206. https://doi.org/10.1080/10508406.2010.508029
Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, *102*(February 2009), 202–223. https://doi.org/10.1016/j.compedu.2016.09.001
Hatano, G., & Oura, Y. (2003). Commentary: Reconceptualizing School Learning Using Insight From Expertise Research. *Educational Researcher*, *32*(8), 26–29. https://doi.org/10.3102/0013189X032008026
Hays R.T. (2005). *The effectiveness of instructional games: a literature review and discussion.* Orlando, FL: Naval Air Warfare Center. Retrieved from: http://handle.dtic.mil/100.2/ADA441935
Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *ZDM - International Journal on Mathematics Education*, *41*(5), 535–540. https://doi.org/10.1007/s11858-009-0214-4
Heirdsfield, A. M., & Cooper, T. J. (2004). Factors affecting the process of proficient mental addition and subtraction: Case studies of flexible and inflexible computers. *Journal of Mathematical Behavior*, *23*(4), 443–463. https://doi.org/10.1016/j.jmathb.2004.09.005
Heirdsfield, A. M. (2011). Teaching mental computation strategies in early mathematics. *Young Children*, *66*(2), 96–102. Retrieved from: https://eprints.qut.edu.au/42451/
Hickendorff, M. (2018). Dutch sixth graders' use of shortcut strategies in solving multidigit arithmetic problems. *European Journal of Psychology of Education,* 33 (4), 577–594. https://doi.org/10.1007/s10212-017-0357-6
Ilgaz, H., Hassinger-Das, B., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). Making the Case for Playful Learning (pp. 1245–1263). Springer, Dordrecht. https://doi.org/10.1007/978-94-024-0927-7_64
Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education*, *51*(4), 1609–1620. https://doi.org/10.1016/j.compedu.2008.03.003
Ke, F. (2009). A Qualitative Meta-Analysis of Computer Games as Learning Tools. In: Ferdig, R. E. (Ed.) *Handbook of Research on Effective Electronic Gaming in Education*, 1–32. Hershey, New York: Information Science Reference.
Ke, F., & Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology*, *44*(2), 225–242. https://doi.org/10.1111/j.1467-8535.2012.01326.x
Kebritchi, M., & Hirumi, A. (2008). Examining the pedagogical foundations of modern educational computer games. *Computers & Education*, *51*(4), 1729–1743. https://doi.org/10.1016/j.compedu.2008.05.004
Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus gradual program design. *Journal for Research in Mathematics Education, 29*, 443-464. doi:10.2307/749861
Laski, E. V., & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology*, *50*, 853-864. doi:10.1037/a0034321
Leemkuil, H. & de Jong, T. (2012). Adaptive advice in learning with a computer based strategy game. *Academy of Management Learning and Education*, 11, 653-665. https://doi.org/10.5465/amle.2010.0141
Lehtinen, E., Hannula-Sormunen, M., McMullen, J., & Gruber, H. (2017). Cultivating mathematical skills: from drill-and-practice to deliberate practice. *ZDM - Mathematics Education*, *49*(4), 625–636. https://doi.org/10.1007/s11858-017-

