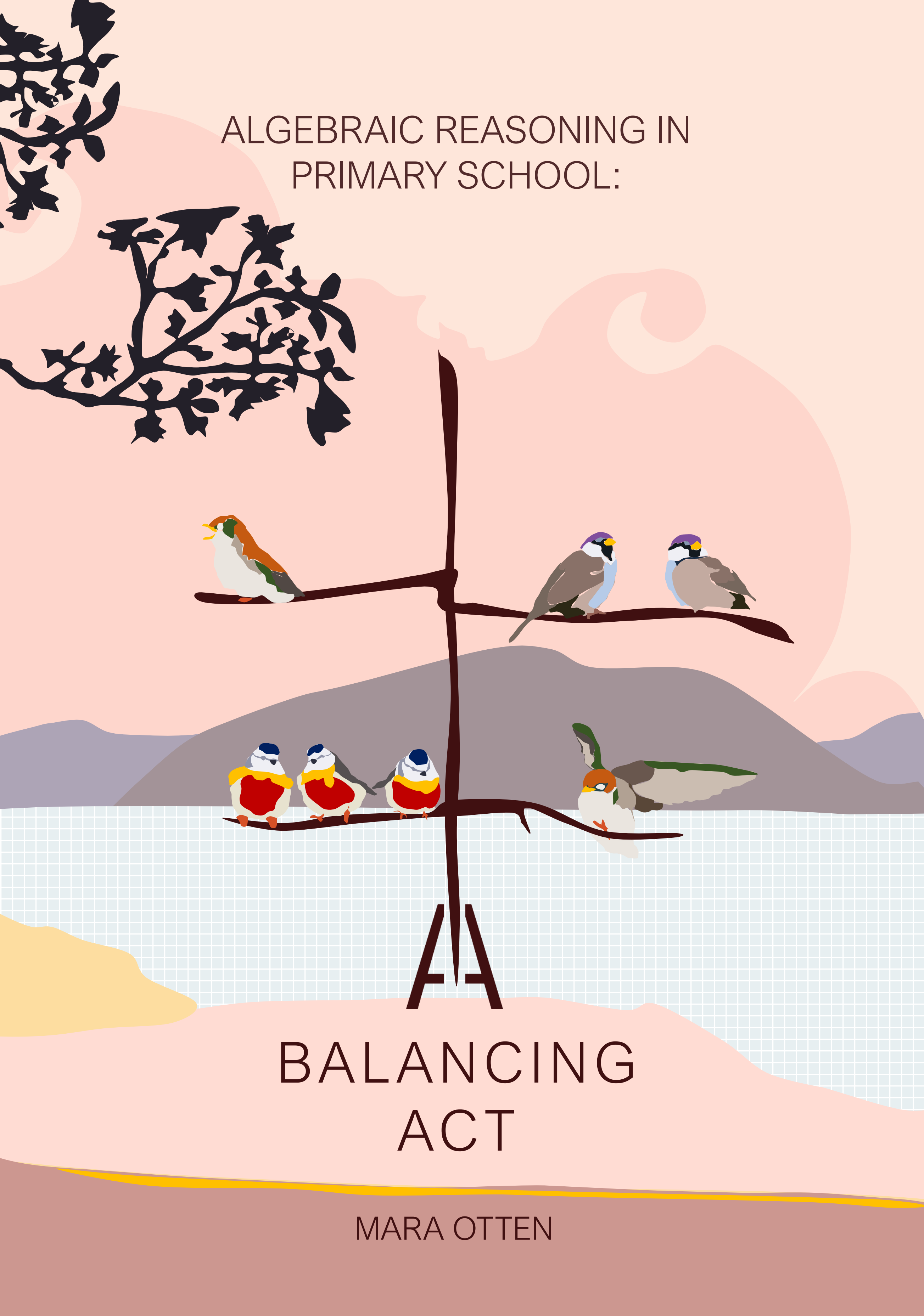
ALGEBRAIC REASONING IN PRIMARY SCHOOL:



BALANCING ACT

MARA OTTEN

Algebraic reasoning in primary school:
A balancing act

Mara Otten

Cover design and lay-out
Carolien Duijzer & Mara Otten

Print
Ridderprint, Alblasserdam
ISBN: 978-94-6416-113-7

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

The research reported in this dissertation was funded by the Netherlands Initiative for Education Research (Nationaal Regieorgaan Onderwijsonderzoek, NRO; projectnumber 405-14-303).

Algebraic reasoning in primary school: A balancing act

Algebraïsch redeneren op de basisschool:
Zoeken naar balans
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 30 oktober 2020 des ochtends te 11.00 uur

door

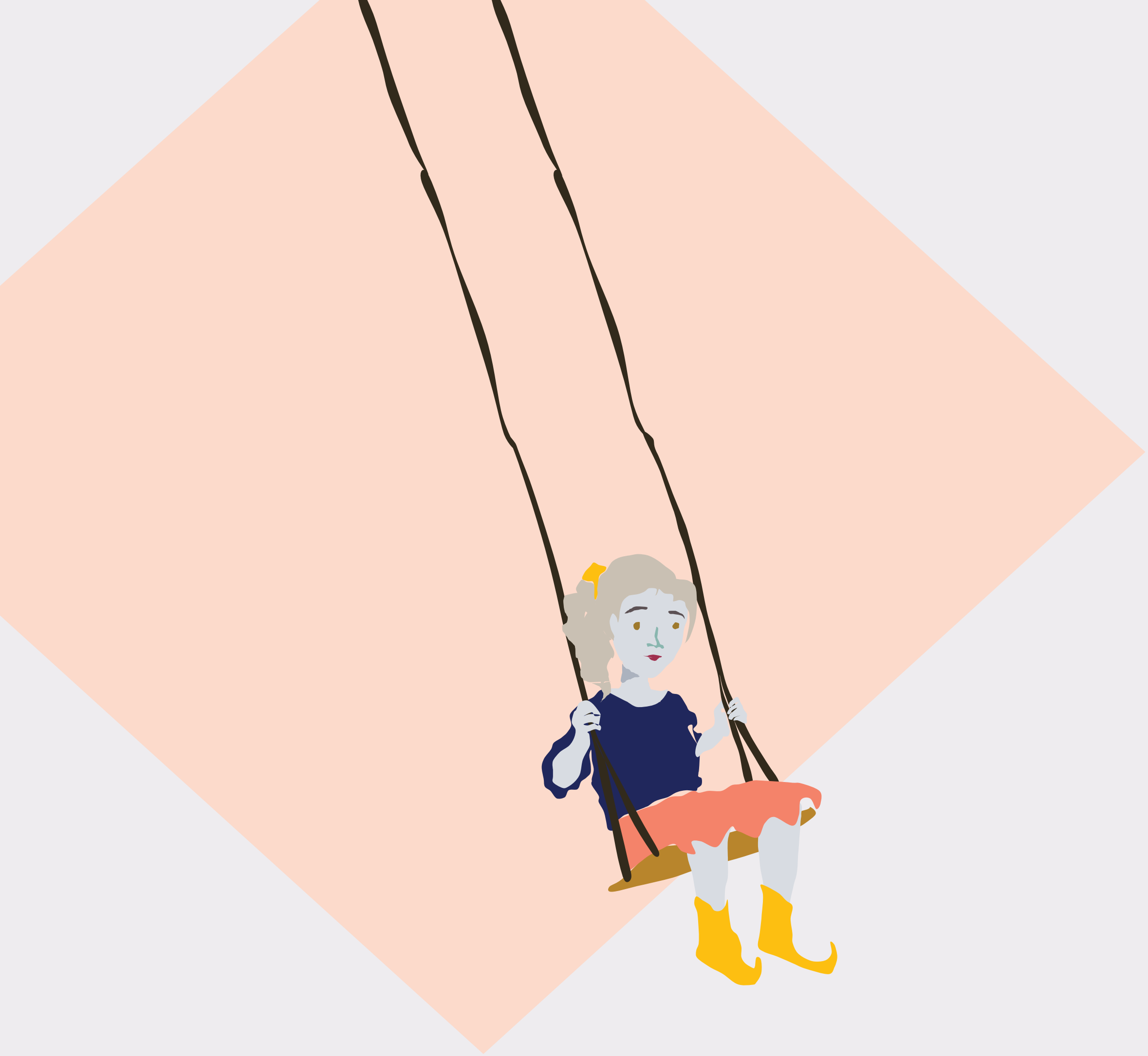
Mara Otten

geboren op 14 december 1990
te Haarlem

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

Copromotor:
Dr. M. Veldhuis

Table of contents



CHAPTER 1

Introduction

Introduction

Haarlem, December 1999

> Mara: Dad, dad, please give me a problem to solve!

When I was young, my two younger brothers and I would sneak into my parents' bed early in the weekend mornings. During those moments, I often asked my father for a problem to solve. Every time he came up with new problems and as I got older the problems got more complicated. The ones I liked most were the problems that were some sort of a puzzle to me. For example, one day he came up with a problem related to my favorite computer game: RollerCoaster Tycoon©. The purpose of this computer game was to create a theme park by building attractions and taking care of the visitors. These visitors had to pay an entrance fee and in addition they had to pay for each ride.

> Dad: Yesterday you were playing RollerCoaster Tycoon, right? Now imagine the following problem: A family of four comes to visit your theme park. They all pay the entrance fee and they all go three times into the same rollercoaster. The price of one rollercoaster ride is half the entrance fee. In total, they have to pay 100 guilders. Can you figure out the price of one rollercoaster ride?

In this example, my father made use of an attractive problem, situated in a context that was very meaningful to me. Due to the many experiences I had playing this computer game, as well as other experiences such as going somewhere with my own family and having to pay a total price for several things combined, I could imagine what this problem implied. In other words, I already had an intuitive understanding of the problem, which was helpful to deduce the underlying mathematical structure. The problem, which could be represented by two linear equations (four entrance tickets and twelve rides together cost 100 guilders; the price of one ride is half the price of an entrance ticket), could be solved in different ways. As a little girl, I solved the problem by making use of various context-connected strategies. I started, for example, with reasoning that if four people had to pay 100 guilders, the costs per person thus had to be 25 guilders. As each person paid the entrance fee and three rollercoaster rides, this implied that these four things combined had to cost 25 guilders. While keeping in mind that the price of one ride was half the price of an entrance ticket, from this point on, I could try various prices until obtaining the total price of 100 guilders. After solving the problem, my dad and I would discuss the steps I had taken to solve it and which alternative solution strategies would have been possible. In this case, for example, replacing a part of the first equation on the basis of information from the second equation (i.e., making use of substitution) could have been a strategy as well.

What I did not know at that time, but what I do realize now, is that by solving these puzzles, I was actually doing mathematics. The problems my father gave me required all kinds of elements of informal forms of algebraic reasoning, such as understanding of relationships, combining relationships, examining ideas, drawing inferences, justifying actions, and reasoning logically. While, at that time (I was about nine years old), I had not received instruction on algebra at school, I was very well able to reason about and to solve such problems — which also have been described as "early algebra" problems (Carraher et al., 2008).

Developing algebraic reasoning is one of the major goals of mathematics education (e.g., National Council of Teachers of Mathematics [NCTM], 2000). Algebra is considered a gateway to understanding of science, statistics, business, or technology (Katz, 2007; Schoenfeld, 1995). It moreover is viewed as crucial for later achievements in our global economy of the 21st century (Vogel, 2008). Acquiring these rather sophisticated reasoning skills, which have also been referred to as 21st century skills (e.g., Binkley et al., 2012) or higher-order thinking (HOT) skills (e.g., Lewis & Smith, 1993) has recently received increased attention, also within the domain of mathematics (e.g., Alexander et al., 2011). Internationally, there is consensus that a foundation for HOT skills should be laid in primary school (Common Core State Standards Initiative [CCSSI], 2010; Goldenberg et al., 2003; NCTM, 2000; Organization for Economic Co-operation and Development [OECD], 2003). For the domain of algebra, several studies have provided evidence that primary school students can be engaged in algebraic reasoning activities and that algebraic reasoning can be integrated in the primary school mathematics classrooms (e.g., Brizuela & Schliemann, 2004; Kaput et al., 2008). In this way, students' HOT skills can be fostered at an early stage.

Compared to the international situation, in the Netherlands it took a little longer before attention was paid to offering elementary-grade students opportunities for developing mathematical reasoning through, for example, engaging them in solving informal algebraic problems or other non-routine problems. Only when a small-scale study (Van den Heuvel-Panhuizen & Bodin-Baarends, 2004) revealed that even the highest achievers in mathematics in grade four had difficulties to solve problems like the "theme park" problem, it became clear that this was a blind spot in the Dutch primary school curriculum. A textbook analysis of six primary school mathematics textbook series confirmed this (Kolovou et al., 2009). The textbooks contained almost no problems for developing students' HOT skills. To explore possibilities for how primary school students can be engaged in algebraic reasoning activities, an online game was developed, evoking students to deal with covarying quantities. Participating in this game resulted in positive effects on students' performance on early algebra tasks (Kolovou et al., 2013; Van den Heuvel-Panhuizen et al., 2013).

This PhD thesis emerged more or less from these earlier studies. The main goal was to gain insight in whether, in what ways, and to what extent primary school students' early algebraic reasoning can be fostered as an approach to incorporate HOT in Dutch classrooms. Within this thesis, we focused on fifth-grade students' reasoning about solving linear equations; a topic of which various studies have revealed that primary school students can deal with it (e.g., Blanton et al., 2015; Brizuela & Schliemann, 2004).

1. Theoretical background

1.1 Higher-order thinking skills in primary education

In today's rapidly changing world, with an abundance of information available and with advanced communication technologies all around, it is essential to develop sophisticated skills such as interpreting information and drawing conclusions, identifying and analyzing arguments, synthesizing information, evaluating statements, making inferences, and explaining and justifying procedures (Binkley et al., 2012; Forster, 2004). These skills are important in this time and age and are therefore frequently referred to as 21st century skills (Binkley et al., 2012; Dede, 2010). Developing 21st century skills requires HOT, and the expressions 21st century skills and HOT skills are often used interchangeably. Resnick (1987) defined HOT as "a cluster of elaborative mental processes requiring nuanced judgement and analysis" (p. 44). Thomas and Thorne (2009) provided the following definition of HOT:

> Higher order thinking (HOT) is thinking on a level that is higher than memorizing facts or telling something back to someone exactly the way it was told to you. HOT takes thinking to higher levels than restating the facts and requires students to do something with the facts — understand them, infer from them, connect them to other facts and concepts, categorize them,

> manipulate them, put them together in new or novel ways, and apply them as we seek new solutions to new problems (p. 1).

Within both these definitions, it comes to the fore how HOT differs from lower-order thinking skills such as memorization or routine application of previously learned knowledge (Newmann, 1991). A similar division can be found in the Taxonomy of Educational Objectives of Bloom et al. (1956), which is also often referred to in the context of HOT. This taxonomy consists of an ordering of cognitive skills, knowledge, comprehension, application, analysis, synthesis, and evaluation, with the last three levels being classified as higher order.

Most conceptualizations of HOT highlight the historical tendency to conceptualize HOT in a domain-general way, reflecting the view that HOT is about applying general principles of (logical) reasoning to any academic domain (e.g., Leighton, 2004). Yet, others have taken an opposite position, advocating that thinking becomes higher-order due to expertise people have within a particular academic domain (e.g., Alexander et al., 2011; Ericsson, 2003). Per this domain-specific view, HOT originates from — and is entangled with — specific academic domains, such as the domain of mathematics. Alexander and et al. (2011) therefore emphasize the necessity of including the role of the domain in the description of HOT. In their view, HOT "exhibits distinctive qualities arising from the nature of the domain within which the task or activity is situated" (p. 53).

Challenging mathematics activities could offer a fruitful starting point for teaching HOT. Several researchers have pointed out that activities which are prone to elicit HOT, such as non-routine problem-solving activities, should be part of mathematics education (e.g., Goldenberg et al., 2015; Schoenfeld, 2013). In order to gain knowledge about how HOT can be supported within mathematics education, the research within this thesis focussed on stimulating primary school students' reasoning related to solving linear equations. This domain offers many possibilities for developing HOT. Consider, as an example, the "theme park" problem in the beginning of this chapter. Reasoning about such an informal mathematical problem requires skills like understanding of relationships, interpreting and connecting information, reasoning adaptively, and drawing inferences. Such skills, especially for young students without experience in algebra, can be considered HOT skills (e.g., Goldenberg et al., 2015).

1.2 Early algebra and linear equation solving

Early algebra refers to the introduction of algebraic thinking in the elementary grades (Kaput, et al., 2008). In this way, algebra is treated as a longitudinal strand throughout K-12 mathematics, starting in the elementary grades with solving informal algebraic problems that build on students' natural ways of thinking and understanding of mathematical patterns and relationships and evolving into more formal algebra in later grades (e.g., Carraher et al., 2008; Kieran et al., 2016; NCTM, 2000). Introducing early algebra activities within the elementary grades and connecting these experiences with advanced algebraic concepts in later grades, provides students a basis for developing algebraic understanding in the course of school education (e.g., Carraher & Schliemann, 2014).

Kaput (2008) proposed that algebraic thinking comprises two core aspects: making and expressing generalizations in increasingly formal and conventional symbol systems, and reasoning with symbolic forms. He further argued that these two practices can take place within three content strands: generalized arithmetic, functional thinking, and modeling. Within the first content strand, the generalized arithmetic strand, the focus is on the structures and relations arising in arithmetic. It includes generalizing arithmetic operations and their properties, and building generalizations and reasoning about numbers or relationships. Examples of algebraic concepts that are embedded in this content strand are numbers, operations, properties, equality, expressions, and equations (Chimoni et al., 2018; Kieran et al., 2016). Within the functional thinking strand, the focus is on generalizing towards the idea of function. Activities with identifying and completing patterns often take a central position within this strand, with a key role for the ideas of systematic variation and change (Kaput, 2008). Associated algebraic concepts are, for example, change, covariance, variable, and equation (Chimoni et al., 2018; Kieran et al., 2016). Lastly, the third content strand contains different types of modeling activities. This strand is often interwoven with the first two strands by means of various problem contexts (Kieran et al., 2016), making that all the above concepts are also incorporated in this content strand (Chimoni et al., 2018). Engaging students in early algebra activities implies that, from early on, students can be involved in problem-solving activities related to these content strands.

Linear equation solving is an important topic within school algebra. An equation is a mathematical statement in which the expressions on both sides of the equal sign represent the same value. In that sense, both sides of the equation are equal (e.g., Kieran, 1981). The use of context-based problems, which can be represented by an equation, opens up possibilities for stimulating young students' reasoning related to linear equation solving (e.g., Brizuela & Schliemann, 2004; Van Reeuwijk, 1995). The "theme park" problem presented at the beginning of this chapter is an example of a context-based problem which, in this case, can be represented by two linear equations. When solving this problem, students first can come up with a hypothesis or an idea about the underlying problem statement, which can be translated into a mathematical relationship. Subsequently, students can solve this problem through reasoning which is still closely linked to the problem context (like I did as a little girl). When students get experience in algebra and improve in their level of understanding, they can transform a context problem into a symbolic problem (for the "theme park" problem: $4x + 12y = 100$ and $0.5x = y$, with x representing the entrance fee and y representing the price of one rollercoaster ride) and solve it by symbolic manipulations.

When solving an equation, the goal is to determine the relationships between unknowns and numbers for which the expressions on both sides of the equation are equal. To this aim, the unknown needs to be isolated, which can be achieved through performing operations on the expressions on both sides of the equation while making sure that their equality is maintained. The notion of equality and the strategies for maintaining this equality fulfill a crucial role within linear equation solving (e.g., Bush & Karp, 2013; Kieran et al., 2016). Students' understanding of this concept is particularly reflected in their understanding of the equal sign; students need to have a relational understanding of the equal sign, which means that they perceive it as a symbol reflecting the sameness of the expressions on both sides (Knuth et al., 2006). A relational understanding of this symbol is positively related to performances in linear equation solving (Knuth et al., 2006). However, students often hold an operational view, meaning that they perceive the equal sign as symbol directing them to calculate an answer (e.g., Behr et al., 1980; Carpenter et al., 2003; Falkner et al., 1999).

When solving a system of linear equations with multiple unknowns, relationships between unknowns and numbers need to be determined, making sure that all equations within the system are satisfied. This requires understanding of the relationship between the unknowns and their pattern of covariation, that is, how changes in the value of one unknown must result in changes in the value of the other unknown in order for their sum to remain the same (Thompson & Carlson, 2017). Often-used strategies for solving a system of equations are substitution, replacing an expression by another expression of the same value, or elimination, subtracting equations from each other (Arcavi et al., 2016).

1.3 Fostering students' understanding of linear equations

Realistic Mathematics Education (RME) is a domain-specific instruction theory for mathematics (e.g., Treffers, 1978, 1987; see also Van den Heuvel-Panhuizen & Drijvers, 2020), which has its roots in Freudenthal's (1973) idea of mathematics as a human activity. Per this view, mathematics should be perceived as an activity of mathematizing, which means that mathematics can be seen as the activity of solving problems in our surrounding world, or more generally, as the activity of mathematically organizing and structuring reality (Freudenthal, 1968). Within RME, "realistic" situations, situations (or: contexts) of which students can imagine what is happening, serve as a source for developing understanding of mathematical concepts and procedures. Within these realistic contexts, students can solve problems with context-connected strategies, which later can gradually evolve into more formal solution processes (Van den Heuvel-Panhuizen & Drijvers, 2020). This process of progressive mathematization can be supported by making use of models: representations of mathematical problem situations in which the essential mathematics concepts and relevant aspects of the problem situation are reflected. Such models can bridge the gap between informal context-connected understanding on the one side and understanding of formal systems on the other (Van den Heuvel-Panhuizen, 2003). By first being a *model of* a particular situation in which the model has a close connection to a specific problem and later involving into a *model for* similar problem situations in other contexts, the model can be used to organize and solve new problem situations (Streefland, 1985, 2003). Allowing students to play an active role in their learning process, letting them explore the model and invent (informal) mathematical strategies, can moreover support their development in understanding (e.g., Van den Heuvel-Panhuizen & Drijvers, 2020).

Models that can be used to teach students linear equation solving include area models, linear models, and balance models. In area models (e.g., Filloy & Rojano, 1989; Van Amerom, 2002), the areas of rectangular shapes are used to represent the elements in equations. In the case of linear models, the unknown quantities are represented by the length of a line or strip (e.g., Dickinson & Eade, 2004; Warren & Cooper, 2009). Balance models (e.g., Van Amerom, 2002; Vlassis, 2002) are based on the idea of having a scale with equal weights on both sides. The balance model has a long didactical history. The philosopher and mathematician Gottfried Wilhelm Leibniz (1646-1716) already referred to the correspondence between equality in a mathematical situation and a balance with equal things on both sides (Leibniz, 1989). The balance model can be used for demonstrating the idea of equality in an equation (e.g., Taylor-Cox, 2003), an idea which is often enforced by referring to the metaphor of two sides being "in balance" (Antle et al., 2013). This model is therefore also deemed suitable for enhancing understanding of the equal sign as a symbol for representing equality (e.g., Pirie & Martin, 1997). The way in which balance models are represented varies, ranging from physical models (or: concrete models) being very close to real-world balance scales on which students' can perform operations, to abstract representations (e.g., Warren & Cooper, 2005).

We all have some sort of intuitive understanding of what "balance" means. Our idea of balance is understood through a set of closely related everyday-life experiences, such as walking without falling, standing up and sitting down, holding objects of similar and different weights, and so on (Alessandroni, 2018). In other words, we primarily understand balance "with our bodies and not by grasping a set of rules or concepts" (Johnson, 1987, p. 74). The idea that higher-order cognitive processes such as mathematical thinking processes, are influenced and shaped by our body and its interaction with the environment, is more broadly reflected in theories of embodied cognition (e.g., Barsalou, 2008; Gallese & Lakoff, 2005; Lakoff & Johnson, 1980; Núñez et al., 1999; Wilson, 2002). Conceptual metaphor theory (e.g., Lakoff & Johnson, 1980) provides an explanation for how mathematical reasoning and the understanding of abstract mathematical concepts can develop through physical experiences that we acquire by constant interaction with our environment. According to this theory, we form so called *image schemas*, knowledge structures that are derived from repeated perceptual-motor experiences within our surrounding world (Johnson, 1987), which can be used to give meaning to more abstract domains (Gibbs Jr, 2006). The linking between the bodily based experience of being in balance and the concept of equality in an equation can be explained by looking into conceptual metaphor theory: repeated bodily experiences related to maintaining balance together give rise to the "balance" image schema (Gibbs Jr, 2006; Johnson, 1987), which helps to understand the abstract idea of equality in an equation (Antle et al., 2013).

The idea that physical experiences are beneficial for learning mathematics has implications for teaching. Embodied learning environments (Duijzer et al., 2019) are environments in which students' physical experiences are a vital part of the learning activities. In this way, students are provided opportunities to ground abstract (mathematical) concepts in concrete bodily experiences (e.g., Abrahamson, 2017; Glenberg, 2010). In the case of teaching linear equation solving, physical experiences related to maintaining balance, for example by means of a physical balance model, might thus be beneficial for understanding equality and the equal sign in a linear equation.

2. The PhD thesis

The research described in this PhD thesis is part of the *Beyond Flatland* project. This research project was carried out with a grant (405-14-303) from the Netherlands Initiative for Education Research (NRO). The aim of this research project was to explore the possibilities for enriching the Dutch “flat” primary school mathematics curriculum as an approach to stimulate primary school students’ HOT. In this way, this research project is in line with the recently launched recommendations of the Dutch Association for the Development of Mathematics Education (Nederlandse Vereniging voor de Ontwikkeling van Reken-WiskundeOnderwijs [NVORWO], 2019; 2020) to support students’ development of mathematical HOT within the primary school classroom, and with the current ideas for the revision of the Dutch curriculum for primary and secondary education (curriculum.nu, 2019).

The *Beyond Flatland* project consisted of three part-projects, which each addressed another mathematical content domain: graphing motion, probability, and early algebra. This PhD thesis is the result of the part-project on early algebra. The main goal of this project was to gain insight in how primary school students’ algebraic reasoning can be stimulated, and in how this can create opportunities for fostering students’ HOT in the mathematics classroom. In this project, we concentrated on students’ reasoning related to linear equations. In order to stimulate fifth-grade students’ algebraic reasoning, a six-lesson teaching sequence about linear equations was developed in which a balance model played a key role. During the lessons, students’ explorations were taken as a basis for developing their reasoning related to solving (systems of) linear equations. The empirical data collected in this research project comprised students’ reasoning on lesson-specific assessment tasks administered at the end of each lesson (i.e., data about students’ reasoning on micro-level) and students’ reasoning on assessment tasks administered four times over the school year (i.e., data about students’ reasoning on macro-level).

Overall, in this PhD project we aimed to investigate (1) the role of the balance model in teaching linear equation solving; (2) the potential of using (various representations of) the balance model for supporting primary school students' algebraic understanding of linear equations; and (3) whether stimulating primary school students' algebraic reasoning related to solving linear equations also offers opportunities to promote students' reasoning in a related mathematical domain: graphing motion.

3. Structure of the PhD thesis

This PhD thesis comprises a number of journal articles, formatted as chapters, each addressing a different part of this research project on stimulating primary school students' algebraic reasoning related to linear equation solving as an approach to incorporate HOT in the mathematics classroom. Table 1 illustrates the structure of this thesis and provides for each chapter the title and the research question.

Table 1
Structure of the PhD thesis

Chapter	Title	Research question(s)
1	Introduction	
2	The balance model for teaching linear equations: A systematic literature review	What role does the balance model play in studies on teaching linear equation solving?
3	Developing algebraic reasoning in primary school using a hanging mobile as a learning supportive tool	How does fifth-grade students' algebraic reasoning for solving linear equations develop in an embodied learning environment? How are students' experiences in the learning environment related to their use of algebraic strategies?
4	Are physical experiences with the balance model beneficial for students' algebraic reasoning? An evaluation of two learning environments for linear equations	What is the effect of an intervention program based on the balance model on students' development of algebraic reasoning? How does the representation of balance model influence the development of students' algebraic reasoning?
5	Fifth-grade students' reasoning on linear equations and graphs of motion	To what extent does a teaching sequence on linear equations affect the development of students' algebraic and graphical reasoning?
6	Summary and discussion	

A systematic literature review was conducted in order to learn more about the often-used balance model as an aid for teaching linear equation solving. This review is described in *Chapter 2*. The purpose of this review was to gain more insight on *why* the balance model is used for teaching linear equation solving, *what* types of models are used, and *when* this model is used. In addition, we were interested in students' learning outcomes when teaching linear equation solving by means of this model. This chapter presents an overview of the role of the balance model in teaching linear equation solving, which can be a source for teachers, researchers, and developers of instructional materials for making informed instructional decisions about choosing this model for teaching linear equation solving.

Chapter 3 reports on the development of fifth-grade students' algebraic reasoning during an intervention consisting of a six-lesson teaching sequence about linear equations. The aim of this teaching sequence was to stimulate students' algebraic reasoning by providing them a learning environment in which they were able to invent, in an informal way, strategies for solving linear equations. A physical balance model, in the form of a *hanging mobile*, played a key role within this learning environment. Students' perceptual-motor experiences with this physical balance model were expected to be beneficial for the development of their algebraic reasoning. To examine students' reasoning, videos of classroom interactions, students' written work during the lessons, and students' responses to lesson-specific assessment tasks were analyzed. In this way, we could investigate the development of students' reasoning and their writing down of this reasoning over the lessons, and the relationship between students' experiences in the learning environment and their use of algebraic strategies such as restructuring, isolation, and substitution.

In a quasi-experimental study, described in *Chapter 4*, we further examined the effect of our six-lesson teaching sequence with the balance model on the development of students' HOT, operationalized as their algebraic reasoning related to solving systems of informal linear equations. Students of three classes participated in the teaching sequence with only pictorial representations of the balance model, students of three classes participated in the teaching sequence with physical and pictorial representations of the balance model, and students of the three classes participating in the control condition did not receive any instruction on linear equation solving. A staged comparison design was used: students' algebraic reasoning was repeatedly assessed four times over the school year and the teaching sequence was provided to the students in between two of these measurements in three successive cohorts, one class per cohort for each condition. In this study, we investigated the effect of participating in the intervention with the balance model on the development of students' algebraic reasoning over the school year. Moreover, we were particularly interested in whether different representations of the balance model had a differential effect on the development of students' reasoning.

Chapter 5 reports on a study in which we investigated whether our teaching sequence stimulating students' domain-specific mathematical HOT related to linear equation solving also offers opportunities to promote HOT within the related mathematical domain of graphing motion. HOT within both these mathematical domains is characterized by reasoning about covariation. As described in *Chapter 4*, students' algebraic reasoning was assessed four times over the school year, while students in between two of these measurements participated in the teaching sequence. At these same measurement moments students' graphical reasoning related to interpreting and constructing graphs of motion was examined as well. This allowed us to investigate the effect of our teaching sequence on students' reasoning in this other mathematical domain. With this study we aimed at knowing more about whether and how mathematical HOT can be stimulated across mathematics domains.

Finally, in *Chapter 6*, the findings from all four studies belonging to this PhD project are summarized and related to each other. Implications of the findings are discussed, limitations of this thesis are addressed, and suggestions for further research are proposed.

References

Abrahamson, D. (2017). Embodiment and mathematics learning. In K. Peppler (Ed.), T*he SAGE encyclopedia of out-of-school learning* (pp. 247–252). SAGE Publications.

Alessandroni, N. (2018). Varieties of embodiment in cognitive science. *Theory & Psychology, 28*(2)*,* 227–248. https://doi.org/10.1177/0959354317745589

Alexander, P. A., Dinsmore, D. L., Fox, E., Grossnickle, E. M., Loughlin, S. M., Maggioni, L., Parkinson, M. M., & Winters, F. I. (2011). Higher order thinking and knowledge: Domain-general and domain-specific trends and future directions. In G. Schraw, & D. R. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 47–88). Information Age.

Antle, A. N., Corness, G., & Bevans, A. (2013). Balancing justice: Comparing whole body and controller-based interaction for an abstract domain. *International Journal of Arts and Technology, 6*(4)*,* 388–409. https://doi.org/10.1504/IJART.2013.058285

Arcavi, A., Drijvers, P., & Stacey, K. (2016). *The learning and teaching of algebra: Ideas, insights and activities*. Routledge.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Behr, M., Erlwanger, S., & Nichols, E. (1980). How children view the equals sign. *Mathematics Teaching, 92*(1)*,* 13–15.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Springer.

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J. S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education*, *46*(1), 39–87. https://doi.org/10.5951/jresematheduc.46.1.0039

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: Handbook I. Cognitive domain.* Addison Wesley.

Brizuela, B., & Schliemann, A. (2004). Ten-year-old students solving linear equations. *For the Learning of Mathematics*, *24*(2), 33–40.

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school.* Heineman.

Carraher, D., & Schliemann, A. D. (2014). Early algebra teaching and learning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 193–196). Springer.

Carraher, D. W., Schliemann, A. D., & Schwartz, J. (2008). Early algebra is not the same as algebra early. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 235–272). Lawrence Erlbaum Associates.

Chimoni, M., Pitta-Pantazi, D., & Christou, C. (2018). Examining early algebraic thinking: insights from empirical data. *Educational Studies in Mathematics*, *98*(1), 57–76. https://doi.org/10.1007/s10649-018-9803-x

Common Core State Standards State Initiative [CCSSI] (2010). *Preparing America's students for college and career*. National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Curriculum.nu (2019, October). Leergebied rekenen & wiskunde [Learning domain mathematics]. Retrieved from https://www.curriculum.nu/voorstellen/rekenen-wiskunde/

Dede, C. (2010). Comparing Frameworks for 21st Century Skills. In J. Bellanca & R. Brandt (Eds.), *21st Century Skills* (pp. 51–76). Solution Tree Press.

Dickinson, P., & Eade, F. (2004). Using the number line to investigate the solving of linear equations. *For the Learning of Mathematics*, *24*(2), 41–47.

Duijzer, C., Van den Heuvel-Panhuizen, M., Veldhuis, M., Doorman, M., & Leseman, P. (2019). Embodied learning environments for graphing motion: A systematic literature review. *Educational Psychology Review, 31*(3), 597–629. https://doi.org/10.1007/s10648-019-09471-7

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 31–83). Cambridge University Press.

Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children's understanding of equality: A foundation for algebra. *Teaching Children Mathematics, 6*(4), 232–236.

Filloy, E., Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics, 9*(2), 19–25.

Forster, M. (2004). Higher order thinking skills. *Research Developments, 11*(1), 10–15.

Freudenthal, H. (1968). Why to teach mathematics so as to be useful? *Educational Studies in Mathematics, 1*(1–2), 3–8. https://doi.org/10.1007/BF00426224

Freudenthal, H. (1973). *Mathematics as an educational task*. Reidel Publishing Company.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3–4), 455–479. https://doi.org/10.1080/02643290442000310

Gibbs Jr, R. W. (2006). *Embodiment and cognitive science*. Cambridge University Press.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 586–596. https://doi.org/10.1002/wcs.55.

Goldenberg, P. E., Mark, J., Kang, J., Fries, M., Carter, C. J., & Cordner, T. (2015). *Making sense of algebra: Developing students' mathematical habits of mind.* Heinemann.

Goldenberg, E. P., Shteingold, N., & Feurzeig, N. (2003). Mathematical habits of mind for young children. In F.K. Lester & R.I. Charles (Eds.), *Teaching mathematics through problem solving: Prekindergarten-grade* (pp. 15–29). National Council of Teachers of Mathematics.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* University of Chicago Press.

Kaput, J. J. (2008). What is algebra? What is algebraic reasoning? In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 5–18). Lawrence Erlbaum Associates.

Kaput, J. J., Carraher, D. W., & Blanton, M. L. (2008). *Algebra in the early grades.* Lawrence Erlbaum Associates.

Katz, V. J. (Ed.). (2007). *Algebra: Gateway to a technological future.* Mathematical Association of America.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Kieran, C., Pang, J., Schifter, D., & Ng, S. F. (2016). *Early algebra: Research into its nature, its learning, its teaching.* Springer (open access eBook). https://doi.org/10.1007/978-3-319-32258-2

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4)*,* 297–312. https://www.jstor.org/stable/30034852

Kolovou, A., Van Den Heuvel-Panhuizen, M., & Bakker, A. (2009). Non-routine problem solving tasks in primary school mathematics textbooks–A needle in a haystack. *Mediterranean Journal for Research in Mathematics Education, 8*(2), 31–68.

Kolovou, A., Van den Heuvel-Panhuizen, M., & Köller, O. (2013). An intervention including an online game to improve Grade 6 students' performance in early algebra. *Journal for Research in Mathematics Education, 44*(3), 510–549. https://doi.org/10.5951/jresematheduc.44.3.0510

Lakoff, G., & Johnson, M. (1980). Conceptual metaphor in everyday language. *The Journal of Philosophy, 77*(8), 453–486. https://doi.org/10.2307/2025464

Leibniz, G. W. (1989). To queen Sophie Charlotte of Prussia, on what is independent of sense and matter (1702). In R. Ariew & D. Garber (Eds., Trans.), *Philosophical essays* (pp. 186–192). Hackett Publishing Company.

Leighton, J.P. (2004). Defining and describing reason. In J. Leighton & R. Sternberg (Eds.), *The nature of reasoning* (pp. 3–11). Cambridge University Press.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice*, *32*(3), 131–137. https://doi.org/10.1080/00405849309543588

National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics.* NCTM.

Nederlandse Vereniging voor de Ontwikkeling van Reken-Wiskunde Onderwijs [NVORWO] (2019, March). *Welke reken-wiskundige bagage heb je nodig om volwaardig te kunnen participeren in de maatschappij?* [Which mathematics equipment is necessary in order to fully participate in the society?]. Retrieved from https://www.nvorwo.nl/welke-reken-wiskundige-bagage-heb-je-nodig-om-volwaardig-te-kunnen-participeren-in-de-maatschappij/

Nederlandse Vereniging voor de Ontwikkeling van Reken-Wiskunde Onderwijs [NVORWO] (2020, January). *Position paper – Reactie NVORWO op eindproduct Curriculum.nu* [Position paper – Reply NVORWO on final product Curriculum.nu]. Retrieved from https://www.nvorwo.nl/proposition-paper-reactie-nvorwo-op-eindproduct-curriculum-nu/

Newmann, F. (1991). Higher order thinking in the teaching of social studies: Connections between theory and practice. In D. Perkins, J. Segal, & J. Voss (Eds.), *Informal reasoning and education*. Lawrence Erlbaum Associates

Núñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics, 39*(1–3), 45–65. https://doi.org/10.1023/A:1003759711966.

Organization for Economic Co-operation and Development [OECD] (2003). *The PISA 2003 assessment framework – Mathematics, reading, science and problem solving knowledge and skills*. OECD.

Pirie, S. E., & Martin, L. (1997). The equation, the whole equation and nothing but the equation! One approach to the teaching of linear equations. *Educational Studies in Mathematics*, *34*(2), 159–181. https://doi.org/10.1023/A:1003051829991

Resnick, L. B. (1987). *Education and learning to think*. National Academies Press.

Schoenfeld, A. H. (1995). Report of Working Group 1. In C. B. Lacampagne, W. Blair, & J. Kaput (Eds.), *The algebra initiative colloquium* (Vol. 2, pp.11–18). U.S. DOE, OERI.

Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, *10*(1), 9–34.

Streefland, L. (1985). Wiskunde als activiteit en de realiteit als bron [Mathematics as an activity and the reality as a source]. *Tijdschrift voor Nederlands Wiskundeonderwijs (Nieuwe Wiskrant)*, *5*(1), 60–67. Retrieved from http://www.fisme.science.uu.nl/wiskrant/artikelen/artikelen00-10/051/0501september_streefland.pdf

Streefland, L. (2003). Learning from history for teaching in the future. *Educational Studies in Mathematics, 54*(1)*,* 37–62. https://doi.org/10.1023/B:EDUC.0000005236.36499.fc

Taylor-Cox, J. (2003). Algebra in the early years. *Young Children, 58*(1), 14–21.

Thomas, A., & Thorne, G. (2009). *How to increase higher-order thinking*. Center for Development and Learning. Retrieved from http://www.thekeytotorah.com/uploads/2/5/5/8/25587179/how_to_increase_higher_order_thinking0001.pdf

Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Handbook of research in mathematics education* (pp. 421–456). National Council of Teachers of Mathematics.

Treffers, A. (1978). *Wiskobas doelgericht* [Wiskobas goal-directed]. Instituut voor Ontwikkeling van het Wiskunde Onderwijs.

Treffers, A (1987). *Three dimensions. A model of goal and theory description in mathematics instruction – The Wiskobas project*. Reidel Publishing Company.

Van Amerom, B. A. (2002). *Reinvention early algebra. Developmental research on the transition from arithmetic to algebra*. CD-ß Press / Freudenthal Institute.

Van den Heuvel-Panhuizen, M. (2003). The didactical use of models in realistic mathematics education: An example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics, 54*(1), 9–35. https://doi.org/10.1023/B:EDUC.0000005212.03219.dc.

Van den Heuvel-Panhuizen, M., & Bodin-Baarends, C. (2004). All or nothing: Problem solving by high achievers in mathematics. *Journal of the Korea Society of Mathematical Education*, *8*(3), 115–121.

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2020). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 713–717). Springer. https://doi.org/10.1007/978-3-030-15789-0_170

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

Van Reeuwijk, M. (1995, April). The role of realistic situations in developing tools for solving systems of equations. *Paper presented at the annual meeting of the American Educational Research Association,* San Francisco, CA.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics, 49*(3)*,* 341–359. https://doi.org/10.1023/A:1020229023965

Vogel, C. (2008). Algebra: Changing the equation. *District Administration*, *44*(6), 34–38.

Warren, E., & Cooper, T. J. (2005). Young children's ability to use the balance strategy to solve for unknowns. *Mathematics Education Research Journal, 17*(1), 58–72. https://doi.org/10.1007/BF03217409

Warren, E., & Cooper, T. J. (2009). Developing mathematics understanding and abstraction: The case of equivalence in the elementary years. *Mathematics Education Research Journal, 21*(2)*,* 76–95. https://doi.org/10.1007/BF03217546

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/BF03196322

CHAPTER 2

The balance model for teaching linear equations: A systematic literature review

Mara Otten, Marja van den Heuvel-Panhuizen,
and Michiel Veldhuis

Published in *International Journal of STEM Education, 2019, 6*(1), 30–51.

The balance model for teaching linear equations: A systematic literature review

Abstract

This paper reports a systematic literature review of the balance model, an often-used aid to teach linear equations. The purpose of the review was to report *why* such a model is used, *what* types of models are used, and *when* they are used. In total 34 peer-reviewed journal articles were analyzed, resulting in a comprehensive overview of described rationales for using the balance model, its appearances, situations in which it was used, and the gained learning outcomes. Some trends appeared about how rationales, appearances, situations, and learning outcomes are related. However, a clear pattern could not be identified. Our study shows that this seemingly simple model actually is a rather complex didactic tool of which in-depth knowledge is lacking. Further systematic research is needed for making informed instructional decisions on when and how balance models can be used effectively for teaching linear equation solving.

Keywords: Algebra, Teaching linear equations, Balance model.

1. Introduction

A substantial component of learning algebra is learning to solve algebraic equations. Within the algebra curriculum, solving linear equations is one of the foundational topics in which students make the transition from reasoning with numbers to reasoning with unknowns (e.g., Filloy & Rojano, 1989). Similarly, early algebra has been described as a "shift from thinking about relations among particular numbers and measures towards thinking about relations among sets of numbers and measures, from computing numerical answers to describing and representing relations among variables" (Carraher et al., 2008, p. 266). Solving linear equations as a basic skill (Ballheim, 1999) is a substantial part of the middle school mathematics program (Huntley & Terrell, 2014). However, many students do not achieve mastery of this basic skill and experience difficulties in learning the concepts and skills related to solving equations (e.g., Kieran, 2007).

Solving linear equations means that the values of the unknown quantities have to be found based on the equality of two given mathematical expressions — one on each side of the equal sign. The essence of an equation is that these mathematical expressions represent the same value (Alibali, 1999), which makes *equality* a key concept in solving linear equations (e.g., Bush & Karp, 2013) and understanding equality one of the main conceptual demands associated with equation solving (Kieran, 1997; Kieran et al., 2016). Students need to understand that in an equation the expressions on both sides of the equal sign have the same value and that this equality should always be maintained in the process of solving an equation (e.g., Kieran et al., 2016).

Misconceptions related to the concept of equality in linear equation solving are well documented. These misconceptions are particularly reflected in students' interpretations of the equal sign. Instead of perceiving it as a relational symbol meaning "is the same as", students often have an operational view of the equal sign, that is, they view it as a sign to "do something" or to "calculate the answer" (e.g., Knuth et al., 2006). For example, when solving the problem $8 + 4 = \_\_ + 5$, a common mistake is adding the numbers on the left side of the equation and putting a 12 in the blank (Falkner et al., 1999). Such interpretation of the equal sign can begin in the elementary grades and can persist through middle school (e.g., Alibali et al., 2007).

One way to help students gain conceptual understanding in equation solving is through the use of models as "ways of thinking about abstract concepts" (Warren & Cooper, 2009, p. 77). Such didactical models can be seen as representations of mathematical problem situations in which the essential mathematical concepts and aspects of the problem situation are reflected, and through which the concrete situation is connected to the more formal mathematics (Van den Heuvel-Panhuizen, 2003). By first being a *model of* a concrete situation where the model is very closely related to a specific problem and later evolving to a *model for* similar problems that are situated in another context, the model can be applied for solving all kinds of new problems (Streefland, 2003).

In mathematics education, several didactical models are used to give students access to particular mathematical concepts, such as the number line or the bar model for teaching fractions. The balance model is another commonly used didactical model. This model is often used to support students' understanding of linear equation solving. Characteristic of the balance model is that its form serves as a model for its function in solving linear equations: the balance can be used to refer to the situation of equality of the expressions on the two sides of an equation. The philosopher and mathematician Gottfried Wilhelm Leibniz (1646-1716) already made this connection when he mentioned the relation between equality in a mathematical situation and a balance with equal things on both sides (Leibniz, 1989).

In the context of a larger research project on algebraic reasoning, we wanted to find indications for setting up a teaching sequence on linear equation solving. We searched for information about the use of the balance model as a possible aid to assist students in developing understanding of solving linear equations. The diverse and scattered picture we got from this initial search prompted us to investigate this more systematically. Therefore, we planned to carry out a systematic review of the literature of how the balance model turns up in the large body of research and professional articles that has been published about teaching linear equation solving. With this review, we aimed at answering the following research question: What role does the balance model play in studies on teaching linear equation solving?

In general, to learn more about a didactical model for teaching students (mathematical) concepts, it is essential to gain insight in various important aspects of a model. The specific way of representing the model is important to take into account, but also information related to possible reasons for choosing this particular model and timing of using the model in a teaching and learning trajectory contribute to getting a complete picture of how the model can be used. Lastly, to be able to evaluate the use of a didactical model for fostering students' conceptual understanding, it is important to incorporate students' learning outcomes as well. To determine the role balance models play in studies on teaching linear equation solving, we looked into what authors reported about *why* such a model is used, *what* types of models are used, *when* they are used, and what learning outcomes are associated with its use. Knowing this can be helpful for teachers, researchers, and developers of instructional material for making informed decisions about choosing the balance model for teaching linear equation solving.

2. Method

2.1 Article search and selection

For selecting articles for the review, we searched in 93 peer-reviewed research journals in the areas of mathematics education, educational sciences, pedagogics, developmental psychology, special education, and technology in education. The search was conducted in Scopus and ERIC for articles in English. The search query consisted of terms such as equation*, equal sign*, equality, equivalence, balanc*, algebra*, mathematic*, unknown*, and solv*, and combinations thereof (see Appendix 2.1 for the complete search queries). There was no limit with regard to the date of publication.

The search, conducted in March 2017, resulted in 932 hits in Scopus and in 723 hits in ERIC, together resulting in 1655 hits (see Figure 1). After merging, 333 duplicates were identified and removed, resulting in 1322 articles from 92 journals. Thirty-two articles were removed either because they, despite our search query, turned out to be not in English, or because they did not originate from peer-reviewed journals (e.g., were book chapters), resulting in 1290 articles.

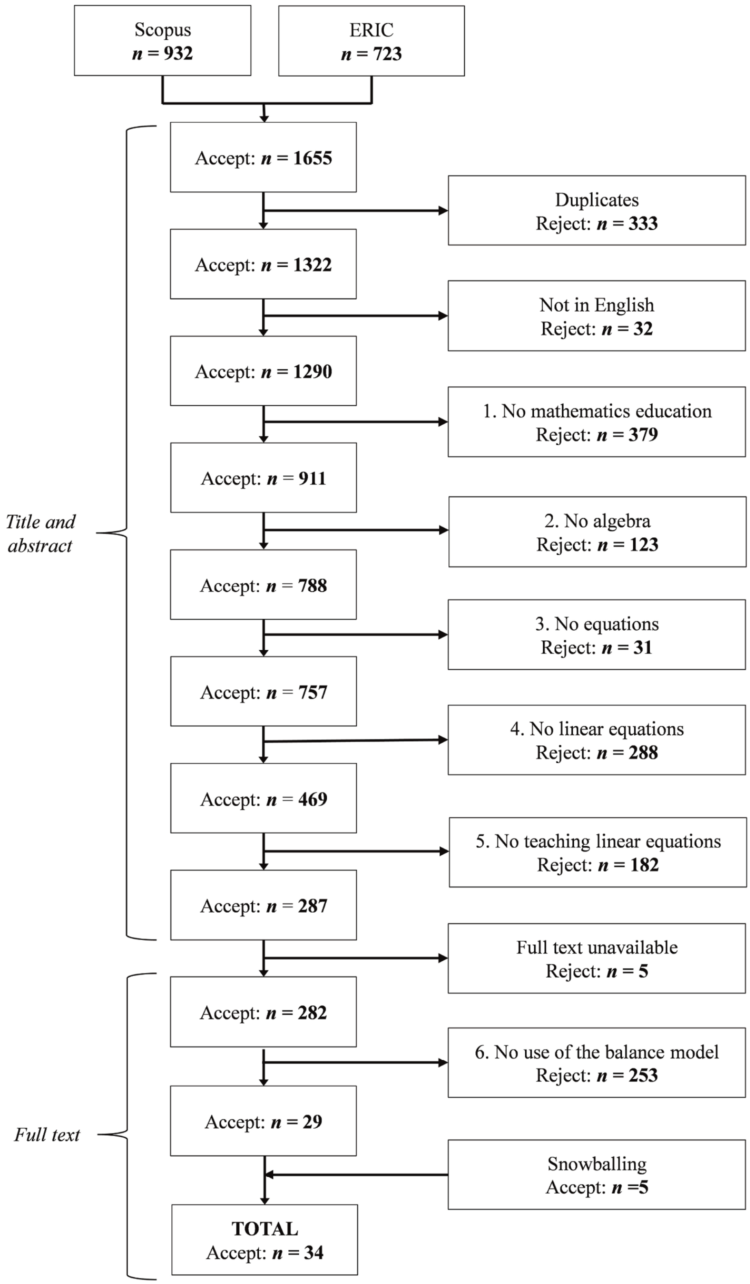


Figure 1. Flow chart illustrating the systematic search process, resulting in 34 articles.

In a six-step procedure, titles and abstracts were screened. Articles that were not in the field of mathematics education, did not touch upon the domain of algebra, did not address equations, were not about linear equations, or did not address teaching or learning linear equations, were excluded. This resulted in 287 articles. In the sixth and final step, the 282 articles of which we could obtain the full text were inspected to make an accurate decision on whether the concept of balance was discussed in relation to linear equation solving. Based on this reading, 29 articles were selected in which the balance model was used to teach linear equation solving. Lastly, snowballing was used to ensure that possible other relevant literature was covered as well, which resulted in five additional articles. Thus, the final collection comprised 34 articles from 22 journals.

2.2 Data extraction

For each of the 34 articles, information was extracted related to the rationales and the limitations of using the balance model, the appearances of the model, the situations in which the model was used, and students' learning outcomes. Information was extracted in case at least one sentence of the article was devoted to either of these four categories. After the inventory of all rationales (in 26 articles) and appearances (in 34 articles), patterns were identified to see whether classes of rationales and types of models could be created. Multiple rationales for using the model and multiple appearances could be extracted from one article. To describe the situations in which the balance model was used (in 34 articles), information was extracted regarding the grade level of the students, the duration of the intervention, the type of tasks students worked on, and the type of instruction. Students' learning outcomes when using a balance model for teaching linear equation solving were discussed in 19 articles. These different aspects of the reviewed articles are summarized in Table 1.

Table 1

Overview of articles in which the balance model (BM) was used for teaching linear equations

Article	Ratio-nale[b]	Appea-rance	Use: Duration inter-vention	Use: Instructional setting	Use: Type of equations[c]	Students involved	Research design[d]	Intervention in comparison group (CG)?	Learning outcomes (on linear equation solving unless otherwise specified)
Alibali (1999)	PE	Drawn	40-min session	Individual instruction by teacher	$3 + 4 + 5 = \_\_ + 5$	Grades 3-5; 143 students	Pre-posttest; BM[a]-group and two comparison groups	CG1: received feedback CG2: explanation solution steps	• 36% of the BM-group improved • BM-group outperformed CG1 • CG2 outperformed BM-group
Andrews (2003)		Physical, drawn	1 lesson	Classroom instruction by teacher	$2x + 5 = x + 8$	Grade 7; 4 students			
Andrews & Sayers (2012)	EQ, LI	Drawn	5 lessons	Classroom instruction by teacher	$x + 7 = 9$ $x - 2 = 10$	Grade 8 3 classes			
Araya et al. (2010)	PE, MR	Drawn	2-h session	Classroom instruction by learning movie	$2x + 1 = 5 + x$	Grade 7; 236 students; no previous algebra instruction	Posttest; BM-group and comparison group	CG: symbolic instruction	• Below average to high achieving students of the BM-group outperformed the CG
Austin & Vollrath (1989)	PE, LI	Physical			$3w + 5 = 11$	"Introductory algebra students"			
Berks & Vlasnik (2014)	MR	Drawn	1 lesson	Classroom instruction by teacher	$4x + 2y = 12$ $y = 2x + 2$	Students with some algebra experience			

Article	Ratio-nale[b]	Appea-rance	Use			Students involved	Research design[d]	Intervention in comparison group (CG)?	Learning outcomes (on linear equation solving unless otherwise specified)
			Duration inter-vention	Instructional setting	Type of equations[c]				
Boulton-Lewis et al. (1997)		Drawn	5 lessons	Classroom instruction by teacher	$2x + 5 = 17$	Grade 8; 21 students	Pre-posttest; BM-group		• Few students could model or solve equations with the BM • Most students could solve equations without the BM • Students preferred not to use the BM
Brodie & Shalem (2011)	EQ	Drawn	3-5 lessons	Classroom instruction by teacher	$3 + x = 5$	Grade 8			
Caglayan & Olive (2010)	MR, LI	Drawn	2 lessons	Classroom instruction by teacher	$2c + 1 = 7$ $2x - 1 = 13$	Grade 8; 24 students	Descriptive; BM-group		• BM gives meaning to equations with addition/multiplication • BM does not give meaning to equations with negative values/subtraction
Cooper & Warren (2008)	MR	Physical; drawn	5 years	Classroom instruction by teacher	? + 11 = 36 ? - 7 = 6	Grades 2-6; 220-270 students	Descriptive; BM-group		• Young students can generalize the balance method[e] for simple equations • Older students can generalize the balance method for all operations and use it to solve equations
Figueira-Sampaio et al. (2009)	EQ, PE, LI	Group 1: Physical Group 2: Virtual	50-min lesson	Group 1: Classroom instruction by teacher Group 2: Working in pairs with computer	$5x + 50 = 3x + 290$	Grade 6; 46 students	Descriptive; two BM-groups		• Virtual BM-group shows more participation, social interaction, motivation, cooperation, discussion, reflection, and a feeling of responsibility, than the physical BM-group

Article	Ratio-nale[b]	Appea-rance	Use: Duration inter-vention	Use: Instructional setting	Use: Type of equations[c]	Students involved	Research design[d]	Intervention in comparison group (CG)?	Learning outcomes (on linear equation solving unless otherwise specified)
Filloy & Rojano (1989)	EQ, MR, LI	Drawn	1 session with 5 problems	Individual instruction by teacher	$3 + 2x = 5x$ $10x - 18 = 4x$	Grade 7; Three classes	Descriptive; BM-group		• With BM, the step from solving equations with unknowns on one side of the equal sign towards solving equations with unknowns on both sides of the equal sign, is smaller than with the geometrical model • The geometrical model is more appropriate than the BM for modeling equations with subtraction • Assigning values to unknowns can hinder students when using the BM
Fyfe et al. (2015)	PE	Physical, drawn	1 lesson	Individual instruction by teacher	2 + 3 = 2 + __	Grades 1-3; 389 students			
Gavin & Sheffield (2015)	EQ	Drawn		Classroom instruction by teacher	$12 + 23 = 13 + n$ $51 - n = 50 - 25$	Grade 6; 305 students			
Jupri et al. (2014)		Drawn	1 item on a test		1kg + 0.5brick = 1brick	Grade 8; 51 students			
Kaplan & Alon (2013)	PE	Virtual	1 session	Individual instruction by teacher and individually working with computer	▲▲ = ●●●	Grades 3-4; 2 students			
Leavy et al. (2013)	EQ	Physical		Classroom instruction by teacher	8 = __ + 3	Grade 3			

Article	Ratio-nale[b]	Appea-rance	Use: Duration inter-vention	Use: Instructional setting	Use: Type of equations[c]	Students involved	Research design[d]	Intervention in comparison group (CG)?	Learning outcomes (on linear equation solving unless otherwise specified)
Linchevski & Herscovics (1996)	EQ, LI	Drawn	1 lesson	Individual instruction by teacher	$8n + 11 =$ $5n + 50$	Grade 7; 6 students	Descriptive; case studies with BM		• BM is suitable for demonstrating cancellation of identical terms on both sides of the equation • BM is not suitable for modeling equations with subtraction
Mann (2004)	EQ	Physical, drawn		Classroom instruction by teacher	●●■ = ■ ■ 5 + 6 = __ + 2	Grade 3; 1 class			
Marschall & Andrews (2015)	EQ, LI	Drawn		Classroom instruction by teacher	$x + 1 = 3$ $4x - 3 = 2x + 5$	Grade 6; 6 classes			
Ngu et al. (2015)		Drawn	40-min lesson	Individual instruction sheet with BM	$5 + 3n = 10$ $3m - 1 = 5$	Grade 8; 71 students	Pre-posttest; BM-group and comparison group	CG: solving equations with inverse operations	• BM-group improved from pre- to posttest • CG improved more than BM-group • Higher cognitive load for BM-group than CG
Ngu & Phan (2016)		Drawn	45-min lesson	Individual instruction sheet with BM	$n / 2 = 7$ $x - 9 = 4$	Grade 7; 63 students	Pre-posttest; BM-group and comparison group	CG: solving equations with inverse operations	• BM-group improved from pre- to posttest • CG improved more than BM-group • Positive relation between performance on procedural knowledge and performance on conceptual knowledge for CG but not for BM-group
Ngu et al. (2018)		Drawn	Two 40-min lessons	Individual instruction sheet with BM	$3x +1 = 2x + 8$ $6 - q = 10$	Grades 8-9; 29 students	Pre-posttest; BM-group and comparison group	CG: solving equations with inverse operations	• BM-group improved from pre- to posttest • CG improved more than BM-group • Higher cognitive load for BM-group than CG

Article	Ratio-nale[b]	Appea-rance	Use: Duration inter-vention	Use: Instructional setting	Use: Type of equations[c]	Students involved	Research design[d]	Intervention in comparison group (CG)?	Learning outcomes (on linear equation solving unless otherwise specified)
Orlov (1971)	PE	Physical	2 years	Classroom instruction by teacher	$5x - x + 2 = 2x + 6$	Grade 8; 200 students	Repeated measures; BM-group and comparison group	CG: experimental program without BM	• BM-group, especially average and above-average students, outperformed CG
Perry et al. (1995)	PE	Physical	1 lesson	Individual instruction by teacher	$3 + 4 + 5 = \_\_ + 5$	Grades 4-5; 56 students	Pre-posttest; BM-group and comparison group	CG: only verbal instruction	• BM-group outperformed CG
Raymond & Leinenbach (2000)		Drawn	26 lessons	Classroom instruction by teacher	$x + 4 = 2x + 3$	Grade 8; 120 students	Descriptive; BM-group		• BM instruction leads to better performance than textbook instruction • Large performance decrease when returning to textbook after BM • Better than expected performances on standardized algebra test after BM
Rystedt et al. (2016)	EQ	Drawn	1 lesson	Classroom instruction by teacher	$4x + 4 = 2x + 8$	Grades 6-7; Five classes			
Smith (1985)		Physical		In pairs with BM	$8w = 120$	Grades 4-6	Descriptive; BM-group		• BM assisted in exploring/learning basic algebraic principles and enhanced motivation
Suh & Moyer (2007)	PE, MR	Group 1: Virtual Group 2: Drawn	5 lessons	Classroom instruction by teacher; students individually with BM	$2x + 2 = 10$	Grade 3; 36 students	Pre-posttest; two BM-groups		• Both BM-groups improved • Each of the BMs showed unique features to support learning
Taylor-Cox (2003)	EQ, PE	Physical	1 lesson	Classroom instruction by teacher	$A + C + B = C + A + B$	Grade 1			

Article	Ratio-nale[b]	Appea-rance	Use			Students involved	Research design[d]	Intervention in comparison group (CG)?	Learning outcomes (on linear equation solving unless otherwise specified)
			Duration inter-vention	Instructional setting	Type of equations[c]				
Vlassis (2002)	EQ, LI	Drawn	16 lessons	Classroom instruction by teacher	$7x + 38 = 3x + 74$ $13x - 24 = 8x + 76$	Grade 8; 40 students	Descriptive; BM-group		• Balance method was used by all students • After BM instruction students made many mistakes related to negative numbers and unknowns
Warren & Cooper (2005)	EQ, MR	Physical; drawn	4 lessons	Classroom instruction by teacher	$? + 7 = 11$ $? - 4 = 13$	Grade 3; 20 students	Descriptive; BM-group		• Most students could represent equations with the BM and translate the model into symbolic equations • BM assisted students in understanding the equal sign and solving for unknowns • Ten students used the balance method for solving a subtraction problem; for others further teaching was necessary
Warren & Cooper (2009)	EQ, PE, MR	Physical; drawn	5 years	Classroom instruction by teacher	$? + 2 = 5$ $? - 3 = 6$	Grades 2-6; 220-270 students;	Descriptive; BM-group		• BM enhanced understanding of language and symbols • Students could generalize balance method for simple equations • Older students could generalize the balance method for all operations
Warren et al. (2009)	EQ, MR	Physical, drawn		Classroom instruction by teacher	$5 + 1 = 2 + 4$	Kinder-garten			

Note. Empty cells indicate that this information was not provided in that article.
[a] BM = Balance model
[b] EQ = Rationales related to the equality aspect, PE = Rationales related to the physical experiences, MR = Rationales related to learning through models and representations, LI = Limitation of using the balance model
[c] The column "type of equations" provides examples of the linear equations which were taught by means of the balance model used in that particular study, such as equations containing only positive numbers and addition, or equations containing also negative values and subtraction
[d] Information about the research design was only included for articles in which the effect of the balance model on students' learning outcomes was evaluated
[e] "Balance method" refers to the method of solving an equation by performing the same operations on both sides of the equation

2

3. Results

3.1 Why was the balance model used?

Rationales for using the balance model were provided in 26 articles. Three main classes of rationales could be identified, which were all related to the specific features of the context of the balance model. Articles constituting the *Equality* class of rationales all *directly* referred to using the balance to enhance students' understanding of the concept of equality. Direct references to equality are directly focused on the mathematical equality, by emphasizing the analogy between the balance model and equality in an equation. Articles in the remaining two classes of rationales made more *indirect* references to using the balance model for enhancing students' understanding of equality. Indirect references to equality are, for example, offering students physical experiences when manipulating a balance model and thus feel, through the experience of balancing, the concept of equality. Such articles that made a reference to previous or concurrent physical experiences related to the balance model fell in the *Physical Experiences* class of rationales. Articles that fell into the *Models and Representations* class of rationales referred to the use of models and representations for enhancing students' conceptual understanding in linear equation solving. Finally, limitations of using the balance model for teaching linear equation solving were also extracted.