0856-6
Li, Q., & Ma, X. (2010). A Meta-analysis of the Effects of Computer Technology on School Students' Mathematics Learning. *Educational Psychology Review*, *22*(3), 215–243. https://doi.org/10.1007/s10648-010-9125-8
Mayo, M. J. (2009). Video Games : A Route to Large-Scale. *Science.* 323(5910), 79–82. https://doi.org/10.1126/science.1166900
McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual. Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside. Retrived from: https://www.wjscore.com/Files/WJIVTechnicalManual.PDF
McMullen, J., Brezovszky, B., Hannula-Sormunen, M. M., Veermans, K., Rodríguez-Aflecht, G., Pongsakdi, N., & Lehtinen, E. (2017). Adaptive number knowledge and its relation to arithmetic and pre-algebra knowledge. *Learning and Instruction*, *49*, 178–187. https://doi.org/10.1016/j.learninstruc.2017.02.001
National Council of Teachers of Mathematics (2014). *Principles to actions: Ensuring mathematical success for all.* Reston.
Nunes, T., Dorneles, B. V., Lin, P.-J., & Rathgeb-Schnierer, E. (2016). *Teaching and Learning About Whole Numbers in Primary School*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-45113-8
OECD (2010), *Inspired by Technology, Driven by Pedagogy: A Systemic Approach to Technology-Based School Innovations*, Educational Research and Innovation, OECD Publishing, Paris, Retrieved from: https://doi.org/10.1787/9789264094437-en
OECD (2018), *Teaching for the Future: Effective Classroom Practices To Transform Education*, OECD Publishing, Paris, Retrieved from: https://doi.org/10.1787/9789264293243-en
Pellegrini, A. (2009). *The Role of Play in Human Development*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367324.001.0001
Pepin, B., & Rösken-Winter, B. (Eds.) (2015). *From beliefs to dynamic affect systems in mathematics education: exploring a mosaic of relationships and interactions*. (Advances in mathematics education). Cham: Springer. DOI: 10.1007/978-3-319-06808-4
Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.
Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of Game-Based Learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533
Pope, H., & Mangram, C. (2015). Wuzzit Trouble: The Influence of a Digital Math Game on Student Number Sense. *International Journal of Serious Games*, *2*(4). https://doi.org/10.17083/ijsg.v2i4.88
Reed, H. C., Stevenson, C., Broens-Paffen, M., Kirschner, P. A., & Jolles, J. (2015). Third graders' verbal reports of multiplication strategy use: How valid are they? *Learning and Individual Differences*, *37*, 107–117. https://doi.org/10.1016/J.LINDIF.2014.11.010
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561
Rittle-Johnson, B., & Star, J. R. (2009). Compared With What? The Effects of Different Comparisons on Conceptual Knowledge and Procedural Flexibility for Equation Solving. *Journal of Educational Psychology*, *101*(3), 529–544. https://doi.org/10.1037/a0014224
Rodríguez-Aflecht, G. (2018). *Exploring Motivational Effects of a Mathematics Serious Game.* (Unpublished doctoral dissertation). University of Turku, Turku, Finland. Retrieved from: http://urn.fi/URN:ISBN:978-951-29-7298-2
Rogers, C. S., & Sawyers, J. K. (1988). *Play in the lives of children*. National Association for the Education of Young Children.
Salen, K., & Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*. The MIT Press.
Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, *47*(6), 1525–1538. https://doi.org/10.1037/a0024997
Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games-but not circular ones-improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*, 545-560. doi:10.1037/a0014239
Slavin, R. E., & Lake, C. (2008). Effective Programs in Elementary Mathematics : A Best-Evidence Synthesis. *Review of Educational Research*. 78(3), 427-515. DOI: 10.3 102/00346543083 17473
Sowder, J. (1992). Estimation and number sense. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 371–389). New York: Macmillan
Squire, K. D. (2008). Video game–based learning: An emerging paradigm for instruction. *Performance Improvement Quarterly*, *21*(2), 7–36. https://doi.org/10.1002/piq.20020
Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*,

18(6), 565–579. https://doi.org/10.1016/j.learninstruc.2007.09.018

Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology*, *31*(3), 280–300. https://doi.org/10.1016/j.cedpsych.2005.08.001

Sutton-Smith, B. (1997). *The ambiguity of play*. Harvard University Press.

ter Vrugte, J., & de Jong, T. (2012). How to adapt games for learning: The potential role of instructional support. In S. Wannemacker, S. Vandercruysse, & G. Clarebout (Eds.), *Serious games: The challenge* (Vol. 280, pp. 1-5). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-642-33814-4_1

Threlfall, J. (2002). Flexible Mental Calculation. *Educational Studies in Mathematics*, *50*(1), 29–47. https://doi.org/10.1023/A:1020572803437

Threlfall, J. (2009). Strategies and flexibility in mental calculation. *ZDM*, *41*(5), 541–555. https://doi.org/10.1007/s11858-009-0195-3

Tobias, S., & Fletcher, D. (2012). Learning from Computer Games: A Research Review (pp. 6–17). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33814-4_2