3.1.1 Rationales related to the equality concept

A majority of 15 articles (three from the same research project) mentioned rationales for using the balance model related to the concept of equality. It was often stated that understanding the concept of equality can be enhanced by making use of the model of a balance (e.g., Gavin & Sheffield, 2015; Leavy et al., 2013; Mann, 2004; Taylor-Cox, 2003; Warren et al., 2009). Because both sides of a balance model are of equal value and thus exchangeable, the model was described as being very suitable for demonstrating the idea of equality or equilibrium (Figueira-Sampaio et al., 2009) and quantitative sameness (e.g., Warren & Cooper, 2005). In line with this, several authors referred to the use of the balance model to enhance the understanding of the equal sign as a symbol for representing equality (e.g., Vlassis, 2002; Warren & Cooper, 2009). Accordingly, the balance model has often been described as suitable to demonstrate the strategy of doing the same thing on both sides of the equation, in which emphasis on the concept of balance is crucial (e.g., Andrews & Sayers, 2012; Figueira-Sampaio et al., 2009; Marschall & Andrews, 2015), thereby helping students in forming, according to Vlassis (2002), a mental picture of the operations they have to apply. Another mentioned advantage of the balance model is the possibility to keep track "of the entire numerical relationship expressed by the equation while it is being subjected to transformations" (Linchevski & Herscovics, 1996, p. 52), which makes it suitable for demonstration of the cancelation of identical terms on both sides of the equation (e.g., Filloy & Rojano, 1989).

3.1.2 Rationales related to the physical experiences

The second class of rationales that was identified, mentioned in 11 articles (all from different research projects), was related to learning through physical experiences. In several articles, a reference was made to previous physical experiences related to maintaining balance. Araya et al. (2010) argued that the process of maintaining balance has a primary biological basis and is therefore common physical knowledge for all human beings. Through using the balance model, this biologically primary knowledge can be connected to the abstract idea of maintaining equality in an equation. Others emphasized the similarity between the model and a teeter-totter (or see-saw) and referred to children's (playing) experiences with this object (Alibali, 1999; Kaplan & Alon, 2013).

In other articles, the contribution of concurrent physical experiences with the balance model was pointed out as being beneficial to the learning of linear equations. Warren and Cooper (2009) underlined the importance of movement (for example by acting out a balance) and gestures during the learning trajectory to develop mental models of mathematical ideas. They argued that referring to these experiences in later stages of the learning process can be beneficial. Also, the importance of physical experiences with concrete objects for developing understanding of linear equations was mentioned in several articles. Offering young students experiences with manipulation of balance scales, because through this manipulation, equality can be recognized, defined, created, and maintained, could enhance students' understanding of this concept (Taylor-Cox, 2003). Suh and Moyer (2007) mentioned that using manipulable concrete objects have a sense-making function, through connecting procedural knowledge (manipulations on the objects) and conceptual knowledge of algebraic equations. However, at the same time these authors pointed out that caution with using such manipulatives for teaching formal equation solving is necessary, because not all students automatically connect their actions on manipulatives with their manipulations on abstract symbols. Also Orlov (1971) commented that the balance model as a physical instrument can help in forming abstract mathematical thought, because it represents an intermediate degree between immediate sensory data and mathematical abstraction. In this same line, Fyfe et al. (2015) recommended a sequence based on fading concreteness, where instruction begins with concrete material and fades into abstract mathematical symbols. The real-time feedback some models provide about being in balance, which allows students to verify the results of their manipulations and their reasoning processes and as such to construct knowledge, was also deemed important (Austin & Vollrath, 1989). When combined with social experiences, physical experiences were also said to contribute to the construction of knowledge (Figueira-Sampaio et al., 2009), for example, because it creates shared meaning between teacher and students (Perry et al.,1995).

3.1.3 Rationales related to learning through models and representations

The third class of rationales, mentioned in eight articles (four from the same research project), included a more general argumentation and referred to learning through the use of models and representations. According to Filloy and Rojano (1989), models such as the balance model can provide an opportunity to semantically and syntactically set a foundation for linear equation solving. Here, the meaning of equality and algebraic operations can first be derived from the context, while after students have gone through a process of abstraction, meaning at syntactic level is linked to this meaning of the context. Researchers involved in the Australian Early Algebraic Thinking Project (Cooper & Warren, 2008; Warren & Cooper, 2009) argued that, through models, mathematical ideas are presented externally as concrete material, by iconic representations, language, or symbols, while comprehension of these ideas occurs internally, in mental models or internal cognitive representations of mathematical ideas underlying the external representations. From this point of view, mathematical understanding is determined by the number and strength of the connections in the student's internal network of representations. Also the use of multiple representations in teaching abstract mathematical concepts or strategies was advocated (e.g., Berks & Vlasnik, 2014), because experiencing different modes of representation and making connections between and within these different modes of representation could enhance deep mathematical understanding (Suh & Moyer, 2007). The sense-making function of representations was elaborated on by Caglayan and Olive (2010), who reasoned that students can make sense of abstract symbolic equations through connecting this symbolic equation with the equation as expressed by its representation.

Also other reasons for using representations of the balance model were suggested. For example that it can create a shared language base which students can use when explaining their solutions (Berks & Vlasnik, 2014; Warren & Cooper, 2005; Warren et al., 2009) or that it is supposed to lower students' cognitive load during equation solving processes (Araya et al., 2010). The latter contrasts with Boulton-Lewis et al. (1997), who hypothesized an increased processing load caused by the use of concrete representations. This might depend on the students' experience and the type of equation problems they have to solve: if students do not really need the help of a concrete representation of the balance model anymore and they still have to use it, this could indeed increase processing load.

3.1.4 Limitations of the balance model

Limitations of the balance model were described in eight articles (all from different research projects). In her well-known article, Vlassis (2002) described how eight-grade students were taught formal linear equation solving by making use of the balance model and concluded that although the balance model was able to provide students an "operative mental image" (p. 355) of the to-be-applied equation solving strategies, this model also had some shortcomings. For example, the model was not helpful for equations containing negative numbers or for other equations that are "detached from the model" (p. 354) and that no longer refer to a concrete model. Also, several other articles referred to the restricted possibilities the model has to represent equations with negative quantities or subtractions (e.g., Filloy & Rojano, 1989; Linchevski & Herscovics, 1996). As soon as negative values are involved, such as in the case of the equation $x + 5 = 3$, or equations with subtraction, such as $2x - 3 = 5$, the solution is difficult to express in terms of physical weight which makes it difficult to construct meaning for these equations (Caglayan & Olive, 2010).

3.1.5 Discussion of the findings regarding why the model was used

Although the three classes of rationales all have unique characteristics based on which they can be differentiated, they are also interrelated. The most often mentioned rationale was related to equality; understanding equality is regarded as one of the main conceptual demands associated with linear equation solving (e.g., Kieran et al., 2016). Inherent properties of the balance were connected to the concept of equality and the strategies that can be applied while maintaining the balance. The two remaining rationales were less often mentioned. These rationales contained indirect references to using the balance model for enhancing students' understanding of equality in an equation, through referring to learning through physical experiences or to learning through models and representations.

Articles in the class of rationales related to physical experiences referred either to the biological basis of maintaining balance or to other physical experiences with balance (such as with a teeter-totter), which, through using the balance model, could be connected to the idea of maintaining the balance in an equation. These previous physical experiences with balance could foster students' understanding of equality in an equation. This can be explained from the perspective of embodied cognition theory, stating that the connection of perceptual and physical experiences that we have when we interact with the world is fundamental for developing conceptual knowledge and cognitive learning processes (e.g., Barsalou, 2008; Wilson, 2002). Perceptual-motor experiences are considered essential for developing mathematical concepts (e.g., Alibali & Nathan, 2012; Núñez et al., 1999), and mathematical reasoning is viewed as intricately linked with embodied actions (Abrahamson, 2017; Alibali & Nathan, 2012). When applying embodied cognition theory to teaching and learning solving linear equations, it is assumed that perceptual-motor knowledge about the action of balancing is a necessary foundation for developing understanding of the mathematical concept of equality (e.g., Antle et al., 2013). This perceptual-motor knowledge is built up from the very pervasive physical experiences we have with balancing from a young age on (Gibbs Jr, 2006), through walking without falling, standing up and sitting down, or holding objects of different weights (Alessandroni, 2018). Furthermore, the other articles in this class of rationales stressed the contribution of concurrent physical experiences with the balance model to the learning of linear equations. Through manipulation of the model, students explore how to maintain its balance; these strategies for maintaining the balance of the model could later be connected to strategies for maintaining equality in an equation. This is also in line with embodied cognition theory: offering students opportunities to revitalize the basic perceptual-motor knowledge through making use of a model of a balance through which they can imagine (or experience anew) the situation of balancing, could be beneficial for supporting students' understanding of equality in an equation and therefore for teaching linear equation solving.

Articles in the class of rationales related to learning through models and representations included more general arguments for enhancing students' understanding of equality in an equation. However, these rationales have some overlap with the rationales related to physical experiences. Both classes are related to perceptual-motor experiences with balance. In the case of the *Models and Representations* class, this experience is more related to what the balance looks like. The balance as a device with two arms and a fulcrum in the middle can be used to represent an equation with on two sides of the equal sign an expression of equal value. Learning through models and representations can be connected to ideas of

Realistic Mathematics Education (RME). One of the main instructional principles of RME is the use of didactical models with the purpose to bridge the gap between informal, context-related solution methods and the more formal ones, and in this way, to stimulate students to come to higher levels of understanding (e.g., Van den Heuvel-Panhuizen, 2003).

3.2 What types of balance models were used?

Three types of appearances of the balance models came forward in the reviewed articles: physical, virtual, and drawn balance models. Physical balance models are concrete balance scales. On these scales, students can represent equations by placing real objects, standing for knowns and unknowns, on both sides of the model. Characteristic for these models is that they are dynamic, which means the students can operate on them and get real-time feedback on their actions. In virtual balance models, the balance is implemented in a digital environment. These models are mostly dynamic, in that sense that the balance tilts in response to students' (digital) manipulations and in this way gives real-time feedback. In drawn balance models, a schematic version of a balance is presented on paper or on the blackboard. The representations of these balance models are static: students cannot manipulate them and cannot receive real-time feedback. Whereas in most articles only one type of appearance of the balance model was used, in other articles different types appeared (e.g., Figueira-Sampaio et al., 2009) or a sequence of different appearances was presented, starting with the use of a physical model followed by a drawn balance model (e.g., Fyfe et al., 2015).

3.2.1 Physical balance models

Physical balance models appeared in 14 articles (three from the same research project). We drew schematic versions of several of these physical balance models. These drawings are shown in Figure 2. The balance displayed in Figure 2a was used by Fyfe et al. (2015) to represent, for example, $3 + 2 = 1 + 1 + \_\_$. Here, students could put three red and two yellow bears on the left side and one red and one yellow on the right, and then add the missing number to get the scales to balance (for similar models, see e.g., Warren et al., 2009). In Austin and Vollrath's balance model (1989; Figure 2b), the equation $3x + 5 = 11$ is portrayed by, on the left side, three containers with unknown content and five washers and eleven washers on the right side (for similar models, see e.g., Andrews, 2003). A more complex example of a balance scale was utilized by Orlov (1971; Figure 2c). His model contains four scales, two on each side. For example, by putting a weight on the left tray of the left part of the scale, the left arm of the balance scale goes up. In this way, negative numbers and unknowns can also be handled by this model. The last type of described physical balance model is a balance model in which the distance of the objects to the fulcrum can be adapted to represent linear equations such as 8 + 4 + 2 = 4 + 4 + __ (Perry et al., 1995; Figure 2d; for a similar model, see Smith, 1985). Here all objects have the same weight, but by putting them at a particular position on the beam they represent a particular value.

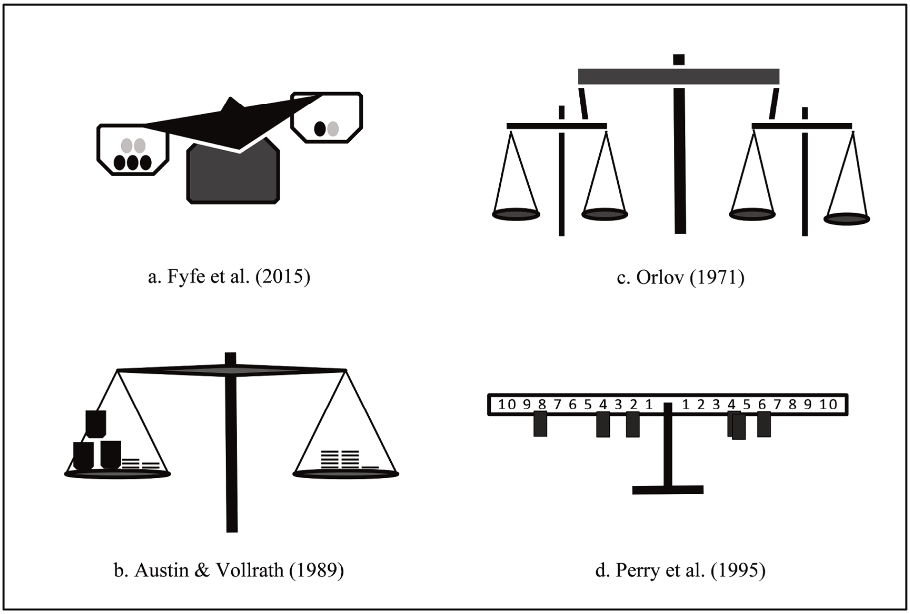


Figure 2. Physical balance models, examples from four articles (a-d).

3.2.2 Virtual balance models

Virtual balance models appeared in three articles (from different research projects). Drawings of the used virtual balance models are shown in Figure 3. Most of these models display a balance scale quite similar to the physical balance models. However, the digital environment enables more possibilities in representations and functions of the model.

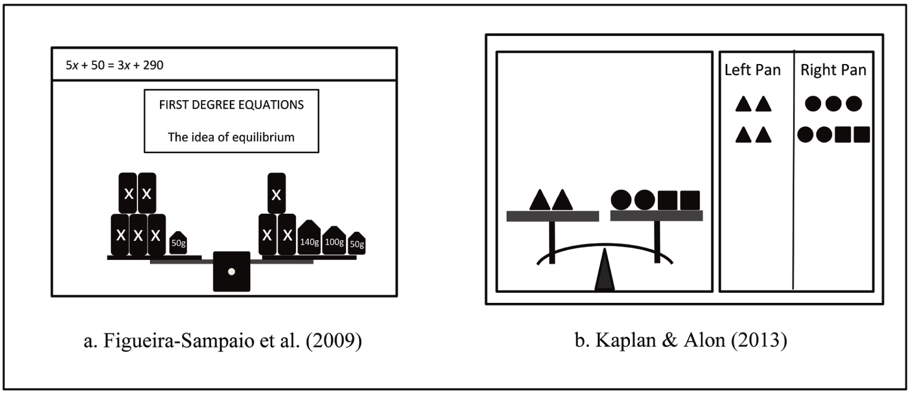


Figure 3. Virtual balance models, examples from two articles (a-b).

In the digital model used by Figueira-Sampaio et al. (2009; Figure 3a), the equation $5x + 50 = 3x + 290$ is represented by cans with the letter *X* depicting the unknowns, and small labeled weights (e.g., 50 g, 100 g) depicting the numbers (for a similar model, see Suh & Moyer, 2007). Here, while students manipulate the virtual balance scale, the corresponding equation is shown in formal algebraic symbols, which makes the link between these manipulations and the changes in the corresponding symbolic equation explicit. A further type of virtual balance model was found in the article of Kaplan and Alon (2013; Figure 3b). In this model, students can explore relationships between different shapes of unknowns and find new equations based on given ones. For example, on the basis of the equations▲ ▲ = ●●● and ▲ ▲ = ●●■■, a third equation can be created.

3.2.3 Drawn balance models

Drawn balance models appeared in 26 articles (four and three from the same research projects). Drawings of the used drawn balance models are shown in Figure 4. Here, it is noticeable that some drawn balance models are depicted more realistically (Figure 4a-c) and others more schematically (Figure 4d-f), with pictures of objects or symbolic expressions to represent the knowns and unknowns.

While drawn balance models were present in many articles (e.g., Brodie & Shalem, 2011; Mann, 2004; Vlassis, 2002), the way in which the equations are represented in these models varied widely. In the drawn balance model found in the article of Vlassis (2002; Figure 4a), the equation $7x + 38 = 3x + 74$ is represented by squares for each x and circles in which the numbers are indicated. The unknowns in this model are depicted in an expanded way (i.e., $7x$ and $3x$ are represented as seven separate x's and three separate x's). While in most models all unknowns are depicted separately, in the model of Linchevski and Herscovics (1996), the unknowns and knowns in the equation $8n + 11 = 5n + 50$ are partially shown in an expanded, respectively decomposed way, leading to the equation $5n + 3n + 11 = 5n + 11 + 39$. In this way, students can see that the terms $5n$ and 11 appear on both sides of the equation, which can cancel each other out. In the balances of Marschall and Andrews (2015; Figure 4b) and Warren and Cooper (2009; Figure 4c), equations with negative values and subtractions can also be represented. In Figure 4b, the subtraction in the equation $4x - 3 = 2x + 5$ is represented by an arrow going down from one of the scales, so that the action of "taking away" is made visible. Alternatively, in Figure 4c, a minus sign is included.

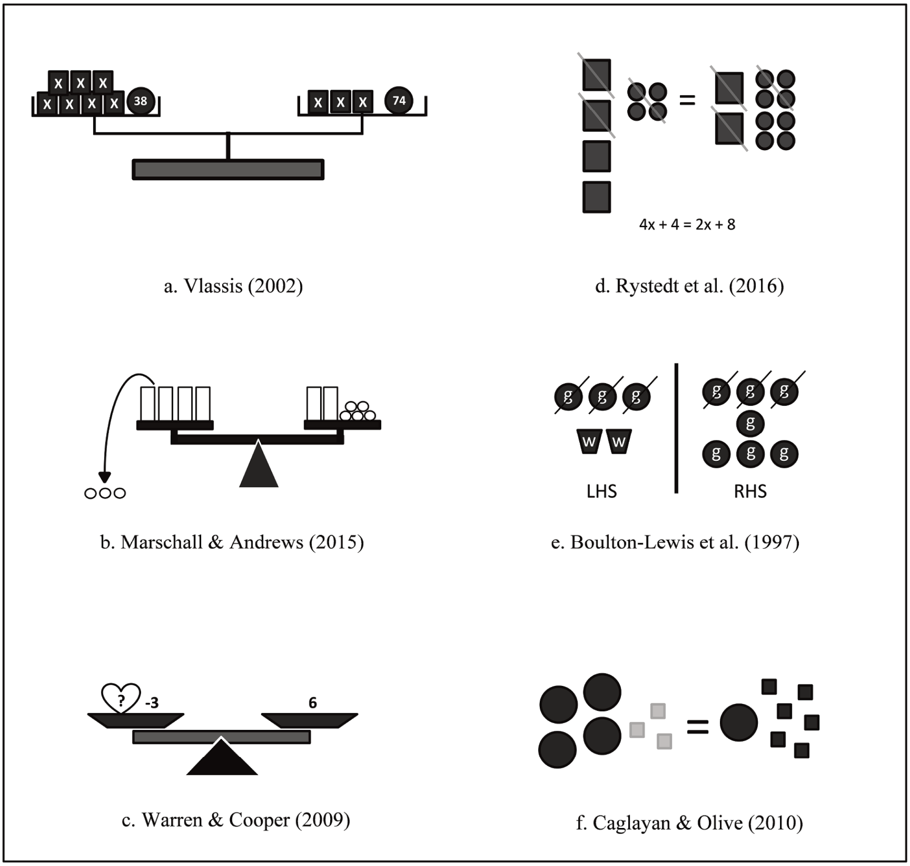


Figure 4. Drawn balance models, examples from six articles (a-f).

Another way in which drawn balance models appeared in the articles is as an abstract drawing. Here, the balance functions as a metaphor to point students' attention to the concept of equality. In Rystedt et al. (2016, Figure 4d) the equation $4x + 4 = 2x + 8$ is represented with boxes for unknowns and dots for numbers. In articles in which such a metaphorical use of the balance model was present (e.g., Caglayan & Olive, 2010), this use was often accompanied by the instruction that the *balance* in the equation had to be *maintained* when solving the equation (Boulton-Lewis et al., 1997), or by gestures representing a balance scale (Rystedt et al., 2016). The use of a drawn balance model, especially for models with an abstract drawing, often went together with the use of manipulatives. For example, in the model of Boulton-Lewis et al. (1997; Figure 4e), the schematically notated equation $2x + 3 = 7$ is represented by two white cups and three green counters on the left-hand side (indicated by LHS) and seven green counters on the right-hand side (indicated by RHS), while other colored cups and counters are used to represent subtractions or negative unknowns and numbers (for a similar approach, see e.g., Suh & Moyer, 2007). Another example is the drawn balance model used by Caglayan and Olive (2010; Figure 4f), where in the equation $4x - 3 = x + 6$ the "-3" is represented by gray tiles instead of black ones. Moreover, the equal sign is directly represented in this model.

3.2.4 Discussion of the findings regarding the types of used balance models

Drawn models appeared the most and virtual models the least, while the use of a physical model was often followed by the use of a drawn model. When looking into the relationship between the rationales and the appearances of the models, it seems that the use of a physical balance model most often goes together with rationales related to learning through physical experiences and the equality aspect. For the virtual models, all rationales appear more or less equally often, and the drawn balance models go most often together with the equality aspect rationale and rationales related to learning through models and representations. Except for rationales related to learning through physical experiences, the remaining two classes of rationales most often go together with the use of a drawn balance model. The drawn model appears to be the most flexible model, which means that it was used with all classes of rationales.

Although all three appearances of the model have the balance as a basic concept, they differ in their nature. Whereas the physical balance model and partly the virtual balance have a dynamic nature and as such can provide real-time feedback to the students about their actions, the drawn balance model is static. Drawn models, either presented on paper or on the blackboard, can nonetheless be extended with dynamic aspects by using manipulatives. For all three types of appearances of the model it applies that most models consist at least of a fulcrum, a horizontal balancing beam, and on both sides a scale. In addition to this configuration of the balance model, in other models, extra features are added. Through the addition of these features, the reach of the balance model is extended to represent a wider range of problems. For example, the additional scales in the physical model of Orlov (1971; Figure 2c), the arrow going down from the scales of the drawn balance model in the article of Marschall and Andrews (2015; Figure 4b), and the different colored manipulatives added to the drawn model of Boulton-Lewis et al. (1997; Figure 4e), are all examples of variations of the balance model allowing the representation of negative numbers and unknowns. Such additional features provide a solution for the restricted possibilities that this model has (e.g., Vlassis, 2002), for example by allowing for the representation of equations with negative quantities or subtractions. In fact, this flexibility of the balance model is exactly how models should work. When used as didactical models (Van den Heuvel-Panhuizen, 2003), models should be flexible and not only suitable for solving one type of equation. One way of ensuring this flexibility is by allowing for adaptations without losing its primary function. However, bearing in mind the concept of *model of … – model for …* (Streefland, 2003), didactical models are not meant as a tool that everlasting has to be used for problem solving at a concrete, context-connected level. Instead, the idea is that in a later phase of the learning process, when a basis is laid for solving linear equations and the students have to solve more difficult equations, the student's thinking can still be supported by, and related to, the model without concretely representing the equation in a physical model.

3.3 When was the balance model used?

The situations in which the balance model was used in the articles when describing the teaching of linear equation solving, varied considerably with respect to the grade level of the students involved, the duration of the intervention with the model, the type of equation problems that students worked on, and the type of instruction that was provided to the students.

3.3.1 Grade levels and intervention duration

The balance model was used to teach linear equation solving to students from Kindergarten to Grade 9. Students up to Grade 6, who do not have previous experience with algebra, had their first encounter with linear equations through the balance model, which came forward in different studies (e.g., Warren & Cooper, 2005). In studies with students from Grades 7-9, who already have some basic experience with linear equation solving (with the exception of the seventh-grade students in the study by Araya et al., 2010), the balance model was introduced as a tool for solving equations (Vlassis, 2002) or used to illustrate the balance method (i.e., perform the same operations on both sides of the equation; Ngu & Phan, 2016). The duration of the interventions in which the balance model was used was also very diverse. The shortest interventions comprised one activity or one lesson (e.g., Figueira-Sampaio et al., 2009; Rystedt et al., 2016), while in other studies the balance model was integrated in a multiple-year teaching trajectory (e.g., Orlov, 1971; Warren & Cooper, 2009).

3.3.2 Type of equation problems

With very young students (e.g., Kindergarten, Grades 1-2), the balance model was mostly used for exploration of the first ideas of equality and the equal sign (e.g., Taylor-Cox, 2003; Warren et al., 2009). Students' task was, for example, to weigh different objects to find out which were the same and which were different. For somewhat older students (e.g., Grades 3-6), the balance model was, for example, used to assist them in solving simple addition problems such as $8 = \_\_ + 3$ (e.g., Leavy et al., 2013). Here, eight objects were put on the left side of the balance and three objects on the right side, and the students' task was to figure out what they could do to make both sides equal. The model was also used to introduce algebraic symbols to students without prior algebra experience, so that they could link the model to the abstract symbols. Then students' task was, for example, to manipulate the objects on the scales in such a way that they could determine the weight of the unknown object, while in the meantime in the digital environment the corresponding symbolic equation was shown (e.g., Figueira-Sampaio et al., 2009, see Figure 3a; Suh & Moyer, 2007). In research with students with some algebra experience (i.e., from Grade 7 on), students' task was, for example, to represent symbolic equations by making use of the balance model and to use this representation to transform and solve the equations (Caglayan & Olive, 2010; see Figure 4f). Or students' task was to solve an equation by making use of a physical balance model, while subsequently to represent the equation and the solution steps symbolically (Andrews, 2003). There were also articles in which two balance models with different unknowns were presented simultaneously to create a system of equations and to evoke the algebraic strategy of substitution (e.g., Austin & Vollrath, 1989; Berks & Vlasnik, 2014). Here, students' task was to combine the information of the equations to find the values of the unknowns.

In most studies, students' task was to determine the value of the unknown(s). However, there were also articles in which the main purpose was to discover different possibilities to maintain the balance of the model, without focusing on finding values of unknowns. For example, in the study by Kaplan and Alon (2013), the goal was to create multiple balanced scales and to analyze the relationships between unknowns (see Figure 3b). Also in other articles, the balance model was used to discover different possibilities to maintain equality (Orlov, 1971) or to discover which "legal moves" (Raymond & Leinenbach, p. 288) could be made without disturbing the balance.

Lastly, there were large varieties between studies concerning maintaining the balance model when teaching equations. For example, in Warren and Cooper (2005), first a physical balance model and later a drawn balance model were used to model equations containing positive values and additive operations (e.g., $? + 7 = 11$). After some lessons, these students also solved equations containing subtraction (e.g., $? - 4 = 13$), but these equations were not represented with the balance model. In other studies the use of the balance model was maintained longer during the learning process. For example, one of the teachers in the study by Marschall and Andrews (2015) did not only use the model for teaching equations containing positive values and addition, but extended the use of the model to represent equations such as $4x - 3 = 2x + 5$ (see Figure 4b; for using the model for other type of equations, see e.g., Boulton-Lewis et al., 1997, see Figure 4e; Orlov, 1971, see Figure 2c).

3.3.3 Type of instruction

When working with the balance model, students either received classroom instruction by a teacher (e.g., Warren & Cooper, 2009) or via a learning movie (Araya et al., 2010), or they received individual instruction by a teacher (e.g., Perry et al., 1995), through instruction sheets (Ngu et al., 2015), or through working individually or in pairs with the balance (e.g., students working with the virtual balance in Figueira-Sampaio et al., 2009). Classroom instruction often concerned the teacher manipulating a balance model in front of the classroom (e.g., students working with the physical balance model in Figueira-Sampaio et al., 2009), while during individual instruction, students more often got opportunities to actively work with the balance model themselves (e.g., Suh & Moyer, 2007).

3.3.4 Discussion of the findings regarding when the balance model was used

In what situations the balance model was used was very diverse in the different studies. For what equation problems the balance model was used appeared to be related to students' experience with solving linear equations. For students up to Grade 6, without previous experience with algebra, most tasks concentrated around exploring the basic ideas of balance and solving simple equations (e.g., 8 = __ + 3), which went hand in hand with the rationale that such activities can be beneficial for developing understanding of equality and a relational understanding of the equal sign. Physical and virtual balance models were relatively often used to teach linear equation solving to students without prior algebra experience. In most of these studies, equations only contained positive values and additive operations. The studies conducted with students without prior experience in general underpinned the use of the balance model for teaching linear equation solving more thoroughly than studies with students with some algebra experience. The rationale that was relatively often mentioned in relation to teaching students without prior algebra experiences is the rationale related to the physical experiences, which fits the using of the physical balance model to teach these students. This also aligns with the common trend of using concrete materials for teaching young students rather than for teaching older students and with research showing that the use of concrete materials in mathematics education is in particular beneficial for children aged 7-11, in the mathematical domains of fractions and algebra (Carbonneau et al., 2013).

With regard to studies conducted with students *with* prior algebra experience (in general students from Grades 7 and higher), students' tasks when working with the balance model were most often to model, to transform, and to solve equations by means of the balance model. Also in these studies, the rationale related to the equality aspect was most prominent. On the contrary, most of the studies in which no rationale for using the model was provided were also conducted with students with prior algebra experience. Most studies in which a limitation of using the balance model was mentioned involved these students. Drawn balance models were mostly used to teach students with prior algebra experience and in more than half of these studies, students were also taught equations containing negative values and subtraction.

3.4 Learning outcomes

Nineteen articles evaluated students' learning outcomes related to the use of the balance model. The research design of these studies and the most important learning outcomes are summarized in Table 1. Most studies were descriptive in nature and less than one-third of the studies used a pre-posttest design combined with a comparison group. As described in section "3.3 When was the balance model used?", the studies showed large variations as regards the age and algebra experience of the students in their sample, the duration of the intervention, the tasks students worked on, and the type of instruction students received. Similar variations were detected upon examining the learning outcomes of the different studies. For example, Araya et al. (2010) found very positive results of using a learning movie with a drawn balance model in Grade 7 with students without prior algebra experience. These students outperformed students in the comparison group who received symbolic linear equation solving instruction. Also, Suh and Moyer (2007) reported positive effects of using balance models to teach third-grade students linear equation solving. Contrastingly, Boulton-Lewis et al. (1997) found that eighth-grade students had difficulties with modeling and solving linear equations when making use of the balance model. These students preferred not to use the model. The studies by Ngu et al. (2015, 2016, 2018) consistently showed similar or lower performances for Grade 7-9 students who used the method of performing the same operations on both sides of the equation, which was taught by making use of an instruction sheet with the balance model, compared to students who used the inverse method, which was taught as by referring to the change side, change sign-rule for solving equations. In this latter approach, in which, for example, $x - 4 = 6$ becomes $x = 6 + 4$, students can conceptualize the inverse operation of -4 becoming +4 as a means to preserve the equality of equations. Therefore, the understanding of this inverse principle at a structural level is considered to be very relevant for students' learning of algebraic thinking (e.g., Ding, 2016). Interesting to notice here is that, although viewed superficially, the balance method differs from the inverse method, this latter method bears much resemblance to "doing the same on both sides". When taking the example of $x - 4 = 6$, then this rule means that on both sides 4 has to be added. This makes $x - 4 + 4 = 6 + 4$, which after simplifying results into $x = 6 + 4$. In other words, the main difference between "doing the same things on both sides" and "change sides, change sign" involves that one comes directly to the result by skipping the intermediate step of adding 4 on both sides. However, despite the close relationship between these two approaches and the related underlying principles, in only a few articles of our review study when authors refer to the use of the balance model, they also refer to the inverse method. This indicates that there has not been much research in which both approaches have been put in relation or contrasted.

The large variation between studies in which the balance model was used and the lack of studies with an experimental research design make it very difficult to draw unequivocal conclusions about the effects of using the balance model on students' learning outcomes. Nevertheless, some trends can be identified. Overall, most mixed and negative results are found for studies with somewhat older students (Grades 7-9) who already had some (basic) experience in solving linear equations (e.g., Ngu et al., 2018; Vlassis, 2002). The main reasons for this finding could be that the balance models in these studies, which were all drawn, were used for teaching a broad range of equations, including more difficult equations such as equations containing negative numbers and unknowns (e.g., Boulton-Lewis et al., 1997; Caglayan & Olive, 2010, Vlassis, 2002). In general, more positive results were found for studies conducted with younger students (e.g., Suh & Moyer, 2007; Warren & Cooper, 2005) or with students without prior knowledge on equation solving (e.g., Araya et al., 2010). In these studies more often a physical model (e.g., Perry et al., 1995; see Figure 2d) or a virtual model (e.g., Figueira-Sampaio et al., 2009; see Figure 3a) was used, which in some cases in later stages was followed by a drawn model (e.g., Warren & Cooper, 2005). In most of these studies, the balance model was used to teach linear equations containing only positive values and addition. However, there were some exceptions. For example, Orlov (1971) found positive results for teaching different types of linear equations (including negative values and subtraction) to eighth-grade students by making use of a physical balance model (see Figure 2c).

3.4.1 Discussion of the learning outcomes

Overall, the balance model seems to have more positive effects on learning outcomes related to linear equation solving for (younger) students without prior knowledge on linear equation solving. A possible explanation might be that for younger students, the balance model is used for laying a conceptual basis for linear equation solving, while for older students, who already have such a basis in solving linear equations, the model is more often used to revitalize this basis. Younger students have their first experience with exploring the concept of equality and with linear equation solving by means of the balance model. The tasks of older students when working with the balance model are more often to model, to transform, or to solve equations. In other words, the balance model then is used to revitalize their knowledge on linear equation solving and assist in solving all kinds of new equations. Warren and Cooper (2005) provide an example of using the balance model to support students in solving equations containing subtraction. In their teaching sequence, they first used a physical model to let students develop understanding of the concept of equality as “balance” and the strategy of doing the same things on both sides. Later, students could use this strategy for solving symbolically notated problems on paper, which also contained subtraction.

4. Conclusion

Our systematic review reveals a rather kaleidoscopic image of the balance model as an aid for teaching linear equations. The findings on *why* a balance model was used, *what* types of models were used, *when* they were used, and what learning outcomes were associated with its use, are diverse. Nevertheless, we could identify some trends within this scattered picture. Physical and virtual balance models were more often mentioned in the articles for teaching students during their first encounter with linear equation solving. Also, authors of these articles were more explicit about their rationales for using the balance model, with most rationales related to the equality aspect and students’ physical experiences. The equations taught to these students mostly only contained positive values and addition, and these studies in general reported positive effects of using the balance model on students’ learning outcomes of linear equation solving. Drawn balance models were more often used for students who already had some previous algebra experience. Additional features (such as manipulatives) were often added to these models, so that a wider range of problems could be represented, such as equations with negative values and subtraction. Articles in which drawn balance models were used were less explicit about their reasons for using the balance model, and in general reported more mixed and negative effects of using the balance model on teaching linear equation solving. However, it is important to note that within these trends, there were still many differences between studies, for example concerning the duration of the intervention and the type of instruction provided to the students.

These trends should of course be interpreted with caution. First of all, our results are entirely based on what the authors of the articles reported. In some articles, the authors did not explicitly report their rationales for using the balance model, which meant that they could not be identified by our analysis. Secondly, although we searched for articles in which the use of the balance model was discussed in 93 peer-reviewed journals to ensure a good coverage of the research literature, we only had a relatively small final sample of 34 articles that met our inclusion criteria. In addition, within the limited time we had available for this study, we could not consider including textbooks or other curriculum documents. Furthermore, we decided to do the review on articles in which the balance model was used for teaching linear equation solving and leave out other mathematical topics in which the balance model could be used. Lastly, a limitation that also should be mentioned is that our study only focused on the balance model and we did not compare it with other often used methods for helping students to solving linear equations such as the change side change sign-rule. Clearly, more research is necessary in this respect.

Our study was meant to create an overview of the role the balance model plays in teaching linear equation solving that might provide teachers, researchers, and developers of instructional materials with a source for making informed instructional decisions. Our analysis of the 34 peer-reviewed journal articles shows that there exists a considerable diversity in the rationales for using the model, the appearances of the model, the situations in which the model is used, and the found learning outcomes. This offers many possibilities for making use of the balance model. However, at the same time, our study reveals a clear lack of in-depth knowledge about when which type of balance models can be used effectively. For gaining this knowledge more research is necessary, in particular (quasi-) experimental studies, allowing to investigate the effects of using models of different appearances (e.g., physical, virtual, and drawn models, with or without additional features such as added scales or the use of manipulatives) and the effects of different situations of using the model (e.g., for students with or without prior algebra experience, a short-time use of the model or a more extensive intervention, with one type of instruction or another) on students' learning outcomes. To provide a more theoretical grounding for the use of the balance model as an aid for teaching linear equations, it is important that the type of model that is used and the situations in which it is used are explicitly related to the rationales for using them. In summary, we can conclude that the balance model, which at first sight may seem to be a rather simple model — and maybe therefore is often used to teach students linear equation solving — is actually a rather complex model, of which still a lot has to be discovered to be used optimally in education.

References

References marked with an asterisk indicate studies included in the systematic literature review.

Abrahamson, D. (2017). Embodiment and mathematics learning. In K. Peppler (Ed.), *The SAGE encyclopedia of out-of-school learning* (pp. 247–252). SAGE Publications, Inc.

Alessandroni, N. (2018). Varieties of embodiment in cognitive science. *Theory & Psychology*, *28*(2), 227–248. https://doi.org/10.1177/0959354317745589

*Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, *35*(1), 127–145. https://doi.org/10.1037/0012-1649.35.1.127

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal look at middle-school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning, 9*(3)*,* 221–247. https://doi.org/10.1080/10986060701360902

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

*Andrews, P. (2003). Opportunities to learn in the Budapest mathematics classroom. *International Journal of Science and Mathematics Education*, *1*(2), 201–225. https://doi.org/10.1023/B:IJMA.0000016874.07706.5a

*Andrews, P., & Sayers, J. (2012). Teaching linear equations: Case studies from Finland, Flanders and Hungary. *The Journal of Mathematical Behavior*, *31*(4), 476–488. https://doi.org/10.1016/j.jmathb.2012.07.002

Antle, A. N., Corness, G., & Bevans, A. (2013). Balancing justice: Comparing whole body and controller-based interaction for an abstract domain. *International Journal of Arts and Technology*, *6*(4), 388–409. https://doi.org/10.1504/IJART.2013.058285

*Araya, R., Calfucura, P., Jiménez, A., Aguirre, C., Palavicino, M. A., Lacourly, N., Soto-Andrade, J., & Dartnell, P. (2010). The effect of analogies on learning to solve algebraic equations. *Pedagogies: An International Journal*, *5*(3), 216–232. https://doi.org/10.1080/1554480X.2010.486160

*Austin, J., & Vollrath, H. J. (1989). Representing, solving, and using algebraic equations. *Mathematics Teacher*, *82*(8), 608–612.

Ballheim, C. (1999, October). Readers respond to what's basic. *Mathematics Education Dialogues*, *3*, 11.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

*Berks, D. R., & Vlasnik, A. N. (2014). Working the system. *Mathematics Teacher*, *107*(7), 542–546. https://doi.org/10.5951/mathteacher.107.7.0542

*Boulton-Lewis, G., Cooper, T. J., Atweh, B., Pillay, H., Wilss, L., & Mutch, S. (1997). Processing load and the use of concrete representations and strategies for solving linear equations. *The Journal of Mathematical Behavior*, *16*(4), 379–397. https://doi.org/10.1016/S0732-3123(97)90014-6

*Brodie, K., & Shalem, Y. (2011). Accountability conversations: Mathematics teachers' learning through challenge and solidarity. *Journal of Mathematics Teacher Education*, *14*(6), 419–439. https://doi.org/10.1007/s10857-011-9178-8

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

*Caglayan, G., & Olive, J. (2010). Eighth grade students' representations of linear equations based on a cups and tiles model. *Educational Studies in Mathematics*, *74*(2), 143–162. https://doi.org/10.1007/s10649-010-9231-z

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

Carraher, D. W., Schliemann, A. D., & Schwartz, J. (2008). Early algebra is not the same as algebra early. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 235–272). Lawrence Erlbaum Associates.

*Cooper, T. J., & Warren, E. (2008). The effect of different representations on Years 3 to 5 students' ability to generalise. *ZDM Mathematics Education*, *40*(1), 23–37. https://doi.org/10.1007/s11858-007-0066-8

Ding, M. (2016). Opportunities to learn: Inverse relations in U.S. and Chinese textbooks. *Mathematical Thinking and Learning, 18*(1), 45–68. https://doi.org/10.1080/10986065.2016.1107819

Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children's understanding of equality: A foundation for algebra. *Teaching Children Mathematics*, *6*(4), 232–236.

*Figueira-Sampaio, A. S., Santos, E. E. F., & Carrijo, G. A. (2009). A constructivist computational tool to assist in learning primary school mathematical equations. *Computers & Education, 53*(2), 484–492. https://doi.org/10.1016/j.compedu.2009.03.012

*Filloy, E., & Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics*, *9*(2), 19–25.

*Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction*, *35*, 104–120. https://doi.org/10.1016/j.learninstruc.2014.10.004

*Gavin, M. K., & Sheffield, L. J. (2015). A balancing act: Making sense of algebra. *Mathematics Teaching in the Middle School, 20*(8), 460–466. https://doi.org/10.5951/mathteacmiddscho.20.8.0460

Gibbs Jr, R. W. (2006). *Embodiment and cognitive science*. Cambridge University Press.

Huntley, M. A., & Terrell, M. S. (2014). One-step and multi-step linear equations: A content analysis of five textbook series. *ZDM Mathematics Education*, *46*(5), 751–766. https://doi.org/10.1007/s11858-014-0627-6

*Jupri, A., Drijvers, P., & Van den Heuvel-Panhuizen, M. (2014). Difficulties in initial algebra learning in Indonesia. *Mathematics Education Research Journal*, *26*(4), 683–710. https://doi.org/10.1007/s13394-013-0097-0

*Kaplan, R. G., & Alon, S. (2013). Using technology to teach equivalence. *Teaching Children Mathematics*, *19*(6), 382–389. https://doi.org/10.5951/teacchilmath.19.6.0382

Kieran, C. (1997). Mathematical concepts at the secondary school level: The learning of algebra and functions. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 133–158). Psychology Press.

Kieran, C. (2007). Learning and teaching algebra at the middle school through college levels. Building meaning for symbols and their manipulation. In, F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, 707–762). Information Age Publishing.

Kieran, C., Pang, J., Schifter, D., & Ng, S. F. (2016). *Early algebra: Research into its nature, its learning, its teaching*. Springer (open access eBook). https://doi.org/10.1007/978-3-319-32258-2

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312. https://www.jstor.org/stable/30034852

*Leavy, A., Hourigan, M., & McMahon, Á. (2013). Early understanding of equality. *Teaching Children Mathematics*, *20*(4), 246–252. https://doi.org/10.5951/teacchilmath.20.4.0246

Leibniz, G. W. (1989). To queen Sophie Charlotte of Prussia, on what is independent of sense and matter (1702). In R. Ariew, & D. Garber (Eds., Trans.), *Philosophical essays* (pp. 186–192). Hackett Publishing Company.

*Linchevski, L., & Herscovics, N. (1996). Crossing the cognitive gap between arithmetic and algebra: Operating on the unknown in the context of equations. *Educational Studies in Mathematics*, *30*(1), 39–65. https://doi.org/10.1007/BF00163752

*Mann, R. L. (2004). Balancing act: The truth behind the equals sign. *Teaching Children Mathematics*, *11*(2), 65–70.

*Marschall, G., & Andrews, P. (2015). Polish teachers' conceptions of and approaches to the teaching of linear equations to grade six students: An exploratory case study. *Research in Mathematics Education*, *17*(3), 220–238. https://doi.org/10.1080/14794802.2015.1107498

*Ngu, B. H., Chung, S. F., & Yeung, A. S. (2015). Cognitive load in algebra: Element interactivity in solving equations. *Educational Psychology*, *35*(3), 271–293. https://doi.org/10.1080/01443410.2013.878019

*Ngu, B. H., & Phan, H. P. (2016). Comparing balance and inverse methods on learning conceptual and procedural knowledge in equation solving: A cognitive load perspective. *Pedagogies: An International Journal*, *11*(1), 63–83. https://doi.org/10.1080/1554480X.2015.1047836

*Ngu, B. H., Phan, H. P., Yeung, A. S., & Chung, S. F. (2018). Managing element interactivity in equation solving. *Educational Psychology Review*, *30*(1), 255–272. https://doi.org/10.1007/s10648-016-9397-8

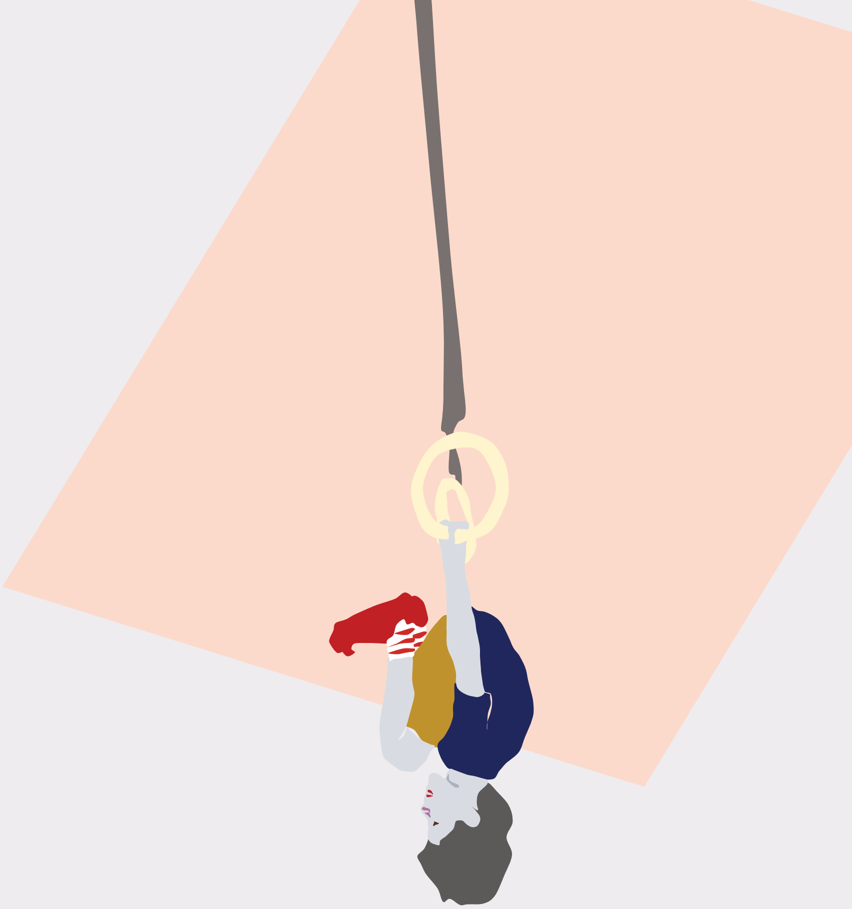
Núñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics*, *39*(1–3), 45–65. https://doi.org/10.1023/A:1003759711966

*Orlov, K. (1971). Experimental verification of the use of the mathematical balance in secondary teaching. *Educational Studies in Mathematics*, *3*(2), 192–205. https://doi.org/10.1007/BF00305448

*Perry, M., Berch, D., & Singleton, J. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues*, *6*, 213–235.

*Raymond, A. M., & Leinenbach, M. (2000). Collaborative action research on the learning and teaching of algebra: A story of one mathematics teacher's development. *Educational Studies in Mathematics*, *41*(3), 283–307. https://doi.org/10.1023/A:1004091631981
*Rystedt, E., Helenius, O., & Kilhamn, C. (2016). Moving in and out of contexts in collaborative reasoning about equations. *The Journal of Mathematical Behavior*, *44*, 50–64. https://doi.org/10.1016/j.jmathb.2016.10.002
*Smith, L. R. (1985). Mathematics on the balance beam. *School Science and Mathematics*, *85*(6), 494–498. https://doi.org/10.1111/j.1949-8594.1985.tb09653.x
Streefland, L. (2003). Learning from history for teaching in the future. *Educational Studies in Mathematics*, *54*(1), 37–62. https://doi.org/10.1023/B:EDUC.0000005236.36499.fc
*Suh, J., & Moyer, P. S. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching*, *26*(2), 155–173.
*Taylor-Cox, J. (2003). Algebra in the early years. *Young Children*, *58*(1), 14–21.
Van den Heuvel-Panhuizen, M. (2003). The didactical use of models in realistic mathematics education: An example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics*, *54*(1), 9–35. https://doi.org/10.1023/B:EDUC.0000005212.03219.dc
*Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*(3), 341–359. https://doi.org/10.1023/A:1020229023965
*Warren, E., & Cooper, T. J. (2005). Young children's ability to use the balance strategy to solve for unknowns. *Mathematics Education Research Journal*, *17*(1), 58–72. https://doi.org/10.1007/BF03217409
*Warren, E., & Cooper, T. J. (2009). Developing mathematics understanding and abstraction: The case of equivalence in the elementary years. *Mathematics Education Research Journal*, *21*(2), 76–95. https://doi.org/10.1007/BF03217546
*Warren, E., Mollinson, A., & Oestrich, K. (2009). Equivalence and equations in early years classrooms. *Australian Primary Mathematics Classroom*, *14*(1), 10–15.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Author contributions: All authors participated in designing the research. MO performed the systematic search and performed the first selection of articles. All authors participated in selection and coding procedures. MO analyzed the data, of which the findings were frequently discussed with all authors. MO prepared the first draft of the manuscript. All authors participated in revising the manuscript and/or provided feedback. All authors read and approved the final manuscript.



CHAPTER 3

Developing algebraic reasoning in primary school using a hanging mobile as a learning supportive tool

Mara Otten, Marja van den Heuvel-Panhuizen, Michiel Veldhuis, and Aiso Heinze

Published in *Journal for the Study of Education and Development, 2019, 42*(3), 615–663.

Developing algebraic reasoning in primary school using a hanging mobile as a learning supportive tool

Abstract

In this study, we investigated the development of algebraic reasoning in 65 fifth-grade students who never had algebra instruction before. In the six-lesson teaching sequence on solving linear equations, a *hanging mobile*, a physical balance model, played the central role. We expected students' perceptual-motor experiences with this hanging mobile to be beneficial for the development of their reasoning related to linear equation solving. To investigate how students' reasoning developed, we analyzed videos of classroom interactions, students' written work during the lessons and students' responses to lesson-specific assessment tasks. Our results reveal that students showed progress in their level of algebraic reasoning and in their writing down of strategies. While working with the hanging mobile, students applied algebraic strategies such as restructuring, isolation and substitution. They later used these algebraic strategies for solving linear equations in new contexts. This suggests that the experiences students gained in the embodied learning environment provided a basis for algebraic reasoning, which appeared to support them when solving systems of symbolically presented linear equations.

Keywords: Early algebra, Solving linear equations, Algebraic reasoning, Embodiment, Qualitative data analysis

1. Introduction

Early algebra refers to the introduction of algebraic thinking in the elementary grades (Kaput et al., 2008) and aims for building a foundation for providing students access to more advanced algebraic concepts in later grades (e.g., Carraher & Schliemann, 2014). Early algebra does not simply mean “teach algebra earlier” and teaching formal algebra to young students (Carraher et al., 2008). Instead, algebra is conceived as a longitudinal strand of thinking and problem solving, starting in kindergarten and continuing in the higher grades (e.g., Kaput, 2008). Early algebra entails a number of big ideas such as equivalence and equations, variable and proportional reasoning (Blanton et al., 2015). From early on, students can be involved in problem-solving activities related to these ideas by drawing on their existing knowledge of skills and procedures in the domain of number (Blanton et al., 2007). Furthermore, it takes students’ everyday experiences and their “intuitive and informal ways of reasoning” (Stephens et al., 2017, p. 143) as the basis for eliciting and developing algebraic thinking that is, for example, necessary for solving elementary linear equations. In this article, we describe how, through participating in embodied activities over the course of a six-lesson teaching sequence on linear equation solving, elementary school students developed context-based algebraic strategies which ultimately enabled them to solve a system of two linear equations with two unknowns.

1.1 Linear equation solving

The main characteristic of an equation is that the expressions on both sides of the equal sign represent the same value. In this sense, both sides of the equation are equal — though they might look different. In the process of solving an equation, this equality should be maintained. Equality is therefore deemed a key concept in linear equation solving (e.g., Bush & Karp, 2013; Knuth et al., 2005; Li et al., 2008). Understanding this concept is one of the main demands when solving linear equations (Kieran, 1997; Kieran et al., 2016).

When solving an equation, the goal is to find the relationships between or the value(s) of the unknown(s) for which the expressions on both sides of the equal sign are equal. This goal can be reached by isolating the unknown. In order to achieve this, students have to operate with the expressions in such a way that their equality is maintained. This is often emphasized by referring to the metaphor of the equation being *in balance* (Ngu & Phan, 2016) and implies performing transformations under the principle of “doing the same on both sides”. Among other things, this involves the preparatory activity of restructuring the given equation by, for example, changing the order of the terms in the expressions by using the commutative and associative properties, replacing expressions by equivalent expressions, or exchanging the two sides, which is possible due to the symmetry of the equality. At a formal level, this restructuring can entail notating unknowns in an expanded (e.g., $A + A$) or condensed (e.g., $2A$) way. Further simplifying an equation to isolate the unknown can, for example, be done by taking away the same elements from both sides or by halving the expressions on both sides. A less targeted method for solving equations is guessing-and-checking. This implies that different values for an unknown are put in an equation until the correct answer is obtained (e.g., Stacey & MacGregor, 1999).

When solving a system of linear equations with multiple unknowns, the information from more than one equation has to be used to find the values of the unknowns. In this case, the isolation method can be applied by eliminating one unknown by means of, for example, subtracting one equation from the other. Yet, the most prominent method to solve a system of linear equations with multiple unknowns is the method of substitution (replacing an expression with another expression of the same value).

Crucial in using these strategies is that students have a good understanding of the concept of equality and the properties of operations. Many of the difficulties students experience when solving (systems of) linear equations have to do with an incorrect interpretation of the equal sign. Students consider it as a symbol for "here comes the answer" or as a sign to "add" (e.g., Behr et al., 1980; Carpenter et al., 2003), instead of as a symbol reflecting the sameness of the expressions on both sides (e.g., Knuth et al., 2006). This latter, proper conception of the equal sign as a relational symbol is positively related to competence in solving linear equations (e.g., Alibali et al., 2007; Knuth et al., 2006; Matthews et al., 2012). Other well-known difficulties for students related to linear equation solving are difficulties with understanding the formal, symbolic representations of equations (e.g., Koedinger & Nathan, 2004) and difficulties with performing operations on the unknowns when solving equations (e.g., Filloy & Rojano, 1989; Herscovics & Linchevski, 1994).

1.2 Teaching linear equation solving

Approaches that have been developed for teaching linear equation solving are all trying, in one way or another, to overcome the aforementioned difficulties. Within Realistic Mathematics Education (RME; e.g., Van den Heuvel-Panhuizen &

Drijvers, 2014), this means starting with meaningful situations in which the students can solve informal equation-like problems with context-connected strategies and then, through a process of progressive mathematization, come to more formal and standard solution processes. In a study by Van Reeuwijk (1995), students were asked to figure out what a T-shirt cost and what a soda cost, when two T-shirts and two sodas cost $44 and one T-shirt and three sodas cost $30. Students are more prone to solving such a problem that is meaningful to them. When students can make sense of a problem, or in other words, when they can imagine what happens in the problem, it can open their action repertoire for solving it (Van den Heuvel-Panhuizen & Drijvers, 2014). In this case, by halving the first set of T-shirts and sodas, students know what one T-shirt and one soda cost, and then they can find that two sodas cost $8. The strategies of isolation (by elimination) and substitution developed for solving this problem can later be used for solving other systems of linear equations. In other words, the meaningful context of the T-shirts and the sodas can become a model for solving other problems.

In addition to this context model, other mathematical models are used for teaching linear equation solving as well, such as area models, linear models and balance models (Van Amerom, 2002). These models are also meant to help students in understanding what it means to solve an equation and develop strategies to find the values of unknowns. Of these models, the balance model has a long didactical history. It was already used by Leibniz (1646-1716), who mentioned the relation between equality in a mathematical situation and a balance with equal things on both sides (Leibniz, 1989).