Torbeyns, J., de Smedt, B., Stassens, N., Ghesquière, P., & Verschaffel, L. (2009). Solving Subtraction Problems by Means of Indirect Addition. *Mathematical Thinking and Learning*, *11*(1–2), 79–91. https://doi.org/10.1080/10986060802583998

Torbeyns, J., Hickendorff, M., & Verschaffel, L. (2017). The use of number-based versus digit-based strategies on multi-digit subtraction: 9–12-year-olds' strategy use profiles and task performance. *Learning and Individual Differences, 58,* 64-74. http://dx.doi.org/10.1016/j.lindif.2017.07.004

Torbeyns, J., Lehtinen, E., & Elen, J. (2015). Describing and Studying Domain-Specific Serious Games: Introduction. In J. Torbeyns, E. Lehtinen, & J. Elen (Eds.). *Describing and studying domain-specific serious games* (pp. 1–6). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20276-1_1

van den Heuvel-Panhuizen, M., Kolovou, A., & Robitzsch, A. (2013). Primary school students' strategies in early algebra problem solving supported by an online game. *Educational Studies in Mathematics*, *84*(3), 281–307. https://doi.org/10.1007/s10649-013-9483-5

Vanden Abeele, V., De Schutter, B., Geurts, L., Desmet, S., Wauters, J., Husson, J., … Geerts, D. (2012). P-III: A Player-Centered, Iterative, Interdisciplinary and Integrated Framework for Serious Game Design and Development. In: De Wannemacker S., Vandercruysse S., Clarebout G. (Eds.) *Serious Games: The Challenge.* (pp. 82-86). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-33814-4_14

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, *24*(3), 335–359. https://doi.org/10.1007/BF03174765

Whitton, N. (2010). *Learning with digital games: A practical guide to engaging students in higher education.* New York, NY: Routledge.

Villarreal, V. (2015). Test Review: Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV Tests of Achievement. *Journal of Psychoeducational Assessment*, 33(4), 391–398. https://doi.org/10.1177/0734282915569447

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., … Conkey, C. (2009). Relationships Between Game Attributes and Learning Outcomes. *Simulation & Gaming*, *40*(2), 217–266. https://doi.org/10.1177/1046878108321866

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of "The number race", an adaptive computer game for remediation of dyscalculia. *Behaioral and Brain Functions, 2*, 1-16. doi:10.1186/1744-9081-2-20

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, *60*(1), 412–425. https://doi.org/10.1016/j.compedu.2012.07.018

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

Young, M. F., Slota, S., Cutter, B., R., Jalette, G., Mullin, G., … Yukhymenko, M. (2012). Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, *82*(1), 61–89. https://doi.org/10.3102/0034654312436980

Young, J. (2017). Technology-enhanced mathematics instruction: A second-order meta-analysis of 30 years of research. *Educational Research Review*, *22*, 19–33. https://doi.org/10.1016/j.edurev.2017.07.001

APPENDICES

Appendix 1.