A recently conducted literature review (Otten et al., 2019, see *Chapter 2* of this thesis) showed that in teaching linear equation solving, various appearances of the balance model are used. It can be a concrete physical device, a drawn balance on a work sheet or a virtual balance in a digital environment. These types of balance models, which in some studies are used in combination, each have their own affordances. Concrete balances are chosen because the physical experiences that students can have when working with such a device are considered to be beneficial for the learning process. A recent meta-analysis on the efficacy of using concrete manipulatives for teaching mathematics found small to moderate effect sizes in favor of using manipulatives compared to providing abstract symbolic instruction (Carbonneau et al., 2013). A theoretical perspective on the working mechanisms behind learning with manipulatives is provided by embodied cognition theory (e.g., Pouw et al., 2014). This theory states that bodily experiences can be advantageous for cognitive learning processes (e.g., Wilson, 2002). Cognition is formed not only by and in our brains, but also by perceptual-motor experiences that we have when our bodies interact with the world (e.g., Gallese & Lakoff, 2005; Wilson, 2002). Therefore, embodied learning environments (short for: learning environments based on embodied cognition theory) are regarded as essential for learning mathematics (Abrahamson, 2017; Lakoff & Núñez, 2000; Núñez et al., 1999). By coupling action and perception, perceptual-motor experiences form the basis for the emergence of mathematical concepts and mathematical reasoning, mediated by conceptual metaphors (Lakoff & Johnson, 1980) and representational redescription (Karmiloff-Smith, 1992). In this way, for example, through physical experiences with a concrete balance, students can revitalize their conception of equality and their physical experiences of how to maintain the balance, and may shift both to a more abstract understanding, which is necessary for linear equation solving at a formal level.

Research in which the balance model was used for teaching linear equation solving indeed has shown that the balance model can be helpful for students. Yet, positive results were not found for all types of balance models. The most positive results have been reported for physical balance models (e.g., Perry et al., 1995; Warren & Cooper, 2005) and for virtual balance models (Figueira-Sampaio et al., 2009; Suh & Moyer, 2007). For drawn balance models, some studies mainly found positive results (e.g., Araya et al., 2010; Cooper & Warren, 2008) while others reported both positive and negative results (e.g., Linchevski & Herscovics, 1996; Vlassis, 2002), or even only negative results (e.g., Ngu & Phan, 2016).

A characteristic of physical balance models is that they are dynamic, which means that students can operate on them and get real-time feedback on their actions. This, in turn, can help students develop strategies for keeping the scale in balance. However, studies in which a physical balance model was used (e.g., Perry et al., 1995) mostly showed only what equations students learned to solve and scarcely revealed how their algebraic reasoning for equation solving developed. The same is true for the algebraic strategies that were used in connection with their working with the physical model. The present study aims to contribute to knowing more about what an embodied learning environment with a physical balance will bring about in students' reasoning.

1.3 Current study

The study made use of a six-lesson teaching sequence on linear equation solving, based on an embodied cognition theory approach and the RME principle of starting with a meaningful context that can evolve into a model for eventually solving a system of two symbolically presented linear equations with two unknowns. In the teaching sequence, algebraic reasoning was elicited by a number of activities in which students were offered embodied experiences with a physical *hanging mobile*: a concrete balance model with, on each side, a number of bags hanging on a chain to represent an equation with unknowns. Bags with different colors were used, which each had a different weight (e.g., white = 50 grams, red = 100 grams, black = 150 grams). These weights were unknown to the students. Students were challenged to discover ways to maintain the *balance* of the mobile while manipulating the differently colored bags.

We theorized that when students worked with the hanging mobile trying to keep its balance, the pervasive everyday-life experiences of balancing (Gibbs Jr, 2006) could be revitalized and linked to the experiences of maintaining the balance of the mobile, representing equality in a linear equation. In this way, students' understanding of the concept of equality could be grounded in the bodily based experiences of maintaining balance, which is also in line with Piaget's statement that children "can only "abstract" the idea of such a relation as equality on the basis of an action of equalization" (Piaget & Inhelder, 1967, p. 43). The possible physical manipulations that maintain the balance of the hanging mobile could act as a metaphorical mapping (Lakoff & Johnson, 1980; Núñez et al., 1999) of the algebraic strategies that can be used to maintain equality in equations. In particular, consider that students' manipulations of the bags can result in the hanging mobile being in or out of balance. Students, for example, can take away similar bags from both sides of the mobile. By doing this, they feel the similar weights of the bags in their hands and concurrently perceive the beam of the hanging mobile remaining in balance. Thus, students physically experience the principle of "doing the same on both sides", which is an important strategy for solving equations (e.g., Arcavi et al., 2016). When working with the hanging mobile, students directly perceive the influence of their actions — such as removing bags — on the status of the balance. In line with embodied cognition theory, we hypothesized that these perceptual-motor experiences could promote the grounding of the concept of equality in the bodily based experience of maintaining balance.

Through maintaining the balance of the hanging mobile, students could intuitively apply informal context-connected algebraic strategies which underlie the conventional strategies for solving equations, such as changing the order of the colored bags (*restructuring*), taking away bags from both sides of the mobile to isolate bags of one color (*simplification/isolation*), and replacing bags with bags of other colors (*substitution*). We hypothesized that through exploring what could be done while keeping the hanging mobile in balance, combined with a number of challenging tasks, students could come more or less spontaneously to the strategies that form the basis of the key algebraic strategies necessary for solving equations. This approach, in which the students have a main role in learning algebraic reasoning, also reflects the didactics of RME, namely that instead of the teacher transferring the strategies to the students, they are active participants in developing these algebraic strategies.

In this study, our aim was to put this teaching sequence to the test and investigate how students' algebraic understanding and reasoning develops, and whether characteristics of their algebraic understanding and reasoning are related to what they are offered in the teaching sequence. More precisely, this resulted in the following two research questions:

(1) *How does fifth-grade students' algebraic reasoning for solving linear equations develop in an embodied learning environment?*
(2) *How are students' experiences in the learning environment related to their use of algebraic strategies?*

Over the teaching sequence, we expected students to improve their algebraic reasoning for solving linear equations. Moreover, we expected students' perceptual-motor experiences with the hanging mobile to be beneficial when solving equations in the context of the hanging mobile, in new informal contexts and when they are formally presented. We therefore expected to find references to (experiences with) the hanging mobile when analyzing students' reasoning.

2. Method

2.1 Participants

To answer these research questions, a study was carried out involving 69 students of three fifth-grade classes in three schools in the Netherlands. Two of the schools, a public one and a Catholic one, were situated in urbanized areas, and the third school was a public school situated in a suburbanized area. The study was approved by the ethical committee of the Faculty of Social and Behavioural Sciences of Utrecht University. Active informed consent was provided by the parents of the students. Permission to use the data was obtained for all students except three. These students were excluded from the analyses. One other student was excluded from the analyses because he missed most of the lessons. This resulted in the final number of 65 students: 27 boys (42%) and 38 girls (58%), with ages between 9 and 11 years old ($M = 10.02$, $SD = 0.45$). They had received no prior instruction on equation solving or other algebra topics, which is in accordance with the Dutch mathematics curriculum, in which teaching algebra starts in the first year of secondary school.

2.2 Materials

2.2.1 Teaching sequence

The teaching sequence (see Figure 1) consisted of six lessons. These lessons were clustered in four episodes, each with their own focus and content. In each of these episodes, the aim was to develop algebraic strategies related to linear equation solving.

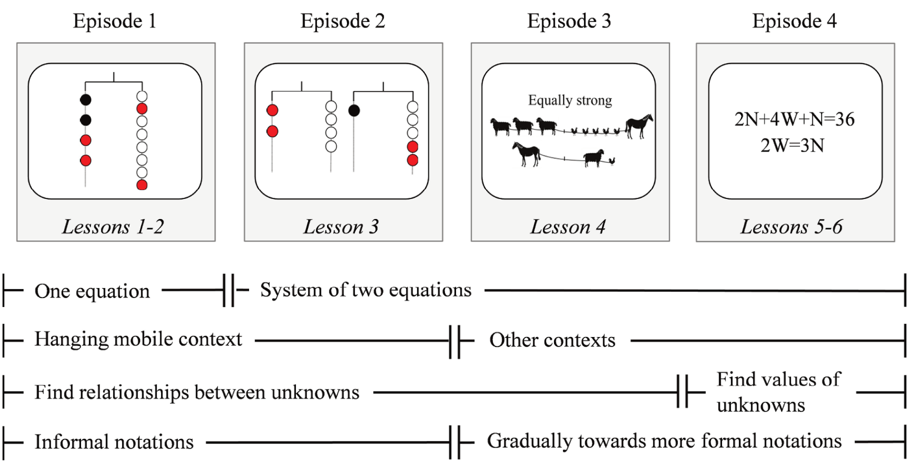


Figure 1. The teaching sequence and its content.

In Episode 1, students gained physical experiences with a physical hanging mobile. The tilting beam of the mobile could be in or out of balance, thus providing students real-time feedback on their actions while manipulating the bags on the mobile.

While students were working with the hanging mobile, the teacher asked questions like: “Can you explain why the beam of your hanging mobile is *straight*?”; “Can you explain why the hanging mobile remains straight after the manipulation you just carried out?”; or “Would it be possible to remove something from your hanging mobile in such a way that it remains straight?” During the ensuing classroom discussion, one hanging mobile was positioned in front of the classroom. The students were asked to mention all possible actions they could perform on the hanging mobile while keeping it straight. These actions were carried out on the mobile, and the teacher registered the possibilities on the blackboard, making use of the students’ own wording (e.g., “one color of bags could be changed for another color of bags” instead of “substitution”). Because students worked with one hanging mobile in this first episode, they were reasoning about one equation.

In the following episodes, the information from two equations had to be combined to discover relationships between unknowns to solve the problems, so that students were reasoning about a system of equations. In Episode 2, problems were still posed in the context of the hanging mobile, both physically and on paper. The teacher started by asking whether it would be possible to combine the information from the *two* hanging mobiles in front of the classroom, to create a new mobile. After gathering students’ thoughts, the teacher invited a student to manipulate the bags on the two existing mobiles to find a new relationship between unknowns and display that on a third mobile. Then, students worked on paper-based tasks in which the information from two mobiles had to be combined.

In Episode 3, the information from two equations had to be combined to discover relationships between unknowns in new contexts, such as a tug-of-war situation. In this episode, the teacher started with a classroom discussion about the similarity between the new context of the tug-of-war situation and the familiar context of the hanging mobile, namely, that in both contexts it is all about maintaining equality. Then, all possible actions to maintain the balance of the hanging mobile were recalled, before students started with solving problems in the new tug-of-war context. Moreover, in this episode students were gradually challenged to use more symbolic notations, such as letters to notate the strength of the animals participating in the tug-of-war game. This was evoked by asking the students to think about alternative ways to notate their explanations, as such avoiding the time-consuming process of drawing the animals or writing down the entire reasoning process.

In the final episode, Episode 4, instead of discovering relationships between unknowns, students had to find the values of unknowns in a system of two linear equations with symbolically notated unknowns. As an introduction, the teacher again discussed the analogy between the hanging mobile and these new, symbolically notated equations. Moreover, she discussed the meaning of the symbolic notations, leading, for example, to discussions about whether "2*N*" would be the same as "$N + N$" or whether "1*N*" would be an alternative way to notate "*N*".

During our teaching sequence students encountered many different equations in which different quantities were used. For example, when students were working with the hanging mobile (Episodes 1 and 2), they were working with the quantity of weight (e.g., three white bags are as heavy as one black bag). In the case of the tug-of-war situations (Episode 3), they were working with the quantity of strength (e.g., one sheep is as strong as three chickens). In Episode 4, they were working with objects and prices (one apple and one banana together cost eight euros), and they ended up with formally notated equations with an equal sign and with unknowns presented as letters ($M + 3L = 25$ and $M = 2L$), which were not related to a given quantity such as weight, strength and cost per object.

2.2.2 Assessment of algebraic reasoning

Students' level of algebraic reasoning related to linear equation solving was assessed by means of paper-and-pencil assessment tasks administered at the end of each lesson (see Appendix 3.1, Figures A1-A3, for examples of the assessment tasks of Episodes 2-4). Each assessment task reflected the goal of the corresponding lesson. The assessment task of Episode 2 contained two balanced hanging mobiles including circles, rectangles and stars, informally representing the formal algebraic equations $3X = Y + 5Z$ and $Y = X + Z$. On the basis of these mobiles, the students had to determine for two other mobiles ($X = 3Z$ and $4Z = Y$) whether they were balanced. In the assessment task of Episode 3, the same equations were used, but with strawberries, bananas and pears, and with an equal sign instead of a hanging mobile. Lastly, in the fourth episode, students had to solve a system of two symbolically presented equations. In this assessment task, they had to discover the values of unknowns M and L by using the equations $M + 3L = 25$ and $2M = 4L$.

For all tasks, in addition to giving the answer to the problem, students were asked to explain their thinking and thus to reveal their reasoning. Answers were scored as incorrect (0) or correct (1), and this dichotomous scoring was used to calculate the success rate — that is, the percentage correct per assessment task. Students' explanations were categorized on their level of reasoning by means of a coding scheme (see section 2.4 Data analysis).

2.3 Procedure

The 50-minute lessons were each taught in six consecutive weeks by the first author of this article, with the assistance of a teaching assistant during the two lessons of the first episode. Episode 1 took place in a separate room in which five hanging mobiles were positioned. In this episode, students worked together in small groups of two or three students. The remaining episodes took place in the students' classroom. During each episode, at least one physical hanging mobile was present in the classroom, even in the case that it was not used for instruction. After the first episode, students worked individually, in pairs or in small groups. Individual or group work alternated with whole-class discussions. In between the lessons the students followed their regular program.

In the first episode, in one class, two pairs of students were video-recorded, so that their manipulations on the physical hanging mobile and their reasoning could be followed more closely. These particular students were recorded because their parents provided permission for this. During the rest of the episodes in this class, the camera was placed in the back of the classroom to give a global overview of the lesson activities. For the other two classes, two pairs of students were video-recorded in the first episode, and also in the following episodes the video-recording of the work and conversations of these students was continued. As such, we could follow these students' reasoning over time. In addition to the focus on these two pairs of students, from Episode 2 an extra camera was placed in the back of the classrooms to give a global overview of the lesson activities in these two classes as well.

2.4 Data analysis

The data consisted of students' responses to the problems to be solved during the lessons and to the assessment tasks, and audio- and video-recorded interactions. To analyze how students' algebraic reasoning developed over time (Research Question 1), we focused on the three consecutive episodes in which students solved systems of linear equations (Episodes 2-4). For each episode, the assessment task that was given to the students at the end of that episode was analyzed (see Appendix 3.1, Figures A1-A3, for examples).

For calculating the success rate on each of these assessment tasks, a student's response was scored as correct when *both* problems of the task were answered correctly. The task belonging to Episode 4 was scored as correct when students correctly reported the value of *both* unknowns.

To categorize students' explanations on the assessment tasks, a coding scheme was developed that was applicable to each of the tasks. The development of this coding scheme followed an iterative process, inspired by the constant comparative method (Glaser, 1965). We commenced with a coding scheme on the basis of the work of only a couple of students. This coding scheme was then taken as a starting point to code the remainder of the data. When examples of student work were encountered that did not fit into the existing scheme, the scheme was adapted. As such, the final coding scheme was established after several rounds of careful examination of the data and the reaching of consensus between the researchers. Subsequently, there was a final round of coding in which all data were recoded on the basis of the final coding scheme.

In each assessment task in Episodes 2-4, students had to solve the problem by reasoning on the basis of a given system of two equations. In the final coding scheme, we distinguished between students who did not use any of these equations in the description of their reasoning (Level R0), students who reasoned on the basis of only one of the two given equations (Level R1) and students who reasoned on the basis of both given equations by combining the information of both of them (Level R2). Within the Levels R1 and R2, we furthermore distinguished between (1) students who did not make clear how they came to their answer or which strategies they used, and (2) students who did give this information. The coding scheme, with examples of students' responses for each assessment task and each level of reasoning, can be found in Appendix 3.2.

Additionally, for each episode, the association between students' level of reasoning and their success rate was evaluated by means of a chi-square test of independence.

To analyze the possible relation between the affordances of the embodied learning environment and students' development of algebraic reasoning (Research Question 2), we looked into the algebraic strategies elicited in each episode by analyzing students' written work and the video- and audio-recorded interactions, and searched for events in which students made use of algebraic strategies when solving the problems in the classroom. In this way, we could relate the students' strategy use to what happened in the lessons during that particular episode.

3. Results

3.1 Students' development of algebraic reasoning in an embodied learning environment

In the course of the teaching sequence, students more often provided the correct answer as shown by the increase in the success rates on the respective assessment tasks, from 44% in Episode 2, through 57% in Episode 3, to 77% in Episode 4. Furthermore, students' level of algebraic reasoning increased over time, as shown in Figure 2. The percentage of students showing reasoning without using the given equations (Level R0) decreased, from 57% in Episode 2 to 20% in Episode 4. Students who showed this type of reasoning provided, for example, the explanation that they "just knew the answer", provided a general description that they "looked at the example" without stating explicitly which information was used to come to that answer, or did not write down anything. The percentage of students that showed reasoning reflecting the use of only one of the given equations (Level R1) remained more or less stable, between 23% and 28%. Julia's answer on the assessment task of Episode 2 (see Appendix 3.1, Figure A1) is an example of reasoning which was categorized as Level R1. She assigned values to the unknowns on the basis of the second given equation alone, while ignoring the information from the first equation. Finally, the percentage of students that showed reasoning in which both given equations were used (Level R2) increased over time, from 17% in Episode 2 to 52% in Episode 4. Julia's reasoning on the assessment tasks of Episodes 3 and 4 (see Appendix 3.1, Figures A2 and A3) are two examples of this level of reasoning. For example, on the assessment task of Episode 4 (see Appendix 3.1, Figure A3), she first transformed the equation $2M = 4L$ into $M = 2L$. Subsequently, she used this information to substitute the unknown M in the other equation by $2L$. In this way, she combined the information of both given equations to come to an answer, which belongs to reasoning Level R2.

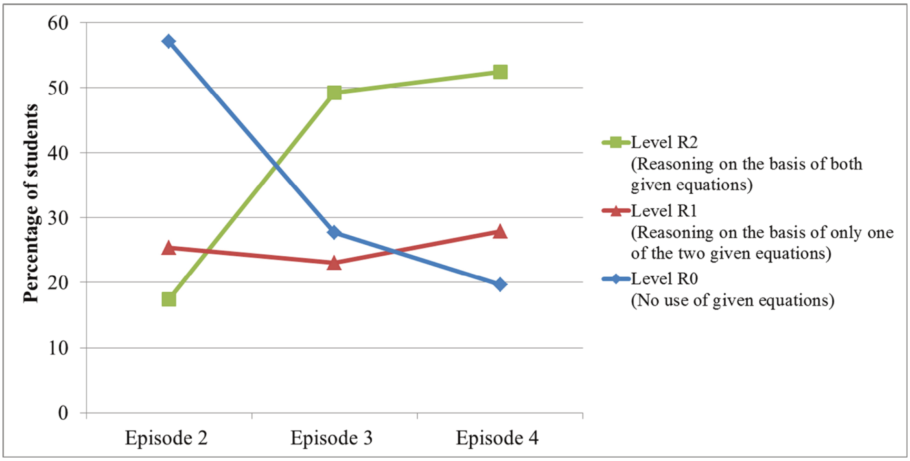


Figure 2. Percentages of students performing at a level of reasoning over the three episodes ($N_{\text{Episode 2}} = 63$; $N_{\text{Episode 3}} = 65$; $N_{\text{Episode 4}} = 61$).

The association between students' level of reasoning and their success rate differed between the episodes. In Episode 2 (χ^2 (2, $n = 63) = 2.37$, $p = .306$) and Episode 3 (χ^2 (2, $n = 65) = 5.97$, $p = .051$), providing a correct answer was unrelated to students' demonstrated level of reasoning, while in Episode 4 (χ^2 (2, $n = 61) = 8.70$ $p = .013$), a high level of reasoning occurred more often with a correct answer and a low level of reasoning with an incorrect answer.

Additionally, the development in level of reasoning was analyzed for each student individually. One student was excluded from this analysis, because she was absent for two of these three assessment tasks. In total, 31 students (48%) showed a pattern of reasoning which improved at least one level over time, while no decline was shown. The level of reasoning of four students (6%) only decreased. Fourteen students (22%) showed reasoning that remained at the same level over time. Finally, the remaining 15 students (24%) showed a fluctuating reasoning pattern, which over time was both increasing and decreasing.

In addition, we also found that there was an increase in the number of students who were able to write how they derived their answer and/or show their applied strategies when solving the problems. This was especially the case for the students in Level R2. Whereas in Episode 2, 16% of the 63 students who belonged to Level R2 gave a description of their reasoning, this percentage increased to 42% for the 61 students involved in Episode 4. For the students performing at Level R1, these percentages ranged from 17% in Episode 2 to 12% in Episode 4.

3.2 Relation between the students' experiences in the learning environment and their use of algebraic strategies

3.2.1 Episode 1: physical experiences with the hanging mobile

In the first episode, all students started with the same, balanced, physical hanging mobile (see Figure 1, Episode 1). Students were not provided with information about how the bags were related. In the first task, students were asked to change the hanging mobile in such a way that it would look different, but would still be hanging straight, while using the same bags. This led students to use different restructuring strategies. For example, Eva commented, as follows, to her partner Jailey:

> **Eva:** We can do exactly the same, but then on the other side. So these bags [1]
> [points to the bags on the left side] to the other side, and these ones [points to the [2]
> bags on the right side] also all to the other side. So all these bags to that side and [3]
> all these bags to the other side as well. Then it is different, but also the same [4]
> [laughs]! [5]

Another restructuring strategy was shown by Iris and Zara. They changed the order of the bags on the right side of the hanging mobile by grouping them by color. Kees and Julia also restructured their hanging mobile, but instead of exchanging all the bags from one side to the other or grouping them by color, they brought three white bags to the left side of the mobile and in exchange one black to the right side. The restructuring strategies identified by the three pairs of students already in their first attempt working with the hanging mobile correspond to important algebraic strategies. Eva and Jailey's strategy of exchanging all the bags between the sides of the hanging mobile corresponds to the symmetry property of equations ($a = b \Leftrightarrow b = a$). The strategy of Iris and Zara, who changed the order of the bags on one side of the hanging mobile, reflects the commutative law of addition. Kees and Julia discovered how the white and the black bags are related and applied already a substitution strategy in restructuring the composition on their hanging mobile. This last way of restructuring was actually shown by most groups of students when working on this task.

After a short group discussion about the resulting compositions on the hanging mobiles after the transformations, students were tasked to find *everything* that could be done while keeping the hanging mobile straight, and this time they were also allowed to add bags and take away bags. The following conversation between Kees and Julia exemplifies the emergence of their use and understanding of the isolation strategy, by first removing identical unknowns from both sides of the equation

(Lines 6-7) and then removing different unknowns from both sides based on the ratio (Lines 9-10). Furthermore, in the second part of the interaction in their discussion with the teacher (T), their use of the substitution strategy is exemplified, as they describe it in words (Lines 13-15) and illustrate it by physically substituting bags on the hanging mobile (Lines 24-25).

Kees: These two [points to the two red bags on the left side] are the same as these two [points to the two red bags on the right side]. They are equally heavy …
Julia: … so then it must be that these two blacks are the same as 1, 2, 3, 4, 5 … 6 [white bags] [counts the white bags on the right side of the hanging mobile]! So then one [black] equals three [whites]
Kees: Yes!

A little later, they had created a balanced hanging mobile with, on the left side, two black bags and, on the right side, two white and two red bags.

T: Can you tell me why this hanging mobile is straight?
Kees: Because this one [points to one red bag] counts for two whites [a ratio they had previously discovered]. And three whites together equal one black. So one red and one white together must be one black.
T: [looks at Julia] Do you agree?
Julia: Yes!
T: So, you actually say, if I would replace this one [points to one red bag] with two whites …
Kees: … it's exactly the same!
T: Yes? Are you sure?
Kees: [nods]
T: Let's try!
Kees: [replaces one red bag with two whites and watches the mobile coming back into balance, see Figure 3] Look!

Figure 3. Kees and Julia substitute a red bag for two whites and watch the physical hanging mobile coming into balance (Episode 1).

In these activities, maintaining the balance of the mobile represents maintaining equality in an equation. The interactions clearly illustrate that students used different algebraic strategies, with which they were gaining perceptual-motor experiences. By restructuring the mobile through exchanging all the bags from one side to the other, the students could physically experience that both sides of the equation are interchangeable and that changing the order of the bags on a side does not disturb the balance. Moreover, by taking away similar bags on both sides, the students could physically experience the possibility of cancelling identical terms from both sides, which can be used to isolate particular unknowns. The third strategy, the substitution strategy, allowed the students to further simplify the composition on the hanging mobile, thus bringing them closer to finding all the ratios. By means of the real-time feedback the tilting beam of the hanging mobile provided, students were constantly able to verify, and if necessary adjust, their reasoning, which was, for example, shown by Kees and Julia when replacing the red bag with two white ones (Lines 24-25).

After students worked with the physical hanging mobile in small groups, in the classroom discussions students shared experiences. During these discussions it came to the fore that most groups of students shared the experience of maintaining balance by using similar strategies.

3.2.2 Episode 2: combining the information from two hanging mobiles

Next, the students were given a task in which they had to combine the information from two hanging mobiles, with the goal of fostering their further development of algebraic strategies, in particular the substitution strategy. The teacher introduced the task by placing three physical hanging mobiles in front of the classroom (see Figure 4).

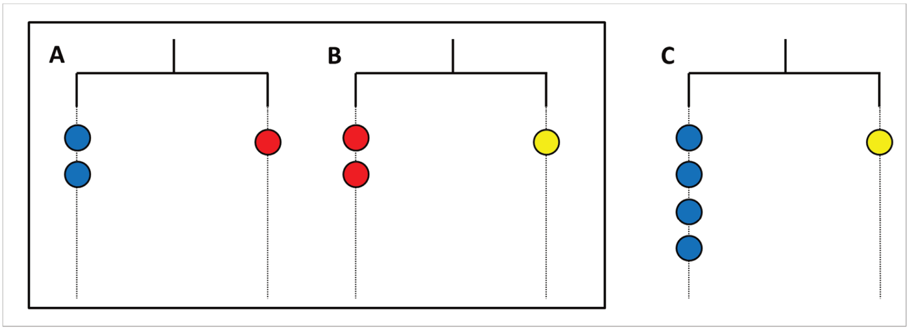
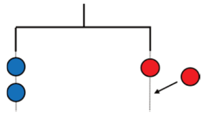


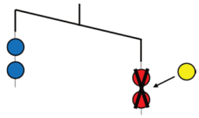
Figure 4. Three hanging mobiles in front of the classroom (Episode 2).

In one classroom, Florien raised her hand when the teacher asked whether it would be possible to combine the two hanging mobiles (A and B) to create the third one (C).

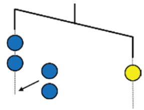
Florien: If you double the red one, [26]



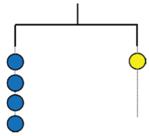
then it becomes one yellow [points to the hanging mobile A]. [27]



And if you then double the blues as well, it results in four [28]



blues .So then you know that four blues equal one yellow. [29]



Florien's strategy was one of the strategies that also came up in the other classes. She prepared her use of the substitution strategy by transforming the first equation (A) by doubling. This enabled her to substitute one side (two red bags) of the equation as a whole, by making use of the information from the second equation (B). After explaining her strategy to the rest of the class, she made it visible to the other students on the physical mobiles in front of the classroom. A second example of substitution was put forward, in which students, based on the information of the first hanging mobile (A), replaced one red bag in the second equation (B) with two blue bags. So here, instead of substituting one side of the equation at once, one unknown (or element) in the equation was substituted for another unknown. Most students used this strategy. Finally, instead of replacing unknowns with other unknowns, a few students also used substitution of unknowns by values, which was exemplified by Mats. He assigned the value of 1 to each blue unknown, the value of 2 to the reds and the value of 4 to the yellow unknowns, which resulted in both sides of each equation adding up to the same number. Although Mats did not explain how he came to these values, his choice for the particular values evidenced that he took all equations into account.

3.2.3 Episode 3: a new context

In this episode, the students were presented with problems that were far away from the problems with the hanging mobile, but nevertheless had some similarity with them. Two groups of animals played a tug-of-war game, and the question was which group would win (see Figure 5). Also, here the focus was on the equality of both sides and figuring out which group was stronger by applying all kind of transformations while keeping the balance.

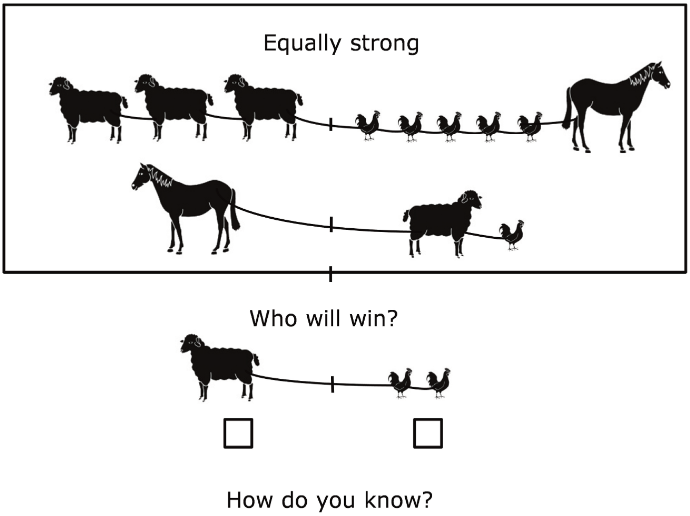


Figure 5. Tug-of-war problem (Episode 3).

Milan explained his strategy for solving this problem to the rest of the class. First, he combined both equations by adding them, as such creating one new equation (Line 30). Subsequently, he applied isolation strategies to solve the equation, by first taking away a horse on both sides (Lines 30-31), then a sheep on both sides (Lines 31-32), and then by removing different unknowns based on the ratio by dividing both sides of the equation by two (Line 33).

> **Milan**: I added the lower row to the upper one. Then I was able to take away both horses. Then on the right side, six chickens and one sheep remained. So, you could take away a sheep at both [sides]. Then there are six [chickens] and two sheep left. Dividing makes three [chickens].

Julia used another strategy for solving this problem. She first converted the problem into two equations with letters, by writing $S + S + S \mid C + C + C + C + C + H$ and $H \mid S + C$. Then, she substituted the H in the first equation for a C and an S, resulting in $S + S + S \mid C + C + C + C + C + C + S$. Subsequently, to isolate the chickens, she crossed out the S on both sides, and lastly she crossed one S on the left side and three Cs on the right side, leaving $S \mid C + C + C$. She rewrote this as

“1$S = CCC$3”. A third strategy that again some students used to solve the problem was assigning values to the unknowns, which was already described in Episode 2.

In this episode, where the students were presented with equation-like problems in a new context, again different strategies for combining the information from both equations came forward. Milan (Lines 30-33), for example, added both equations to create a new equation with some identical unknowns on both sides that he could take away, to subsequently isolate one unknown and express it in another. However, Milan was one of the few students that used this smart way of combining both equations to constitute a new one. Most students used the substitution strategy to solve this problem. This strategy was also used a lot while working with the physical hanging mobile. In Episode 3, this strategy was, for example, shown by Julia, who converted one type of unknown (the horse) to two other types (a sheep and a chicken). She also used this process of substitution of one type of unknown into other types when working with the physical hanging mobile in Episode 1 (Lines 9-10).

The strategies that came up in this episode to solve the new type of equation-like context problems are equivalent to the strategies the students had developed while working on the hanging mobile. When explaining their answers, some students explicitly referred to their experiences of working with the hanging mobile, for example, through expressions such as “the horse can be seen as a green bag”. In one classroom, when the teacher did not pay attention for a short time, the students even used the colored bags of the hanging mobile, which were not used at that moment, to represent the tug-of-war situation on their table. Additionally, students often explained their solving strategies in terms of actions, such as “I replace the horse by a sheep and a chicken, then I take away a sheep on both sides, and then I take half of both sides”. The combination of students using similar algebraic strategies, explicitly referring to the hanging mobile context, and providing explanations on solving the problem in terms of actions suggests that students’ experiences of working on the hanging mobile might have formed a foundation for this later stage of solving linear equations in a new context.

3.2.4 Episode 4: solving a system of symbolic equations

In the last episode of the teaching sequence, students were presented with a system of two symbolic linear equations with two unknowns. As discernible in Figure 6, Kees (see also Episode 1) could solve this problem by showing that $F = 6$ and $H = 12$. He first converted the equations into hanging mobiles, notating the unknowns in an expanded way. Then, by connecting the *F*s on both sides of the hanging mobile, he showed the presence of identical terms on both sides of the equation. Moreover, he crossed these unknowns to reveal the process of cancellation, resulting in the isolation of one *F* on the right side of the equation. Subsequently he could derive the value of this unknown by solving 19 - __ = 13. Lastly, by substituting the value 6 for the unknown *F* in the second hanging mobile (representing the second equation), he was able to calculate the value of *H*.

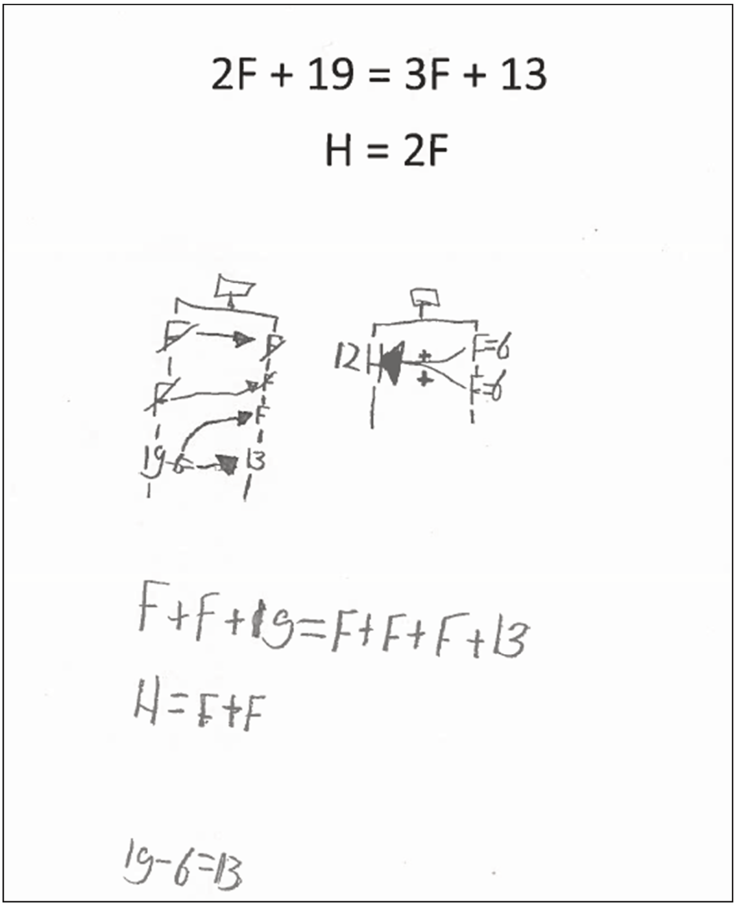


Figure 6. Kees' solution of a system of two symbolic linear equations, combined with the answer that *F* = 6 and *H* = 12 (Episode 4).

Like Kees, many students first converted the equations into a hanging mobile before showing cancellation of identical terms. Some other students reasoned directly on the basis of the symbolically notated equation. For example, Kira first repeated the first symbolic equation, and then wrote down "1*F* is added" implying that she noticed that in the first equation on the left side there are two *F*s, whereas on the right side there are three. She then reasoned that she could do the subtraction "19 - 13 = 6" to arrive at the value of "$1F = 6$". Finally, she used this finding to determine the value of "*H*: $2F = 12$".

In this last episode, it was revealed that most of these students, who were taught linear equation solving in an embodied learning environment, were able to solve systems of two symbolic linear equations. Most students did not seem to have difficulties with understanding the symbolic way of notating the unknowns. Kees, for example, showed understanding that $2F$ is an abbreviated way of notating $F + F$. Nevertheless, he did not solve this problem in a symbolic, formal way. Instead, before applying the isolation strategy and then the strategy of substituting of values, he first went back to the model of the hanging mobile. This is an example showing that at least some students experienced the model of the hanging mobile as a powerful tool that they applied when solving linear equations at a formal level. Some other students, like Kira, reasoned directly on the basis of the symbolic equations.

Kees and Kira, for example, are students who, like most students in our study, did not seem to have difficulties with equations not written in the standard form (e.g., $a + b = c$), such as equations containing unknowns on both sides of the equation (e.g., $2F + 19 = 3F + 13$ or $H = 2F$). The students did not seem to have misconceptions about the equal sign, meaning that instead of interpreting the equal sign as a symbol for "here comes the answer", they considered it as a symbol reflecting the equality of the expressions on both sides. The tasks with the hanging mobile, in which our main focus was to let students develop strategies related to maintaining the equality in equations and as such to provide them with a basis for solving equations later on, seem to have contributed to their understanding of the equal sign, even in formally notated equations.

4. Discussion

In this study, we investigated the development of students' algebraic reasoning in an embodied learning environment and the possible relation between the characteristics of the learning environment and the applied algebraic strategies. In the three consecutive episodes in which students worked on problems involving systems of two linear equations, they clearly showed progress in level of reasoning and success rate. At first more than half of the 65 students reasoned without making use of the information in the given equations. Over the teaching sequence, this decreased to about one-fifth of the students. Similarly, whereas in the beginning less than one-fifth of the students showed the highest level of reasoning, in which they combined the information from both given equations, after the lessons more than half of the students attained this level of reasoning. The individual learning progresses confirmed these findings. The reasoning of most students improved over time without any decline. For 14 students no change was found, 15 students showed a fluctuating pattern, and only four students decreased in their level of reasoning. Further investigation of the students' written work on the assessment tasks revealed that they also improved in their ability to explain their reasoning and their applied strategies to solve these tasks.

Many students appeared to be reticent in writing down their reasoning. In each class, during the teaching sequence, we observed fluctuations regarding students' willingness to write down their reasoning. This is a possible explanation for the declining or fluctuating reasoning pattern of some of the students. Another explanation concerns the nature of the assessment task of Episode 4. Whereas in the assessment tasks of Episodes 2 and 3 students had to find relationships between unknowns, in the assessment task of Episode 4 they had to identify the values of unknowns. It could be that asking students to identify the values of unknowns (compared to asking them about relationships between unknowns) made it seemingly easier for them to find a correct answer, while at the same time they did not write down their reasoning. Because our categorization of students' levels of reasoning was based on their written explanations, we had to assign them to Level R0 ("student does not use any of the given equations") if they did not write anything down. Of the 19 students that showed a fluctuating or declining pattern of reasoning, 11 declined in their level of reasoning from Episode 3 to Episode 4, while 10 of these 11 students answered the assessment task of Episode 4 correctly. This seems to support the latter explanation.

We found some indications for the existence of a relation between the characteristics of the embodied learning environment and the students' algebraic reasoning. Particular ways of reasoning, uses of strategies and forms of notations could be related to the students' (physical) experiences of working with this hanging mobile. The influence of these experiences seemed to be present when the problems were posed in different contexts, and students often used similar strategies to solve the problems as when they were working with the hanging mobile. Thus, the acquired strategies appeared to form a base for solving new problems. This came even more to the fore when students used restructuring, isolation and substitution strategies to solve systems of two symbolic linear equations. The hanging mobile seemed to evolve into a model on which students could rely for solving equations at a formal level. This was exemplified by Kees' reasoning in his solution of the assessment task in Episode 4 (see Figure 6), in which he clearly made use of the model of the hanging mobile to solve a system of two formally notated linear equations.

In line with embodied cognition theory (e.g., Wilson, 2002), we assumed that students' perceptual-motor experiences with the hanging mobile would be beneficial for their linear equation solving abilities. During the teaching sequence, students could directly perceive how the status of the balance — representing equality in a linear equation — was influenced by their manipulations of the bags on the hanging mobile. This action-perception coupling was expected to ground students' understanding of equality, a crucial concept in linear equation solving (e.g., Bush & Karp, 2013; Knuth et al., 2005; Li et al., 2008), in the bodily based experiences of maintaining balance. The result of such a possible coupling appeared to be exemplified by students' reasoning in Episode 3, in which students used the algebraic strategies and provided explanations in terms of actions that could be related to their working on the hanging mobile. However, the design of our study precludes us from truly deciding about the veracity of our assumptions about the working mechanisms governing students' development of algebraic reasoning in interaction with the physical hanging mobile.

In addition to these findings, it was remarkable that most of our students, when they had to solve linear equations, did not seem to show difficulties that are often reported in other studies. For example, students did not show a lack of relational understanding of the equal sign (e.g., Behr et al., 1980; Carpenter et al., 2003). Students showed this relational understanding in the assessment tasks of Episodes 3 and 4 by correctly interpreting the equal sign as a signal for equality instead of as a "to-do" signal. This is striking, because based on the mathematics curriculum, students only have experience with the equal sign in tasks in the standard form (e.g., $a + b = c$). Tasks in a non-standard form (e.g., $c = a + b$ or $a + b = c + d$) that could promote a relational interpretation of the equal sign are scarcely addressed in the curriculum. Obviously, our approach with the hanging mobile, in which the focus was on developing strategies in order to maintaining the equality in equations, helped the students in acquiring a relational understanding of the equal sign.

Furthermore, the students in our study also did not show difficulties with the formal, symbolic way of notating equations (e.g., Koedinger & Nathan, 2004), nor with operations on unknowns (e.g., Filloy & Rojano, 1989; Herscovics & Linchevski, 1994). Being proficient at solving this type of problem means that the students do not only have to understand the given equations, but also have to be able to manipulate and combine them to use this newly gathered information about the unknowns in solving the problems. For students at the primary school age, who have never been taught linear equation solving, this requires significant higher-order thinking. In this way, the findings of our study plainly support the generally agreed ambition (e.g., National Council of Teachers of Mathematics [NCTM], 2000) to lay the foundation for higher-order thinking already in primary education.

In interpreting the results of our study it is important to take several limitations into consideration. First of all, the exploratory nature of this study enabled us to explore the development of algebraic reasoning of the students in our study. However, there are some limitations associated with this design, such as the lack of a control group, that imply the results should be interpreted carefully. Additional research with an experimental study is therefore necessary to further evaluate the influence of experiences in an embodied learning environment, such as the hanging mobile, on algebraic reasoning.

A second issue concerns how we measured students' reasoning. Our classification of students' level of reasoning was based on written answers and not on think-aloud protocols. This means that we might not have fully captured their reasoning. As such, their level of reasoning may have been underestimated. This is also suggested by the observation that some students were very well able to verbally explain their reasoning during the lessons but did not write their thoughts with the same level of sophistication on the assessment tasks. A further aspect to consider is the repeated assessment of solving linear equations, which could have given students more opportunities to learn how to solve these problems. Therefore, it is important to acknowledge that, in addition to the effect of the experiences with the hanging mobile, the mere use of the assessment tasks could have contributed to students' progress in their level of reasoning. Moreover, having assessment tasks that directly followed after the teaching activities and that focused on the same content as addressed in the lessons could have influenced the students' results on the tasks. A delayed assessment might have provided a better view on what the students gained from the lessons.

A next issue of concern is that although our study was carried out in a regular classroom setting, it was the researcher who taught the lessons. This means that one might wonder whether this kind of algebra instruction with concrete apparatus is feasible to be carried out in regular educational practice. In addition one might wonder to what extent primary school classroom teachers are prepared for, and feel confident in, teaching such lessons.

A question that we also did not answer is whether the teaching of linear equation solving by means of the hanging mobile can be implemented in the longer trajectory of teaching algebra. In our study, the topic of linear equation solving was also rather shortly dealt with and consisted only of teaching equations with positive numbers and addition. Additional research is necessary to investigate how the support by the hanging mobile works out for other types of equations including negative numbers or subtraction that cannot be modelled by such a concrete model. As other authors have emphasized (e.g., Fyfe et al., 2014), a process of concreteness fading might be necessary, beginning with visual representations and fading into abstract symbols.

Notwithstanding the aforementioned limitations, our findings illustrate that fifth-grade students can successfully apply algebraic strategies such as restructuring, isolation and substitution when solving linear equations. Thus, the embodied learning environment in our study seems to have laid a basis for mathematical reasoning in later stages when solving systems of symbolically presented linear equations.

References

Abrahamson, D. (2017). Embodiment and mathematics learning. In K. Peppler (Ed.), *The SAGE encyclopedia of out-of-school learning* (pp. 247–252). SAGE Publications, Inc.

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal look at middle-school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning, 9*(3), 221–247. https://doi.org/10.1080/10986060701360902

Araya, R., Calfucura, P., Jiménez, A., Aguirre, C., Palavicino, M. A., Lacourly, N., Soto-Andrade, J., & Dartnell, P. (2010). The effect of analogies on learning to solve algebraic equations. *Pedagogies: An International Journal*, *5*(3), 216–232. https://doi.org/10.1080/1554480X.2010.486160

Arcavi, A., Drijvers, P., & Stacey, K. (2016). *The learning and teaching of algebra: Ideas, insights and activities*. Routledge.

Behr, M., Erlwanger, S., & Nichols, E. (1980). How children view the equals sign. *Mathematics Teaching, 92*(1), 13–15.

Blanton, M., Schifter, D., Inge, V., Lofgren, P., Willis, C., Davis, F., & Confrey, J. (2007). Early algebra. In V. J. Katz (Ed.), *Algebra: Gateway to a technological future* (pp. 7–15). The Mathematical Association of America.

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J. S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education*, *46*(1), 39–87. https://doi.org/10.5951/jresematheduc.46.1.0039

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

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

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school.* Heineman.

Carraher, D., & Schliemann, A. D. (2014). Early algebra teaching and learning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 193–196). Springer.

Carraher, D. W., Schliemann, A. D., & Schwartz, J. (2008). Early algebra is not the same as algebra early. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 235–272). Lawrence Erlbaum Associates.

Cooper, T. J., & Warren, E. (2008). The effect of different representations on Years 3 to 5 students' ability to generalise. *ZDM Mathematics Education*, *40*(1), 23–37. https://doi.org/10.1007/s11858-007-0066-8

Figueira-Sampaio, A. S., Santos, E. E. F., & Carrijo, G. A. (2009). A constructivist computational tool to assist in learning primary school mathematical equations. *Computers & Education, 53*(2), 484–492. https://doi.org/10.1016/j.compedu.2009.03.012

Filloy, E., & Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics*, *9*(2), 19–25.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*(1), 9–25. https://doi.org/10.1007/s10648-014-9249-3

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3-4), 455–479. https://doi.org/10.1080/02643290442000310

Gibbs Jr, R. W. (2006). *Embodiment and cognitive science*. Cambridge University Press.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436–445. https://doi.org/10.2307/798843

Herscovics, N., & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational Studies in Mathematics*, *27*(1), 59–78. https://doi.org/10.1007/BF01284528

Kaput, J. J. (2008). What is algebra? What is algebraic reasoning? In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 5–18). Lawrence Erlbaum Associates.

Kaput, J. J., Carraher, D. W., & Blanton, M. L. (2008). *Algebra in the early grades.* Lawrence Erlbaum Associates.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

Kieran, C. (1997). Mathematical concepts at the secondary school level: The learning of algebra and functions. In T. Nunes, & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 133–158). Psychology Press.

Kieran, C., Pang, J., Schifter, D., & Ng, S. F. (2016). *Early algebra: Research into its nature, its learning, its teaching.* Springer (open access eBook). https://doi.org/10.1007/978-3-319-32258-2

Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens, A. C. (2005). Middle school students‘ understanding of core algebraic concepts: Equivalence & variable. *ZDM Mathematics Education*, *37*(1), 68–76. https://doi.org/10.1007/BF02655899

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4)*,* 297–312. https://www.jstor.org/stable/30034852

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, *13*(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1

Lakoff, G., & Johnson, M. (1980). Conceptual metaphor in everyday language. *The Journal of Philosophy, 77*(8), 453–486. https://doi.org/10.2307/2025464

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. Basic books.

Leibniz, G. W. (1989). To queen Sophie Charlotte of Prussia, on what is independent of sense and matter (1702). In R. Ariew & D. Garber (Eds., Trans.), *Philosophical essays* (pp. 186–192). Hackett Publishing Company.

Li, X., Ding, M., Capraro, M. M., & Capraro, R. M. (2008). Sources of differences in children's understandings of mathematical equality: Comparative analysis of teacher guides and student texts in China and the United States. *Cognition and Instruction*, *26*(2), 195–217. https://doi.org/10.1080/07370000801980845

Linchevski, L., & Herscovics, N. (1996). Crossing the cognitive gap between arithmetic and algebra: Operating on the unknown in the context of equations. *Educational Studies in Mathematics*, *30*(1), 39–65. https://doi.org/10.1007/BF00163752

Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education*, *43*(3), 316–350. https://doi.org/10.5951/jresematheduc.43.3.0316

National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics.* NCTM.

Ngu, B. H., & Phan, H. P. (2016). Comparing balance and inverse methods on learning conceptual and procedural knowledge in equation solving: A cognitive load perspective. *Pedagogies: An International Journal*, *11*(1), 63–83. https://doi.org/10.1080/1554480X.2015.1047836

Núñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics*, *39*(1–3), 45–65. https://doi.org/10.1023/A:1003759711966

Otten, M., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2019). The balance model for teaching linear equations: a systematic literature review. *International Journal of STEM Education*, *6*(1), 30–51. https://doi.org/10.1186/s40594-019-0183-2

Perry, M., Berch, D., & Singleton, J. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues*, *6*, 213–235.

Piaget, J., & Inhelder, B. (1967). *The child's conception of space*. W. W. Norton.

Pouw, W. T., Van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51–72.

Stacey, K., & MacGregor, M. (1999). Learning the algebraic method of solving problems. *The Journal of Mathematical Behavior*, *18*(2), 149–167. https://doi.org/10.1016/S0732-3123(99)00026-7

Stephens, A. C., Fonger, N., Strachota, S., Isler, I., Blanton, M., Knuth, E., & Gardiner, A. (2017). A learning progression for elementary students' functional thinking. *Mathematical Thinking and Learning*, *19*(3), 143–166. https://doi.org/10.1080/10986065.2017.1328636

Suh, J., & Moyer, P. S. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching*, *26*(2), 155–173.

Van Amerom, B. A. (2002). *Reinvention early algebra. Developmental research on the transition from arithmetic to algebra.* CD-ß Press / Freudenthal Institute.

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2014). Realistic mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 521–525). Springer. https://doi.org/10.1007/978-94-007-4978-8_170

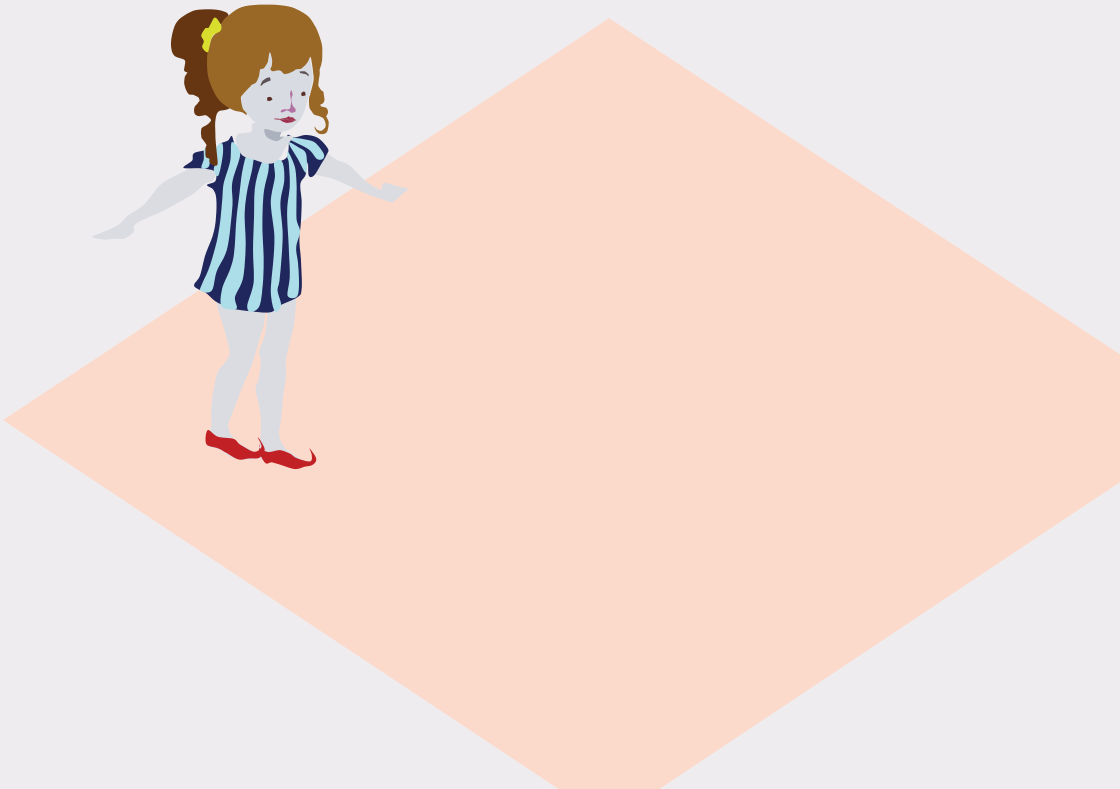
Van Reeuwijk, M. (1995, April). The role of realistic situations in developing tools for solving systems of equations. *Paper presented at the annual meeting of the American Educational Research Association,* San Francisco, CA.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics, 49*(3)*,* 341–359. https://doi.org/10.1023/A:1020229023965

Warren, E., & Cooper, T. J. (2005). Young children's ability to use the balance strategy to solve for unknowns. *Mathematics Education Research Journal, 17*(1), 58–72. https://doi.org/10.1007/BF03217409

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/BF03196322

Author contributions: All authors participated in designing the research. MO gave the lessons and collected the data. All authors developed the assessment tasks and the coding scheme. MO, MV coded the data. MO prepared the first draft of the manuscript. All authors participated in revising the manuscript and/or provided feedback. All authors read and approved the final manuscript.



CHAPTER 4

Are physical experiences with the balance model beneficial for students' algebraic reasoning? An evaluation of two learning environments for linear equations

Mara Otten, Marja van den Heuvel-Panhuizen, Michiel Veldhuis, Jan Boom, and Aiso Heinze

Published in *Education Sciences, 10*(6), 163.

Are physical experiences with the balance model beneficial for students' algebraic reasoning? An evaluation of two learning environments for linear equations

Abstract

The balance model is often used for teaching linear equation solving. Little research has investigated the influence of various representations of this model on students' learning outcomes. In this quasi-experimental study, we examined the effects of two learning environments with balance models on primary school students' reasoning related to solving linear equations. The sample comprised 212 fifth-graders. Students' algebraic reasoning was measured four times over the school year; students received lessons in between two of these measurements. Students in Intervention Condition 1 were taught linear equation solving in a learning environment with only pictorial representations of the balance model, while students in Intervention Condition 2 were taught in a learning environment with both physical and pictorial representations of the balance model, which allowed students to manipulate the model. Multi-group latent variable growth curve modelling revealed a significant improvement in algebraic reasoning after students' participation in either of the two intervention conditions, but no significant differences were found between intervention conditions. The findings suggest that the representation of the balance model did not differentially affect students' reasoning. However, analyzing students' reasoning qualitatively revealed that students who worked with the physical balance model more often used representations of the model or advanced algebraic strategies, suggesting that different representations of the balance model might play a different role in individual learning processes.

Keywords: Early algebra, Linear equation solving, Balance model, Representations, Physical experiences

1. Introduction

Mathematical reasoning is an essential aspect of learning and doing mathematics (Cai et al., 1996; National Council of Teachers of Mathematics [NCTM], 2000). It involves making and evaluating mathematical conjectures, identifying mathematical patterns and relationships, and justifying mathematical thinking and actions (Stein et al., 1996; Thom, 2011). Well-developed mathematical reasoning entails noticing relations both in mathematical contexts and in the world around us, which makes mathematical reasoning a powerful way to gain insight into a wide range of real-world phenomena (NCTM, 2000). This mathematical reasoning is considered "a habit of mind, and like all habits, it must be developed through consistent use in many contexts" (NCTM, 2000, p. 56). Unfortunately, the development of mathematical reasoning is often overlooked in primary mathematics education. Here, the emphasis is predominantly on arithmetic skills, performing operations with particular numbers and quantities, while there is less attention for mathematical reasoning, focusing on relationships between variables or sets of values (e.g., Kilpatrick et al., 2001).

Algebraic reasoning has been recognized as a powerful vehicle to develop children's mathematical reasoning (e.g., Blanton et al., 2015; Brizuela & Schliemann, 2004; Kaput et al., 2008). Learning to reason algebraically means learning to make generalizations on the basis of particular instantiations of mathematical ideas as well as building, justifying, and expressing conjectures about mathematical structures and relationships (e.g., Blanton & Kaput, 2005; Cai & Knuth, 2011). Algebraic reasoning of young students can be fostered by engaging them in solving problems that draw on their existing knowledge and skills (e.g., Blanton et al., 2007; Smith & Thompson, 2008; Stephens et al., 2017). The "Candy Problem" is an example of such a problem which can elicit algebraic reasoning (see Brizuela & Schliemann, 2004). In this problem, two children have the same number of candies: the first child has one box, two tubes, and seven loose candies; the second child has one box, one tube, and 20 loose candies. The number of candies in each of the boxes is the same and the number of candies in each of the tubes as well; the students' task is to figure out the number of candies in a box and a tube. Such a problem, which is meaningful to students (i.e., they can imagine what happens in the problem) and which draws on comprehension and skills they already have, can elicit natural, intuitive context-connected reasoning, which can be considered a first step towards more abstract algebraic reasoning (in this case, solving linear equations with unknowns on both sides of the equal sign).

The current study was initiated to investigate how primary school students' algebraic reasoning could be stimulated. We focused on one particular aspect of algebraic reasoning: reasoning related to solving linear equations (Kaput et al., 2008). To this end, we developed an intervention program consisting of a series of six lessons. In these lessons we aimed to foster students' algebraic reasoning by providing them with a learning environment in which they were able to invent, in an informal way, strategies for solving linear equations. A balance model played a central role in this teaching sequence. More specifically, we used a *hanging mobile,* a physical balance model consisting of a dynamic beam with on each side a number of bags hanging on a chain, representing an equation with unknowns (see for a similar approach the mobile puzzles used by Goldenberg et al., 2015). In an earlier study, we established that students developed informal context-connected algebraic strategies which underlie conventional equation solving strategies such as restructuring, isolation, and substitution through working with this physical hanging mobile (Otten, Van den Heuvel-Panhuizen, Veldhuis, & Heinze, 2019, see *Chapter 3* of this thesis). Students eventually used these strategies for solving informal linear equations in new contexts and even for solving systems of symbolically presented linear equations.

In the current study, we quasi-experimentally investigated the effect of our intervention program with the balance model on students' linear equation solving performance. More specifically, we examined the effect of using a physical balance model in comparison with a pictorial representation of the balance model.

1.1 Using the balance model for linear equation solving

Characteristic of an equation is that the expressions on both sides of the equal sign represent the same value and, in this sense, are equal (Jones et al., 2012; Kieran, 1981). This equality should be maintained when solving an equation. Students' understanding of equality is particularly visible in their interpretation of the meaning of, and reasoning about, the equal sign. A correct understanding of equality and the equal sign is crucial for learning linear equation solving (e.g., Bush & Karp, 2013; Kieran et al., 2016). Whereas the most appropriate interpretation is considering the equal sign as a relational symbol representing equality, it is often misinterpreted with students thinking that the equal sign is a signal for "here comes the answer" or a "do something"-signal (e.g., Behr et al., 1980; Carpenter et al., 2003; Falkner et al., 1999; Knuth et al., 2006; McNeil & Alibali, 2005). Relational understanding of the equal sign can be fostered by referring to the two sides of the equation being "in balance" (Antle et al., 2013).

The balance model is an often-used meaningful context to stimulate and structure students' reasoning related to solving linear equations (e.g., Figueira-Sampaio et al., 2009; Papadopoulos, 2019; Suh & Moyer, 2007; Warren & Cooper, 2005). It resembles familiar objects such as a seesaw (e.g., Alibali, 1999; Kaplan & Alon, 2013), or a kitchen scale, which makes it so that students can imagine what happens when this model is used. The balance model can be used to bring the focus on an equation as representing a mathematical structure linking two different algebraic expressions. It can be utilized to show that both sides of the equation represent the same quantity (or: are *in balance*) and are thus interchangeable (e.g., Bajwa & Perry, 2019; Mann, 2004; Warren & Cooper, 2005). This makes the model particularly deemed suitable for promoting relational reasoning around the idea of equality in an equation, for example by eliciting strategies which keep the model in balance (e.g., Anthony & Burgess, 2014; Vlassis, 2002), and which represent strategies that can be carried out on the equation.