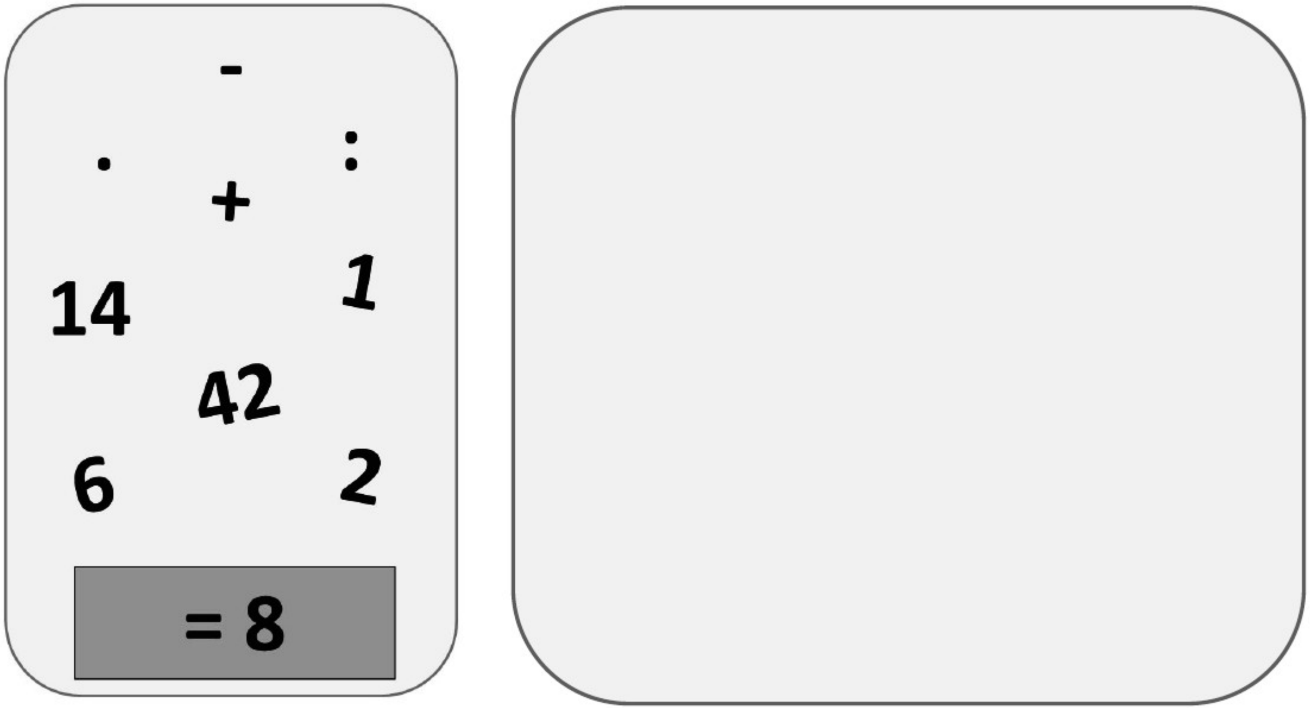
Summary of the main findings, conclusions and suggestions of five extensive reviews/meta-analyses in game-based learning from the past ten years.

Year/Author Period (n studies)	Focus	Main findings	Main conclusions/suggestions
2012/Young et al./ n.a. (39)	- academic performance in five subjects	- strongest results in languages and physical education and no evidence in science and math - lots of variability in the field (type of training, study design) → no meta-analyses can be conducted - general questions regarding the media's effectiveness are not productive	- game-based training should be integrated in the curriculum, and instructional support should be provided - longitudinal study designs are needed - assessment should be both inside and outside the game - game design should match game goals and learning goals
2012 Connolly et al. 2004-2009 (129)	- cognitive gains, skills acquisition and motivation in secondary school	- large variability in the field (type of training, study design, learning outcome) - not enough 'strong' evidence for the educational potential of games - most frequent study design is quasi-experimental with few RCTs	- system of categorisation based on genre, type of technology and learning outcomes is offered - further breakdown of these main themes should be used - more RCTs are needed to draw more reliable conclusions
2013/Wouters et al. 1990-2012	- cognitive and motivational effects of educational games (meta-analyses)	- when compared with traditional instruction game-based learning, was more effective in terms of cognitive gains, but not motivation - stronger effects if training is longer and if additional instructional support is used - quality of study design can affect outcomes (smaller effect sizes for RCTs)	- evaluation should consider alternative methods to capture different aspects of learning - more research is needed in game design and instructional design - future reviews should consider specific game features' effects
2016/Hainey et al. 2000-2013 (45)	- cognitive, affective and behavioural effects of game-based learning in primary school - empirical studies	- lots of variability in genre and platform - lack of longitudinal designs and RCTs - knowledge acquisition and content understanding were the most common learning objectives - PE, math, science, language and social studies were the most popular subjects	- more complex learning aims and specific research questions - more longitudinal study designs and RCTs
2016/Boyle et al. 2009-2014 (143)	- update to Connolly et al., 2012, using the same framework	- STEM games were most popular - compared with Connolly et al., 2012: similar distribution of study designs (mostly quasi-experimental and few RCT); larger number of relevant studies in shorter time frame; still lots of variability (aims, genre, design, domain); there is more focus on exploring different game features	- more complex learning aims and specific research questions - an evaluation framework is needed for the different types of learning outcomes of GBL - clarity in terminology (e.g., game-based learning and serious games)

Appendix 2.

Example item of the Arithmetic Production Task.

*Try to **make as many different problems** where the solution is **8** as you can. Use only the numbers in the box. You can use each number as many times as you want. You can use addition, subtraction, multiplication, and division as many times as you want.*



Appendix 3.

List of items, timing, and Cronbach's alpha (α) reliability scores of the different versions of the Arithmetic Production Task (correct solutions and multi-operational solutions) used in Study II, III and V.

Study	Given numbers	Target	Seconds/item	Correct (α)	Multi-op. (α)
IIa (grade 3-5)	2,4,8,12,32	16	60	.56	.51
	1,2,3,5,30	59			
	3,4,5,6	63			
	2,4,8,10	22			
IIb (adults)	2,4,8,12,32	16	60	.68	.49
	1,2,3,5,30	59			
	120, 180, 3, 5, 30	12			
	6,5,4,3	126			
IIc (grade 6)	2,4,8,12,32	16	90	.55	.67
	1,2,3,5,30	59			
	2,4,8,10	22			
	3,4,5,6	63			
III pre-test (grade 6)	Same as IIc	-	-	-	-
III post-test (grade 6)	2,4,6,16,24	12	90	.74	.72
	1,2,4,5,40	79			
	3,5,30,120,180	12			
	3,4,5,6	126			
V pre-test (grade 4-6)	1,2,3,4 (example item)	6	90	.70	.63
	2,4,8,12,32	16			
	1,2,3,5,30	59			
	2,4,6,16,24	12			
	3,5,30,120,180	12			
V post-test (grade 4-6)	1,2,3,4 (example item)	6	90	.86	.80
	2,3,9,15,36	18			
	1,2,6,14,42	8			
	2,4,6,16,48	24			
	1,2,3,30,36	14			
	2,6,8,32,54	18			
	3,4,5,6	7			
	2,3,6,10,18	38			

Appendix 4.

List of items used in the pre-algebra knowledge test in Study V

Instructions:

- **Pre-test A:** *Choose which number goes in the empty space so that both sides of the equals sign will be the same.*
- **Post-test A:** same as for pre-test
- **Post-test B:** *Write the number in the empty space so that both sides of the equals sign (=) will be the same.*

Items	Pre-test	Post-test	
	A	A	B
1	12 + ___ = 11 + 15 a) 16 b) 5 c) 15 d) 14	___ : 2 + 12 = 15 + 11 a) 16 b) 40 c) 28 d) 14	12 + ___ = 11 + 15
2	6 · ___ = 2 · 15 a) 5 b) 24 c) 6 d) 4	6 · ___ + 6 = 2 · 18 a) 5 b) 12 c) 6 d) 4	6 · ___ = 2 · 15
3	___ · 8 = 2 · 16 a) 3 b) 2 c) 4 d) 16	___ · 8 = 96 : 2 a) 3 b) 2 c) 4 d) 6	__ · 8 = 2 · 16
4	54 - 48 = 18 : ___ a) 48 b) 6 c) 3 d) 2	54 - 36 = 6 · ___ a) 2 b) 3 c) 4 d) 5	54 - 48 = 18 : ___
5	___ : 7 = 6 : 3 a) 8 b) 14 c) 21 d) 7	___ : 7 = 15 : 5 a) 18 b) 21 c) 35 d) 14	__ : 7 = 6 :
6	27 - ___ = 50 : 2 a) 2 b) 7 c) 5 d) 0	31 - ___ = 58 : 2 a) 3 b) 7 c) 5 d) 2	27 - ___ = 50 : 2

Annales Universitatis Turkuensis

UNIVERSITY OF TURKU

Grano Oy - Turku, Finland 2019

ISBN 978-951-29-7621-8 (PRINT)
ISBN 978-951-29-7622-5 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)