A variety of balance models have been used to promote young students' understanding of concepts related to linear equations. For example, Cheeseman et al. (2017) reported on the use of a physical balance model with five- to seven-year-old students. Students experimented with the physical balance model by making use of a range of equipment with different weights. Explorations of this physical model fostered students' understanding of equality, which was, for example, reflected by one student's comment, "I add the same to each side and it stays even" (p. 154). In another study, a computer-based balance model with known weights (e.g., a weight labelled with 50 g) and unknown weights (a weight labeled with *X*) was used to teach sixth-grade students solving equations such as $5x + 50 = 3x + 290$ (Figueira-Sampaio et al., 2009). Manipulations on the virtual model directly resulted in changes in the corresponding symbolic equation, which made this model especially suitable for demonstrating the relationship between the manipulations on the model and the changes in the formal algebraic symbols. Pictorial representations of the balance model can also be used for exploring concepts and strategies related to linear equation solving, such as puzzles on paper that include collections of balanced objects with unknown weights hanging on two sides of a balanced beam (i.e., mobile puzzles; see Goldenberg et al., 2015). These puzzles were, for example, used with sixth-grade students in a study by Papadopoulos (2019). After working with the puzzles, students showed a wide range of reasoning abilities which can be considered as first steps towards the algebraic strategies and conventional steps for solving (systems of) linear equations, such as adding or taking away similar symbols (i.e., unknowns) on both sides, isolating particular symbols, and substitution of symbols by weights or by other symbols.

These studies together indicate the wide variety in representations of balance models used. A recent review study confirmed this apparent diversity in appearances, and additionally showed the different situations in which the model was used for teaching linear equation solving as well as different rationales for using the model (Otten, Van den Heuvel-Panhuizen, & Veldhuis, 2019, see *Chapter 2* of this thesis). In general, most positive effects of using the balance model for linear equation solving were reported for (young) students encountering this algebraic topic for the first time. What this review also suggested was the possibility that different representations of the balance model (e.g., a physical model, a virtual model, or a model presented on paper) might result in different effects on students' learning outcomes; at least a few studies provided some indications for this. For example, Suh and Moyer (2007) compared the effects of using a dynamic virtual balance model on third-grade students' linear equation solving abilities, with the effects of using a static model on paper in combination with manipulatives. Students working with either of the two models improved in solving linear equations (as shown on a combination of pictorial, numerical, and word problems) and gained flexibility in representing their reasoning. However, no significant differences between the two interventions regarding students' learning gains on solving equations were reported. Qualitative analyses showed that both models had unique learning facilitators, such as immediate feedback and a direct link between manipulations on the equation and changes in the corresponding symbolic equation for the virtual model, and tactile features for the model on paper with manipulatives.

There are further studies using balance models but without comparisons of different representation types. For example, Figueira-Sampaio et al. (2009) explored the change of students' activities in the Brazilian classroom when a physical balance model is replaced by virtual balance models. They compared the use of one physical balance model as a demonstration model in the traditional classroom with the use of a virtual balance model in small groups of students. Students using the virtual balance model during group work showed higher participation, social interaction and dialogue, motivation, and reflection than students who had only seen a physical balance model at the front of their classroom. A comparison between the effects on students' ability to solve linear equations of the different representations of the model was not made in this study. Bajwa and Perry (2019) investigated the effect of using various virtual balance models on students' ability to solve problems such as 3 + 4 + 2 = 3 + __ and the meaning of the equal sign. Students who worked with either of the representations of the virtual balance model showed higher learning gains compared to students in the control condition who only solved symbolic problems. In addition, higher learning gains were found for students who worked with a static virtual balance model consisting of only two pans with a number of blocks (with feedback provided by means of an equal or unequal sign) than for students who worked with a dynamic virtual balance model.

Although in some studies on linear equation solving multiple representations of the balance model were used in a sequence starting with a physical model followed by a pictorial model on paper (e.g., Fyfe et al., 2015; Warren & Cooper, 2005), a direct comparison between the effects of using a physical balance model or a pictorial balance model on students' linear equation solving abilities has not been reported on. From the perspective of embodied cognition, this comparison might be worthwhile to investigate, because (mathematical) cognition seems to benefit from physical experiences of our body in interaction with the world around us (e.g., Alibali & Nathan, 2012; Núñez et al., 1999; Wilson, 2002). The physical experience of maintaining balance might be helpful for understanding the abstract concept of equality in a linear equation (Alessandroni, 2018; Antle et al., 2013; Gibbs Jr, 2006; Otten, Van den Heuvel-Panhuizen, Veldhuis, & Heinze, 2019, see *Chapter 3* of this thesis), because the actions performed on the balance model could act as metaphorical mapping for developing strategies to maintain equality in an equation. Moreover, it has previously been established that using concrete materials in learning algebra can help students to move from concrete physical experiences to abstract reasoning (Carbonneau et al., 2013; NCTM, 2000). Additionally, from the perspective of the feasibility of using this model in the classroom, it is interesting to know whether pictorial balance models are as effective as their physical real-world counterparts for teaching linear equation solving.

1.2 Current study

The goal of our study was to learn more about the relevance of different representations of the balance model for developing students' reasoning when solving systems of linear equations. For this, we examined the effect of two learning environments consisting of a teaching sequence with a balance model on students' development of algebraic reasoning about linear equations. More specifically, our interest was in whether a static version of the balance model presented on paper would have a different effect on the development of students' algebraic reasoning about linear equations than a dynamic physical balance model with which students could gain physical experiences with equality. We expected students to benefit from the physical experiences when solving linear equations, resulting in a more frequent use of the model of a balance when solving linear equations in contexts not related to the balance model (either explicitly by making use of the representation of the model or implicitly by making use of algebraic strategies) and in a larger improvement in their algebraic reasoning.

2. Materials and methods

2.1 Participants

The study was carried out with a convenience sample. About 40 schools, which were easily accessible for us, were contacted by email about whether they would like to participate in our study on fostering primary school students' mathematical reasoning. Schools and classes were selected based on availability and on teachers' willingness to participate. Participants included 229 students of nine fifth-grade classes in eight schools in the Netherlands, four public and four of Christian denomination (Catholic or Protestant). We chose fifth-graders for our study because in general Dutch students of this age have no previous experience with solving equations. Parental consent was obtained for all students except 12, who were excluded from the analyses. Five other students were excluded because they missed most of the lessons. The final sample consisted of 212 students (47% boys), with ages ranging from 9 to 11 (average: 10 years old). Students of three classes participated in Intervention Condition 1, in which a balance model on paper was used in the instruction ($n = 67$, 49% boys), students of three classes were in Intervention Condition 2, in which the same intervention program was used with in addition a physical balance model ($n = 65$, 42% boys). We also included three classes ($n = 80$, 50% boys) in a control condition, in which no instruction on algebra but on probability was provided. All students had not received prior instruction on equation solving or other algebra topics. This remained the case throughout the year in which the study took place.

2.2 Conditions

Two parallel versions of the intervention were created which were identical in terms of the length (six lessons), content, task types, and sequence (see Figure 1) but which differed in terms of the used representation of the balance model. A static, pictorial representation of a hanging mobile as balanced model was used in both intervention conditions. In Intervention Condition 1, students only worked with this static version of the hanging mobile on paper. Students in Intervention Condition 2 were taught the same lessons with the same tasks and were presented with the same problems during the lessons, but in addition to the hanging mobile on paper, a physical hanging mobile was provided allowing students to gain physical experiences. Lastly, the students in the control condition participated in an intervention consisting of a six-lesson teaching sequence on probability — a topic which is also not taught at primary school in the Netherlands. This control condition was included to ensure that possible differences between the intervention conditions and the control condition could not be attributed to receiving additional lessons on a (to them) new mathematical topic. Adding this control condition, moreover, assured us that a possible effect of the intervention could not be attributed to, for example, retest effects.

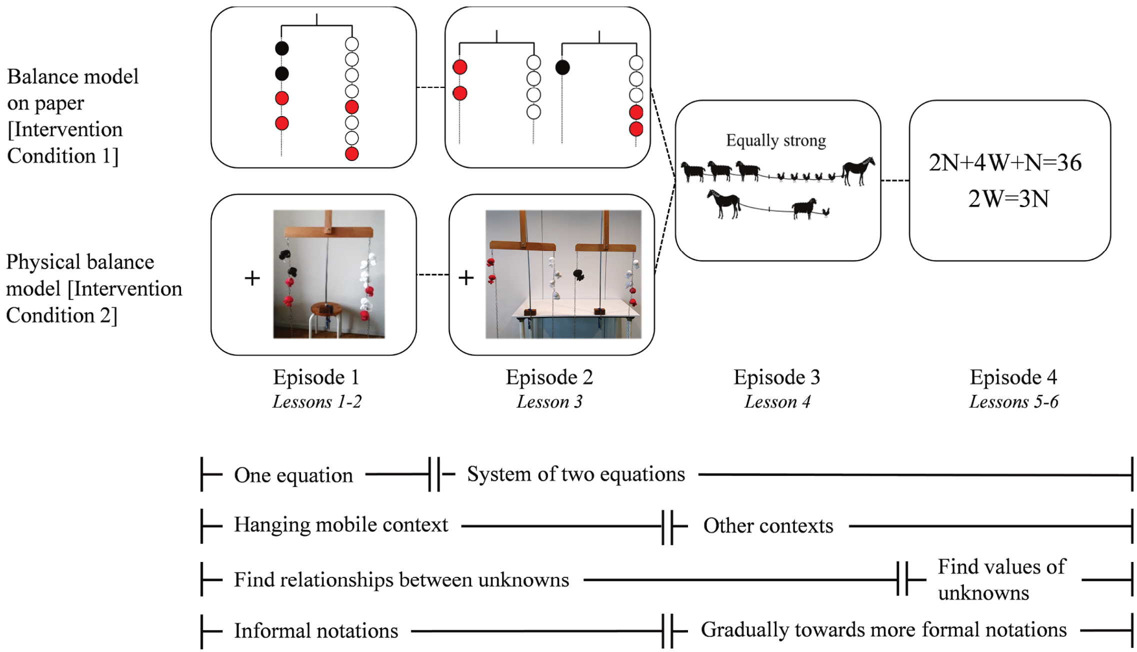


Figure 1. Schematic representation of the intervention and the main elements comprising this intervention (for both intervention conditions).

2.3 Intervention program

The intervention program consisted of a six-lesson teaching sequence on solving linear equations (see also Otten, Van den Heuvel-Panhuizen, Veldhuis, & Heinze, 2019, see *Chapter 3* of this thesis). In the beginning of this teaching sequence, the focus was on solving informal linear equations, that is, equations posed in informal contexts that students directly have a good understanding of. Over the course of the teaching sequence, more formal equations were introduced gradually. The teaching sequence could be clustered into four episodes based on the focus and content of the lessons (see Figure 1). In the first episode, students could develop informal algebraic strategies related to linear equation solving. Instead of the teacher transferring the strategies to the students, the students were active participants in developing the strategies. Students worked in small groups (2-3 students) using a hanging mobile as balance model and reasoned about relationships between unknowns. Their main task was to discover all possible ways to maintain the balance of the mobile (i.e., the equality). Students could, for example, exchange the balls of the left and right side of the mobile to figure out that both sides are interchangeable (i.e., apply a *restructuring* strategy), take away similar balls from both sides (i.e., apply an *isolation* strategy), or substitute one color of balls with another color and as such make use of the relationship between different unknowns (i.e., apply a *substitution* strategy). During the ensuing classroom discussion, students could mention all possibilities they discovered to maintain the balance of the mobile. The teacher wrote students' ideas on the blackboard, which resulted in an overview of the various possibilities. Students' own wordings were used in this overview (e.g., "change one color of bags by bags of another color" instead of "substitution").

From Episode 2 on, the information from two hanging mobiles had to be combined to discover new relationships between unknowns. At this time, students had to reason about a *system* of informal equations in the context of the hanging mobile. In Episode 3, problems were posed in new informal contexts which are often used for eliciting algebraic reasoning, such as a tug-of-war situation (e.g., Kindt et al., 1998). After a classroom discussion about equality being crucial in both the familiar context (the hanging mobile) and the new context (the tug-of-war situation), all possible strategies for maintaining this equality were discussed. Then, students were again invited to discover relationships between unknowns in this new context. In the example of the tug-of-war situation presented in Figure 1, students could, for example, apply a substitution strategy and replace one horse in the first informal equation by a sheep and a chicken (on the basis of the second equation), and take away a sheep from both sides in other to isolate the chickens on the right side of the equation. Moreover, in this third episode, students were gradually challenged to use more symbolic notations when writing down their reasoning. Finally, in Episode 4, students reasoned about systems of formal linear equations. The resemblance between the familiar contexts and the new context was discussed again, as well as the meaning of the algebraic symbols (e.g., what does W stand for?). Then, the students' task was to use all previously discovered strategies for maintaining the equality of an equation, with the goal to combine the information of both symbolically notated equations in order to determine the values of the unknowns. In the example of the system of two formal linear equations presented in Figure 1, students could, for example, restructure the first equation by combining $2N$ and N into $3N$, resulting in $3N + 4W = 36$, then substitute $3N$ by $2W$ on the basis of the second equations (resulting in $6W = 36$), and then further isolate the unknown W by dividing both sides by 6.

2.4 Measures

2.4.1 Algebraic reasoning

Students' algebraic reasoning related to solving linear equations was assessed by a paper-and-pencil test. Open-ended problems were used to explicitly invite students to explain their thinking and thus reveal their reasoning. The test consisted of four problems in which students had to solve (a system of) linear equations. The four problems were part of a larger test that also included problems in two other mathematical domains, namely graphing (four) and probability (five). In this study, we only focus on the problems on linear equation solving.

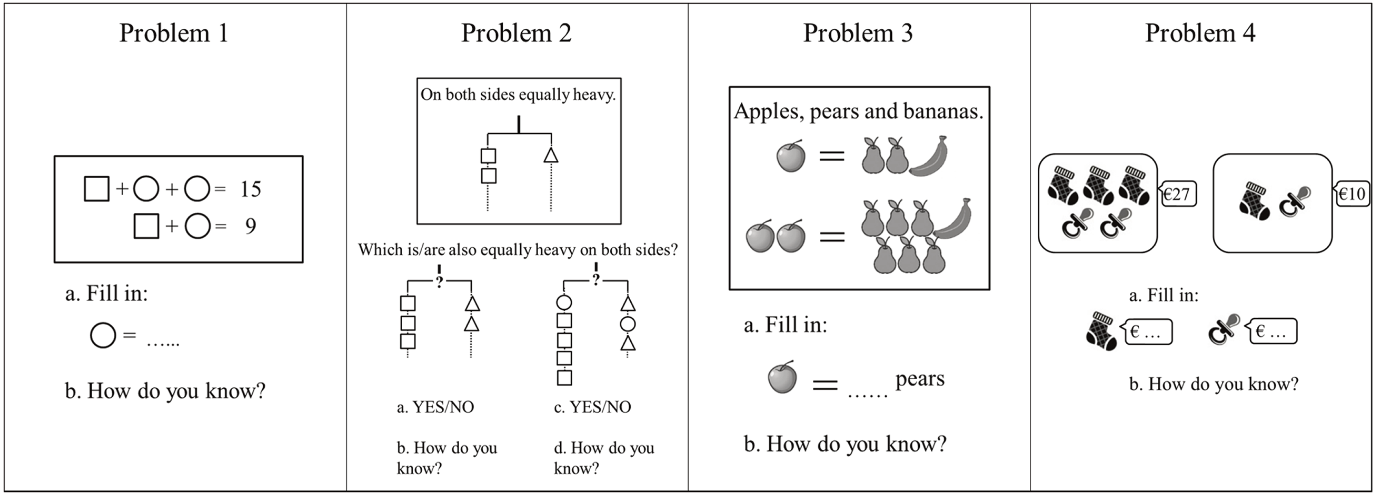


Figure 2. The four problems in the algebraic reasoning test (translated from Dutch).

4

The problems on linear equation solving (see Figure 2) were formulated in such a way that prior instruction in formal linear equation solving was not necessary to solve the problems. Information presented in two informal equations had to be combined. In Problems 1, 3, and 4 a system of two linear equations was presented which had to be solved; in Problem 2 the information from a given equation had to be compared with the information from two other equations. Whereas in Problems 1 and 4 the goal was to find the value of the unknowns, in Problems 2 and 3 the task was to discover a relationship between unknowns. The algebraic strategies of restructuring, isolation, and substitution were needed to solve the problems.

2.4.1.1 Coding

Students' reasoning on each of the problems was categorized by means of a coding scheme. An iterative process, inspired by the constant comparative method (Glaser, 1965), was used for its development. The first version of the coding scheme was developed by taking the work of a couple of students. We first analyzed students' reasoning based on the strategies of restructuring, isolation, and substitution when solving the systems of equations. When students used the strategies of restructuring or isolation, this usually meant that in their reasoning they only integrated the information from one of the two equations in the problem. When students made use of the more advanced substitution strategy, they generally combined the information from both given equations in their reasoning. Sometimes substitution was separately used, but more often in combination with the restructuring and isolation strategies. Another strategy which was often used when combining both equations was the strategy of elimination of unknowns by subtracting one equation from the other.

There appeared to be an intricate relation between the use of relevant algebraic strategies in students' reasoning and the number of equations they referred to while describing their solution of the problem. Focusing on the number of equations in students' reasoning turned out to be the most reliable way of coding students' written work, as this was almost always quite clearly visible, whereas the algebraic strategies were much more indirectly mentioned. In the end, we therefore decided to distinguish between students who did not use any of the two given equations in their reasoning (Level R0), students who reasoned on the basis of only one of the two given equations (Level R1), and students who reasoned on the basis of both given equations by combining the information of both of them (Level R2). Importantly, these levels of students' algebraic reasoning thus reflected both the straightforward number of equations they referred to in their reasoning and also the depth of their reasoning by the use of the algebraic strategies.

After several rounds of coding, 100% consensus was reached between the researchers and a final coding scheme was established, which was used for a final round of data coding. The final coding scheme, with examples of student responses for each problem and each level of reasoning, can be found in Appendix 4.1. Examples of responses in which students did not show the use of any of the equations in their reasoning (Level R0) were: "I made a guess", "Because I think this should be the answer", "?", or "I have no clue". Student responses in which only one of both given equations was used (Level R1), for example in Problem 4 (see Figure 2), only referred to one equation, such as "5 + 5 = 10, so both must be €5", "this fits in the second one, because 7 + 3 = 10", or "in the first one you see 3 socks and 2 pacifiers must be 27, so one sock = 5 and the pacifier = 6". Examples of responses in which students reasoned on the basis of both given equations (Level R2) were in Problem 4, explanations like "3 × 7 = 21, 2 × 3 = 6, 6 + 21 = 27. And 7 + 3 = 10. So, this must be the answer", or "€10 + €10 = €20. Then one sock is left, so the sock must be €7. Then the pacifier = €3". When a student did not provide any reasoning, this was coded as missing.

Inter-rater reliability was established by having an independent second rater recode the responses of 10% of the students. Two or three students were randomly drawn from each class, resulting in a sample of 21 students (84 responses for each problem, 336 responses in total). Inter-rater reliability between coders was high (Cohen's kappa = .92).

2.4.2 General reasoning ability

An abbreviated version of Raven's Standard Progressive Matrices (SPM; Raven et al., 1996), consisting of 18 items (Bilker et al., 2012), was used as a measure of general reasoning ability. The items, which increased in difficulty, consisted of diagrams with one part missing. Students have to reason which part is missing, before selecting this missing part to complete the design among six or eight alternatives. Answers were scored as incorrect (0) or correct (1), resulting in the minimum score of 0 and the maximum score of 18.

2.4.3 General mathematics performance

Students' general mathematics performance was measured by the CITO Monitoring System, a Dutch standardized test for different subjects and grade levels (Janssen et al., 2005). The end-term scores of Grade 4 were obtained.

2.5 Research design and procedure

The study was approved by the ethical committee of Utrecht University. A staged comparison design with two intervention conditions and one control condition was used (see Table 1). Making use of this staged design made it possible that the same teacher taught all experimental lessons in both intervention conditions. We distinguished three cohorts which differed in the timing of the teaching sequence, and a fourth cohort with the control condition who did not receive algebra instruction but lessons on probability. Each of the three cohorts in the intervention conditions was made up of matched pairs of classes, based on characteristics such as the location of the school, the type of school, and the percentages of students going to particular levels of secondary education. Subsequently, within each pair, the classes were randomly divided over the two intervention conditions.

Over the school year, the students' algebraic reasoning related to solving linear equations was assessed four times by means of the same algebraic reasoning test, with approximately two months in between. In this way, the algebraic reasoning of students in each cohort was measured before and after participating in the teaching sequence. The teaching sequence consisted of six lessons, of about 50 min each. The students were taught one lesson a week, during six consecutive weeks. The lessons in both intervention conditions (Cohorts 1-3) were taught by the first author of this paper, while the probability lessons in the control condition (Cohort 4) were taught by another researcher from the same research group. Raven's SPM was administered in each class before the beginning of the study.

Table 1
Research design

	Cohort	*n*	Measurement 1 October 2016	November–December 2016	Measurement 2 Dec. 2016	February–March 2017	Measurement 3 March 2017	May–June 2017	Measurement 4 June 2017
Balance model on paper [Intervention Condition 1]	1	22	M1	Teaching sequence (6 lessons)	M2		M3		M4
	2	21	M1		M2	Teaching sequence (6 lessons)	M3		M4
	3	24	M1		M2		M3	Teaching sequence (6 lessons)	M4
Physical balance model [Intervention Condition 2]	1	22	M1	Teaching sequence (6 lessons)	M2		M3		M4
	2	18	M1		M2	Teaching sequence (6 lessons)	M3		M4
	3	25	M1		M2		M3	Teaching sequence (6 lessons)	M4
Control condition	4	80	M1		M2		M3		M4

2.6 Data analysis
2.6.1 Qualitative analysis
To get insight into students' development in reasoning, we first identified for the whole sample, for each cohort of students, and for each problem the most prevalent patterns of reasoning. The work of two students, whose patterns of reasoning were most prevalent (and thus representative) on that problem, was further analyzed and discussed.

We then compared *all* students' use of the model after participating in an intervention with either a balance model on paper or a physical balance model. In this way, we could frame the use of the balance model of the two students whose reasoning was analyzed more deeply, and we could shed more light on the effects of working with different representations of the balance model. On the measurement directly after the intervention we looked into (1) whether students explicitly used a representation of the balance model (i.e., a drawing of the model) in their reasoning, and (2) whether students implicitly used the model as shown in their use of the algebraic strategies. This was only done for Problems 1, 3, and 4, because in Problem 2 the representation of a balance model was already part of the question. Because not all algebraic strategies were equally easy to discern in the students' reasoning we decided to focus only on the more advanced algebraic strategies for combining both equations: substitution of a part of one equation on the basis of the information from the other or subtracting one equation from the other in order to eliminate unknowns.

2.6.2 Quantitative analysis
2.6.2.1 Descriptive statistics
Analyses of variance (ANOVAs) were performed to compare the three conditions on general reasoning ability and general mathematics performance. Proportions of each level of reasoning (R0, R1, R2) on the algebraic reasoning test were calculated for each cohort of each condition on each of the measurements.

2.6.2.2 Multi-group latent variable growth curve modeling
Latent variable growth curve modeling (LGM) was used to model students' development in algebraic reasoning about linear equations over the four measurements. LGM is a powerful and flexible technique for modeling longitudinal change using repeated measures (Bollen & Curran, 2006). The core of such an LGM is a latent ability, in this case students' reasoning about linear equations, that is different for each participant (inter-individual differences), but which also

possibly changes *within* participants (intra-individual differences) over the four measurements. A cohort sequential multi-group LGM (Duncan et al., 2006) was used in this study, with the cohorts as groups.

Item response theory (IRT) was used to map the likelihood of a level of reasoning (Level R0, R1, or R2) onto students' latent algebraic reasoning ability. This latent ability was modeled as the combination of four partial effects: (1) The *intercept effect*: The baseline over all measurements; (2) The *slope effect*: The linear change from one measurement to the next; (3) The *intervention effect*: The effect of the intervention could only influence the score in the measurements following the intervention (e.g., when the intervention took place between M1 and M2, this would influence M2, M3, and M4); (4) The *weakening effect:* The weakening of the effect of the intervention could only influence the score in the delayed measurements after the intervention (e.g., when the intervention took place between M1 and M2, this would influence M3 and M4). The possibility existed that there were baseline differences in ability between the different cohorts in our study. Therefore, differences between the intercepts of the different cohorts were allowed. Because of the different timing of the intervention, the loadings for intervention and weakening also systematically differed for Cohorts 1-3. Because there was no intervention in Cohort 4, we did not include an intervention or weakening effect in this cohort. All other parameters were modelled exactly the same in all cohorts. Using LGM thus allowed us to disentangle students' possible (linear) development over the four measurements (represented by the intercept and the slope) from the intervention effect.

In addition to these partial effects, three predictors were added to the model: (a) Condition was added as a dummy predictor of the intervention effect (coded as –1 and 1, for Intervention Conditions 1 and 2 respectively); (b) A measure of general reasoning ability (Bilker et al., 2012; Raven et al., 1996) was added in a centered form as a predictor of the intercept; (c) A measure of general mathematics performance (Janssen et al., 2005) was added in a centered form as a predictor of the intercept.

The model was fitted in Mplus 8 (Muthén & Muthén, 1998-2017), with the weighted least squares means and variances adjusted estimator (WLS-MV). Following commonly applied cut-off criteria, model fit was considered acceptable with the Root Mean Square Error of Approximation (RMSEA) below .08 and the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) above .90 (Little, 2013). A PROBIT link was used, which means that differences between difficulty and ability are expressed in units that refer to a standard normal distribution with a mean of zero, with units representing standard deviations.

2.6.3 Missing data

There were four students for whom one of the measurements M1-M3 was completely missing, while subsequent measure(s) were present. We reasoned that the baseline linear change (i.e., the slope effect) could be estimated less reliably when one of the measurements in between was missing. We therefore decided for each case to replace the missing measurement by the subsequent measurement. More specifically, M2 of one student from Cohort 1 was missing; M3 was used as if it were M2 (as the measurement directly after the intervention) and M4 as M3. The same procedure was applied to the three other students, belonging to the control condition, of which M1, M3, and M3 were missing, respectively. Class averages were calculated and imputed for students' missing general mathematics performance scores.

3. Results

In this results section, we first give an idea of what students' development in algebraic reasoning included by providing a qualitative analysis of two students' reasoning over the measurements and their use of the balance model. The patterns of reasoning of these two students were most prevalent (and thus representative) on these two problems. Next, we investigate whether the results of these two students can be generalized to the whole sample. We qualitatively investigate the effect of working with different representations of the balance model on *all* students' use of this model when solving systems of informal linear equations in other contexts (i.e., contexts not related to the balance model). We distinguish between explicit use (i.e., using a representation of the model) and implicit use (i.e., using algebraic strategies).

After presenting the results of this qualitative analysis, we continue with the findings of the quantitative analysis of students' level of reasoning. Here, we investigate the effects of using the balance model on students' levels of reasoning in both the short term and the long term. In addition, we consider the effects of the students' working with a physical versus a pictorial representation of the balance model on their levels of reasoning.

3.1 Results from the qualitative analysis of students' reasoning

3.1.1 Case 1 – Noah

Noah participated in Cohort 1 and worked with the hanging mobile on paper (Intervention Condition 1). His reasoning on this problem consisted of the pattern R1-R2-R2-R2, a pattern which was the most prevalent in this cohort and displayed by 9% of the students. Noah's answers and his reasoning on the measurements right before the intervention (i.e., the pretest M1) and right after the intervention (i.e., the direct posttest M2) are displayed in Figure 3.

On the measurement before the intervention, Noah based his answer on the first equation, in which two pears are displayed as part of the equation. He ignored the rest of the first equation and the entire second equation. This reasoning was therefore categorized as Level R1 (reasoning on the basis of only one equation). No algebraic strategies came to the fore in his reasoning. On the measurement directly after the intervention, Noah first converted both equations into symbols. Subsequently, he showed that the first equation could be subtracted from the second equation, revealing the relationship between the apples and the pears. This reasoning was categorized as Level R2 (reasoning on the basis of both equations). Noah's answer on this problem can be seen as a clear demonstration of the effect of the intervention. In his reasoning he showed understanding of how to combine the information of both equations in the problem: by subtracting one equation from the other, one unknown was isolated. Moreover, he displayed his algebraic reasoning by making use of letters.

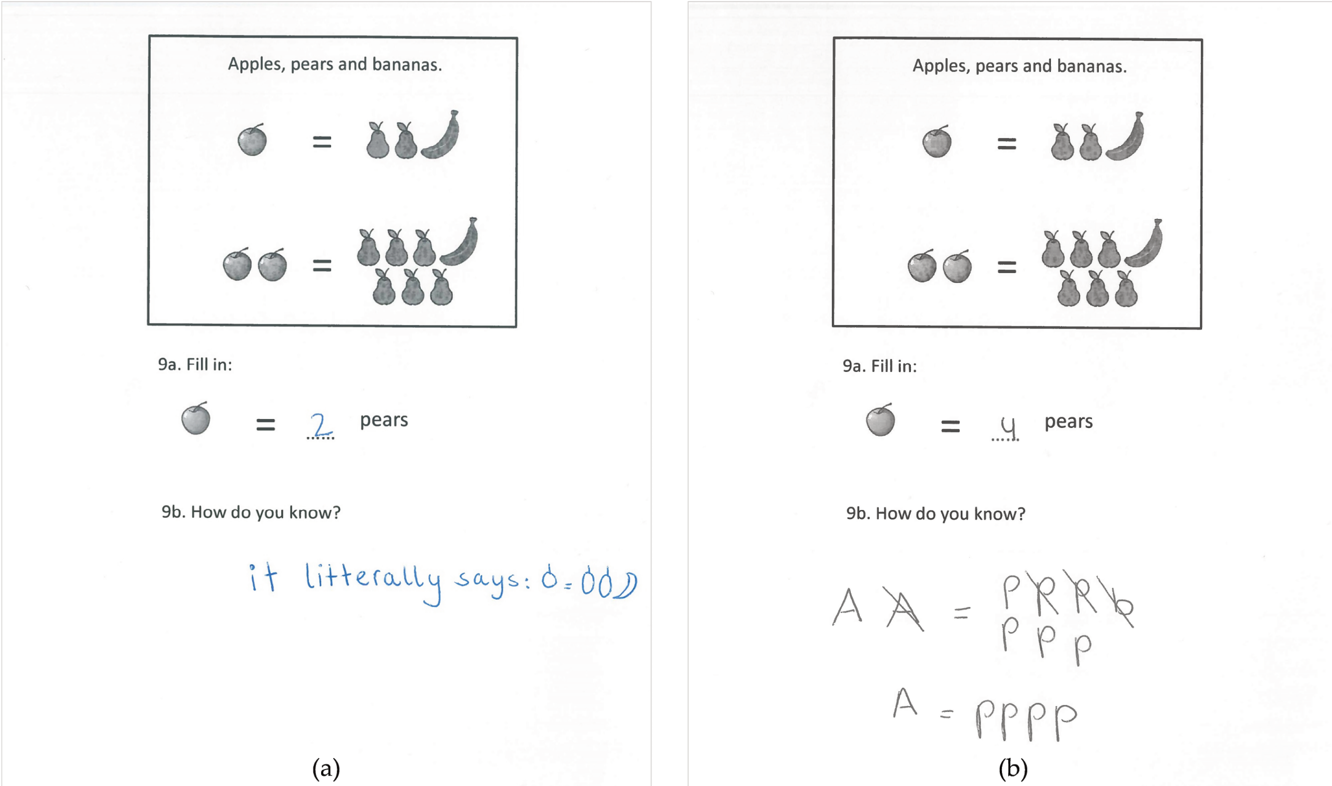


Figure 3. (a) Noah's (Cohort 1) reasoning on Problem 3 on the pretest M1; (b) Noah's reasoning on Problem 3 on the direct posttest M2 (translated from Dutch).

3.1.2 Case 2 – Lea

As a second example, we zoom in on the reasoning of Lea, a student from Cohort 3 who worked with the physical hanging mobile (Intervention Condition 2). Lea demonstrated a pattern of reasoning consisting of only Level R2 on all four measurements of Problem 4. This pattern was most prevalent for this problem in Cohorts 1 and 3 (displayed by 23% and 14% of the students respectively). Lea's answers on M1-M3 were very similar; her reasoning on the measurement before the intervention (i.e., the pretest M3) and after the intervention (i.e., the direct posttest M4) is shown in Figure 4.

On the measurement prior to the intervention, Lea substituted the values in both equations and thus showed that these values add up to the right amounts of €10 and €27. Because she made use of both given equations, her reasoning was categorized as the highest level of reasoning (Level R2). On the measurement after partaking in the teaching sequence, Lea started with converting both equations into hanging mobiles, making her reasoning visible. She then doubled the second equation and subtracted the value of 20 from the first equation. In this way, she isolated the sock and determined its value (€7). Subsequently, she substituted this value of 7 for one sock in the first equation to reveal that two pacifiers must be equal to €6 so one pacifier must be €3. This reasoning was again categorized as Level R2. Although the effect of the intervention is not directly visible in Lea's *pattern* of reasoning levels, we do see an effect when we zoom in on the extensiveness and completeness of her reasoning: whereas Lea in M1-M3 proved the correctness of her answer by substituting both values in both equations, in M4 she clearly made use of various algebraic strategies to come to her answer. She isolated one unknown by subtracting the second equation two times from the first equation and used the strategy of substituting unknowns by values. By converting the equations into hanging mobiles, which she used in combination with pre-formal algebraic symbols, she moreover showed her ability to incorporate different representations into her reasoning and her flexibility in switching between these representations.

For both students, the effect of working with the balance model during the intervention was visible in their algebraic reasoning on the measurement directly following the intervention. Both students, either after working with the model on paper (Noah) or with the physical model (Lea), displayed algebraic reasoning by eliminating unknowns through subtracting one equation from the other (i.e., they took away things on both sides of one equation on the basis of the other equation). Moreover, Lea explicitly made use of the model of the balance in her reasoning.

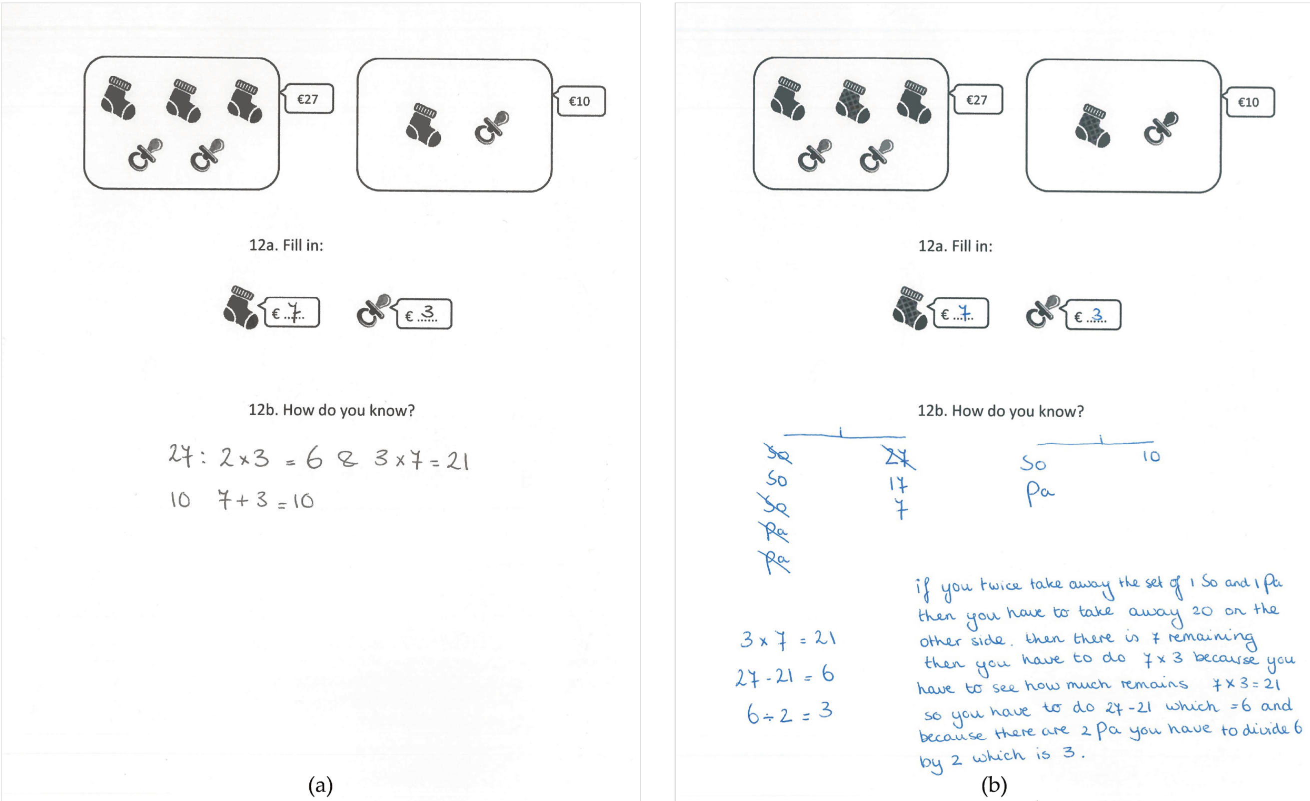


Figure 4. (a) Lea's (Cohort 3) reasoning on Problem 4 on the pretest M3; (b) Lea's reasoning on Problem 4 on the direct posttest M4 (translated from Dutch).

Examination of the work of *all* students who worked with the balance model *on paper* (Intervention Condition 1) revealed that 27 students (40%) made use in at least one of the problems of an advanced algebraic strategy (i.e., substitution or elimination) for combining both equations (like Noah). Only one student (1%) in this intervention condition explicitly made use of the representation of the balance model in their reasoning. On the contrary, 39 students (60%) who worked with the *physical* balance model (Intervention Condition 2) used at least once such an advanced algebraic strategy for combining both equations, and, moreover, 11 students (17%) explicitly used the model of the balance in their reasoning (like Lea).

So far, this analysis demonstrates differences between intervention conditions as regards the use of the balance model on the measurement after the intervention. There might also be differences between both conditions when focusing on students' levels of reasoning. An analysis of these differences is reported in the next section.

3.2 Results from the qualitative analysis of students' reasoning

In this section, we will further analyze students' development in algebraic reasoning. We will start with providing the descriptive statistics, and then present our LGM model.

3.2.1 Descriptive statistics

Students' general reasoning ability ($F(2, 209) = 1.11$, $p = .331$, partial $\eta^2 = .011$) and general mathematics performance ($F(2, 209) = 1.92$, $p = .149$, partial $\eta^2 = .018$) did not significantly differ between students in the three conditions (see Table 2).

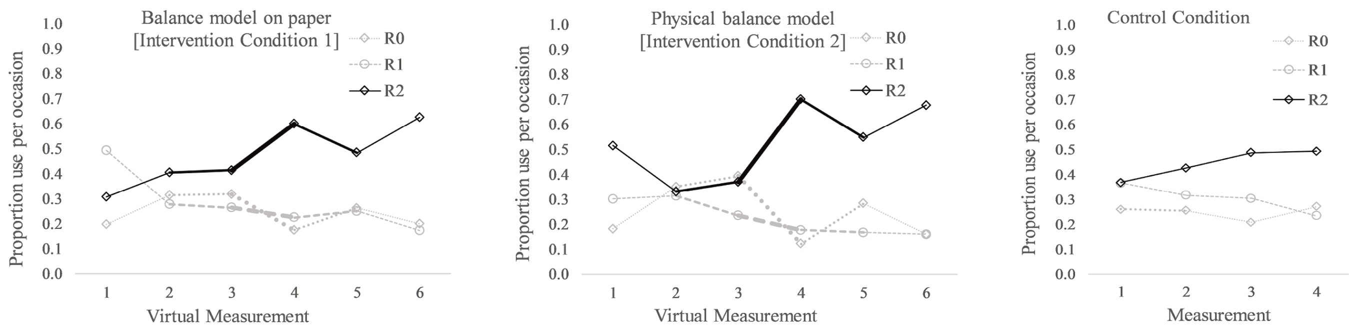
Subsequently, the proportions of each level of reasoning on the four algebraic reasoning problems were calculated. Over the course of the school year, students' algebraic reasoning was measured four times. Recall that the three cohorts, in which students received the lessons, differed in timing. As a consequence, while measurement 2 for students in Cohort 1, for example, was the test given directly after the lessons (i.e., the direct posttest), this same measurement 2 for students in Cohort 2 was the test given before the lessons (i.e., the pretest). To allow for direct (visual) comparison of the measurements and to analyze the change in reasoning of all cohorts of the intervention conditions at the same time, we created six virtual measurements. Students in Cohort 1 were depicted as having participated in virtual measurements 3-6, students in Cohort 2 as having participated in virtual measurements 2-5, and students in Cohort 3 as having participated in virtual measurements 1-4. In this way, for all cohorts together, virtual measurement 3 reflected the measurement directly before the intervention (i.e., the pretest), virtual measurement 4 reflected the measurement directly after the intervention (i.e., the direct posttest), and the change from virtual measurement 3 to 4 reflected the change in reasoning due to the intervention.

Figure 5 shows, for each condition, the proportion of each level of algebraic reasoning on each of the six virtual measurements. Here it is visible that in both intervention conditions the proportion of levels of reasoning R0 and R1 decreased after the intervention compared to before the intervention, while the proportion of level of reasoning R2 increased (Figure 5a,b). Thus, students in both intervention conditions showed more reasoning on the basis of both linear equations after participating in the teaching sequence. Moreover, the proportion of Level R2 increased more in Intervention Condition 2 (.33 increase in proportion) than in the cohorts of Intervention Condition 1 (.18 increase in proportion). Figure 5c shows the proportion of each level of reasoning on the four measurements for the control condition. No virtual measurements needed to be created for this condition to allow for comparison of the reasoning of all students together, because none of the students in this control condition participated in the algebra intervention so there was no shift in timing of an intervention which needed to be controlled for. Figure 5c shows that the proportions of levels of reasoning in the control condition remained more or less stable.

Table 2

Students' scores on general reasoning ability and general mathematics performance for all three conditions

	Cohort	General reasoning ability *M* (*SD*)	General mathematics performance *M* (*SD*)
Balance model on paper [Intervention Condition 1]	1	11.18 (2.84)	102.05 (9.48)
	2	11.00 (2.17)	96.57 (9.19)
	3	10.08 (2.59)	87.00 (10.82)
	Mean	10.73 (2.56)	94.94 (11.65)
Physical balance model [Intervention Condition 2]	1	9.95 (2.01)	95.76 (9.49)
	2	8.94 (2.78)	92.82 (9.21)
	3	10.92 (2.97)	92.48 (9.38)
	Mean	10.05 (2.71)	93.68 (9.35)
Control condition	4	10.49 (2.74)	97.32 (12.62)



(a) (b) (c)

Figure 5. Proportions of level of reasoning (R0, R1, R2) on the algebraic reasoning test, for each (virtual) measurement, for (a) Intervention Condition 1, (b) Intervention Condition 2, and (c) the control condition. The intervention took place between virtual measurements 3 and 4. For the intervention conditions, thin lines reflect the scores of students of one cohort, thicker lines of two cohorts, and the thickest lines are based on the scores of all three cohorts. For the control condition, all lines are based on three cohorts.

3.2.2 Multi-Group latent growth model

A multi-group LGM with a PROBIT link was fitted to investigate the overall effect of the intervention on students' reasoning ability in both the short term and the long term and to investigate the effect of the two different representations of the balance model on this reasoning ability. The model with an intercept, slope, intervention effect, and weakening effect, condition as predictor of the intervention effect and general reasoning ability as predictor of the intercept had an acceptable fit (RMSEA = .066, 90%-CI [.050 - .080], CFI = .926, TLI = .937). Adding general mathematics performance as a predictor resulted in a deterioration of the fit and was therefore disregarded in this analysis. Table 3 shows the parameter estimates of this model. The overall effect of the intervention on students' reasoning ability was significant ($M = 0.67$, $p < .001$). Students' algebraic reasoning thus improved after partaking in the teaching sequence. This effect showed weakening on the delayed measures after the intervention ($M = -0.31$, $p = .001$), which means that students' level of algebraic reasoning decreased a little in the long term. The differential effect of condition (physical vs. pictorial balance model) on the intervention effect turned out to be nonsignificant ($\beta = .33$, $p = .136$). In other words, the representation of the balance model did not differentially affect students' reasoning. Lastly, general reasoning ability was a significant predictor of students' baseline reasoning ability (i.e., the intercept, $\beta = .34$, $p <.001$), which means that a higher general reasoning ability was associated with a higher baseline level of algebraic reasoning.

Table 3
Parameter estimates of multi-group LGM model

Model parameter	*M*	*p*-value	*var*
Intercept			
Cohort 1	@0		0.59
Cohort 2	-0.44	.013	0.59
Cohort 3	0.14	.399	0.59
Control Cohort	0.10	.534	0.59
Slope (mean)	0.06	.048	0.05
Intervention (mean)	0.67	<.001	0.09
Weaken (mean)	-0.31	.001	@0
Predictor regressions (β)			
General reasoning ability on intercept	.34	<.001	
Condition on intervention	.33	.136	

In order to gauge the effect size of the intervention, it is helpful to visualize the results. As an illustration, Figure 6a shows a standard normal distribution (as required for a PROBIT model) representing the hypothetical algebraic reasoning ability of all students on Problem 1 at the measurement directly before the intervention. The total area under the curve is one and is divided in three parts, separated by so-called thresholds, which reflect the likelihood of reasoning in accordance with Levels R0, R1, and R2 respectively. At the measurement just after the intervention, the algebraic reasoning abilities have changed and the curve in the figure has shifted to the right (see Figure 6b). The thresholds do not change. Due to the intervention, the likelihood of reasoning in accordance with Level R0 decreases, as can be seen in Figure 6, while the likelihood of reasoning in accordance with Level R2 increases after the intervention. In other words, this figure visualizes that after partaking in the teaching sequence students' reasoning improves: fewer students use none of the given equations in their reasoning (Level R0), somewhat fewer students reason on the basis of only one of the given equations (Level R1), and more students combine the information of both equations in their reasoning (Level R2).

The effect size of the intervention determined by the whole test can be computed straightforwardly from the model parameters in Table 3. As explained earlier a score is based on four components: the intercept, the slope, the intervention effect, and the weakening effect. As the intercept only influences the first measurement and the weakening effect is still zero at the measurement directly after the intervention, the effect size of the score gain (Cohen's *d*) is reflected by the sum of the intervention and the slope effect from Table 3: 0.67 + 0.06 = 0.73.

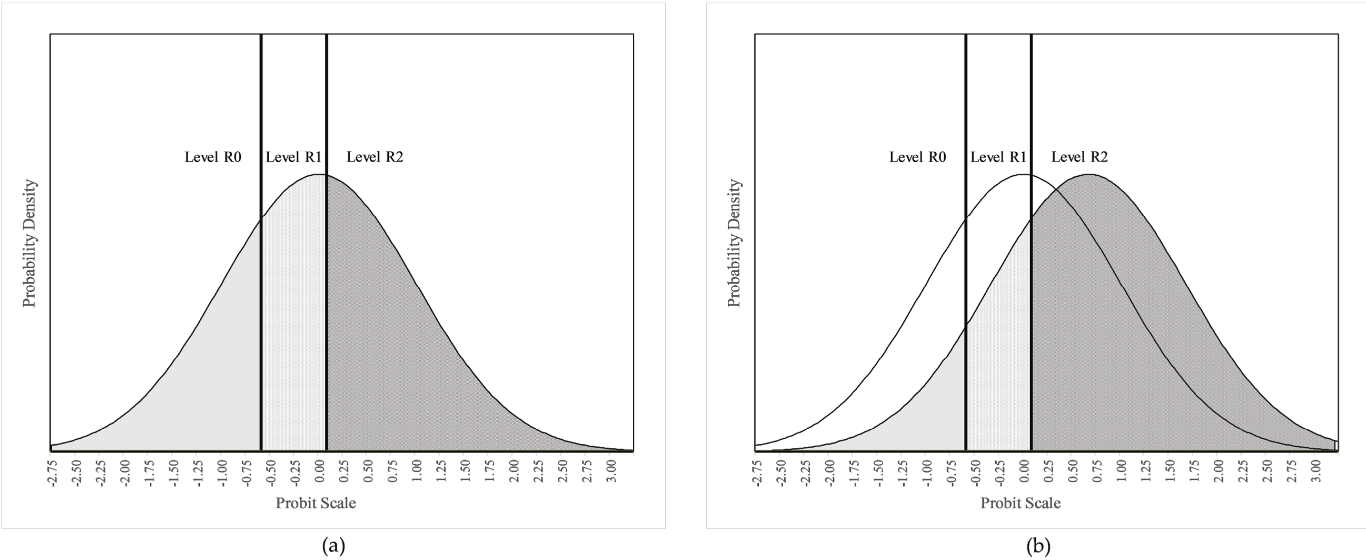


Figure 6. (a) Distribution of students' algebraic reasoning on Problem 1 before the intervention, and (b) Distribution of students' algebraic reasoning on Problem 1 with a shift due to the intervention.

4. Discussion

In this study, we investigated the effects of two learning environments consisting of a teaching sequence with a balance model on the development of primary school students' algebraic reasoning related to linear equation solving. The balance models in these learning environments were implemented in two different representations (dynamic physical vs. static pictorial on paper). The development of students' algebraic reasoning was examined through four assessments over the school year, with students participating in the teaching sequence between two of these measurements. In the assessments, students answered four open-ended tasks, each representing a system of informal linear equations. The sample in this study had no prior instruction on linear equations.

Our results show that fifth-grade students who participated in an intervention based on the balance model showed higher levels of algebraic reasoning when solving systems of informal linear equations. Students improved in their ability to reason by combining the information of both equations (Level R2), instead of reasoning on the basis of only one of the two given equations (Level R1) or none of the given equations (Level R0). This highest level of reasoning was displayed more frequently after the intervention (65%) than before (39%). These results underscore previous research showing that primary school students' algebraic reasoning about (systems of) linear equations can be fostered (e.g., Blanton et al., 2015; Brizuela & Schliemann, 2004; Van Amerom, 2003; Van Reeuwijk, 1995). It moreover underlines the suitability of the balance model for stimulating and structuring this reasoning (e.g., Otten, Van den Heuvel-Panhuizen, & Veldhuis, 2019, see *Chapter 2* of this thesis; Otten, Van den Heuvel-Panhuizen, Veldhuis, & Heinze, 2019, see *Chapter 3* of this thesis; Papadopoulos, 2019; Suh & Moyer, 2007; Warren & Cooper, 2005). In addition, we also systematically investigated the development in reasoning of all students by making use of repeated measurements. This allowed us to examine both the short-term effects of the intervention and its long-term effects: resulting in the finding of a small fading out of the intervention effect.

However, our main interest was whether a static pictorial representation of a balance model had a different effect on the development of students' algebraic reasoning than a physical balance model which students can manipulate and which tilts in response to students' actions. We expected students' perceptual-motor experiences with the physical balance model to be beneficial for their understanding of the abstract mathematical concept of equality in an equation (in line with, e.g., Alessandroni, 2018; Antle et al., 2013; Núñez et al.,1999), which we supposed to positively influence their reasoning about linear equations. Qualitative analyses of students' written responses on the measurement directly after the intervention showed that students who worked with the physical balance model more frequently used the balance model when solving systems of informal linear equations in contexts *not* related to the balance model than students who worked with the model on paper. This use of the model was either explicit, by making use of the representation of the model, or implicit, by making use of substitution of a part of one equation on the basis of the information from the other, or subtracting one equation from the other in order to eliminate unknowns. For students who worked with the physical balance model, 17% used the representation of the model in their reasoning and 60% used advanced algebraic strategies, compared to only 1% and 40% of the students who worked with the model on paper. However, although the descriptive values also suggested a larger improvement in level of reasoning for students who worked with the physical balance model compared to students who worked with the model on paper, the LGM analysis did not yield significant differences in the development of students' levels of algebraic reasoning about equations.

There are several possible explanations for this nonsignificant finding. The teaching sequence in both intervention conditions might have been too similar to affect students' level of algebraic reasoning differently. After all, apart from the used representation, students in both conditions were taught by means of the same didactical model. Such models are meant to elicit students' growth in understanding of mathematics (Van den Heuvel-Panhuizen, 2003). Through the balance model, students in both types of intervention were primed to the equality concept, which is crucial to come up with strategies to solve linear equations and which can assist students to bring the focus on an equation as representing a mathematical structure that links two different algebraic expressions. Possibly, the balance model is a very strong didactical model, which, independent of the representation of the model, is very accessible for students and can help them make sense of the problem situation. Additionally, the difference between both teaching sequences was present mainly in the first three lessons, with students working with only a balance model on paper or in addition a physical balance model. This period of three lessons might have been too short to induce a different effect on students' reasoning. Lastly, it is also possible that *both* representations evoke the idea of balance (in line with Alibali & Nathan, 2012; Núñez et al., 1999; Wilson, 2002), which is strongly grounded in previous physical experiences (Alessandroni, 2018; Gibbs Jr, 2006). Direct perceptual-motor experiences with the physical balance model, or indirect experiences through mental simulation or predicting whether the model on paper will be in or out of balance, can result in the same neural activation patterns (e.g., McCaffrey & Matthews, 2017).

These explanations for the absence of a significant difference between the intervention conditions as regards students' *levels of reasoning* are in contrast with the differences between conditions as regards students' *use of the balance model*, reflected by either their use of a representation of the model and/or their use of advanced algebraic strategies. A possible explanation for this discrepancy is that different representations of the balance model do affect students' algebraic reasoning differently, but only on such a detailed level that our coding scheme was not able to capture these differences. Although our three-level coding scheme proved to be suitable to capture students' level of algebraic reasoning (with a high inter-rater reliability) and although the different levels of reasoning did reflect, to a certain extent, the depth of students' reasoning by the use of algebraic strategies, we, in the end did, lose some of the richness in students' reasoning by means of this straightforward way of coding. Lea's pattern of reasoning provides a good example: although she consistently showed reasoning on the basis of both equations (Level R2), her reasoning after participating in the teaching sequence was much more elaborate and she clearly used more algebraic strategies (or at least she was better able to demonstrate her use of algebraic strategies in her written response). Also, because we did not include think-aloud protocols or other types of live registration of reasoning, we might not have captured the students' full reasoning. We recommend further research to investigate the effects of different representations of the balance model on students' reasoning while making use of live registration of students' reasoning.

Alternatively, because of practical reasons (all lessons were taught by the same teacher) we made use of different cohorts in our study. Within each cohort of each intervention condition we included only one class. This might have resulted in too little power to detect differences in levels of reasoning between conditions using the LGM analysis. The fact that all cohorts of students working with the physical balance model showed higher learning gains could be an indication of this. Ideally, we would have included more than 212 students in our study. This probably would have resulted in a better fitting model and more power to detect potential effects. A design without cohorts or with multiple classes within each cohort would then be preferable to enhance the power of the study.

When interpreting the results of our study, one should also keep in mind its quasi-experimental nature. There was no random assignment of students to conditions. However, we did control for initial differences between classes in our analyses, by including general reasoning ability and general mathematics ability as covariates. Moreover, quasi-experimental designs are considered very appropriate for testing the effectiveness of interventions in natural educational environments (Cook & Campbell, 1979). On the other hand, the study design also has several strengths. First, our staged comparison design with multiple cohorts allowed us to investigate not only the short-term but also the long-term effects of the intervention on students' reasoning. It also created the possibility to take into account the effect of repeatedly assessing students' algebraic reasoning. Second, mixed methods (i.e., quantitative and qualitative analyses) could be used. We went beyond only looking at the correctness of all students' answers, which is often done from a pragmatic point of view, and instead focused on students' reasoning. LGM was subsequently used to model the development of students' algebraic reasoning ability. Lastly, this study could take place in an authentic classroom setting. The high ecological validity makes the results of our study quite easy to translate to educational practice.

In the current study, we demonstrated the effectiveness of a learning environment with the balance model, in a whole-classroom setting, on primary school students' reasoning about solving systems of informal linear equations. Using this model to elicit algebraic reasoning aligns with the objective to commence with stimulating such reasoning in primary school classrooms (NCTM, 2000). No significant differences were found between using a balance model on paper or a physical balance model on the development of students' level of reasoning, suggesting that the representation of the model does not play a role. However, having a closer look at students' reasoning revealed that students who worked with the physical balance model more often made use of the balance model, either by making use of the representation of the model or by making use of algebraic strategies such as substitution or elimination, when solving systems of informal linear equations. This suggests that different representations of the balance model might play a different role in individual learning processes. We recommend for future research to address this discrepancy in findings and to further investigate the possibility that different representations of the balance model might affect students' algebraic reasoning differently, for example by making use of live registration of students' reasoning.

References

Alessandroni, N. (2018). Varieties of embodiment in cognitive science. *Theory & Psychology*, *28*(2), 227–248. https://doi.org/10.1177/0959354317745589

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, *35*(1), 127–145. https://doi.org/10.1037/0012-1649.35.1.127

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

Anthony, G., & Burgess, T. (2014). Solving linear equations. In F. Leung, K. Park, D. Holton, & D. Clarke (Eds.), *Algebra teaching around the world* (pp. 17–37). Sense Publishers.

Antle, A. N., Corness, G., & Bevans, A. (2013). Balancing justice: Comparing whole body and controller-based interaction for an abstract domain. *International Journal of Arts and Technology*, *6*(4), 388–409. https://doi.org/10.1504/IJART.2013.058285

Bajwa, N. P., & Perry, M. (2019). Features of a pan balance that may support students' developing understanding of mathematical equivalence. *Mathematical Thinking and Learning*. Advance online publication. https://doi.org/10.1080/10986065.2020.1700587

Behr, M., Erlwanger, S., & Nichols, E. (1980). How children view the equals sign. *Mathematics Teaching, 92*(1)*,* 13–15.

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354–369. https://doi.org/10.1177/1073191112446655

Blanton, M., Schifter, D., Inge, V., Lofgren, P., Willis, C., Davis, F., & Confrey, J. (2007). Early algebra. In V. J. Katz (Ed.), *Algebra: Gateway to a technological future* (pp. 7–15). The Mathematical Association of America.

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J. S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education*, *46*(1), 39–87. https://doi.org/10.5951/jresematheduc.46.1.0039

Blanton, M. L., & Kaput, J. J. (2005). Characterizing a classroom practice that promotes algebraic reasoning. *Journal for Research in Mathematics Education*, *36*(5), 412–446. https://www.jstor.org/stable/30034944

Bollen, K.A., & Curran, P. J. (2006). *Latent curve models.* Wiley.

Brizuela, B., & Schliemann, A. (2004). Ten-year-old students solving linear equations. *For the Learning of Mathematics, 24*(2), 33–40.

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

Cai, J., Jakabcsin, M. S., & Lane, S. (1996). Assessing students' mathematical communication. *School Science and Mathematics, 96*(5)*,* 238–246. https://doi.org/10.1111/j.1949-8594.1996.tb10235.x

Cai, J., & Knuth, E. (2011). A global dialogue about early algebraization from multiple perspectives. In J. Cai, & E. Knuth (Eds.), *Early algebraization* (pp. vii–x). Springer.

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

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school.* Heineman.

Cheeseman, J., McDonough, A., & Golemac, D. (2017). Investigating children's thinking about suspended balances. *New Zealand Journal of Educational Studies*, *52*(1), 143–158. https://doi.org/10.1007/s40841-016-0073-9

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Rand McNally.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling*. Lawrence Erlbaum Associates.

Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children's understanding of equality: A foundation for algebra. *Teaching Children Mathematics*, *6*(4), 232–236.

Figueira-Sampaio, A. S., Santos, E. E. F., & Carrijo, G. A. (2009). A constructivist computational tool to assist in learning primary school mathematical equations. *Computers & Education, 53*(2), 484–492. https://doi.org/10.1016/j.compedu.2009.03.012

Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction*, *35*, 104–120. https://doi.org/10.1016/j.learninstruc.2014.10.004

Gibbs Jr, R. W. (2006). *Embodiment and cognitive science*. Cambridge University Press.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436–445. https://doi.org/10.2307/798843

Goldenberg, P. E., Mark, J., Kang, J., Fries, M., Carter, C. J., & Cordner, T. (2015). *Making sense of algebra: Developing students' mathematical habits of mind.* Heinemann.

Janssen, J., Scheltens, F., & Kraemer, J. M. (2005). *Leerling- en onderwijsvolgsysteem rekenen-wiskunde [Student monitoring system mathematics]*. Cito.

Jones, I., Inglis, M., Gilmore, C., & Dowens, M. (2012). Substitution and sameness: Two components of a relational conception of the equals sign. *Journal of Experimental Child Psychology, 113*(1), 166–176. https://doi.org/10.1016/j.jecp.2012.05.003

Kaplan, R. G., & Alon, S. (2013). Using technology to teach equivalence. *Teaching Children Mathematics*, *19*(6), 382–389. https://doi.org/10.5951/teacchilmath.19.6.0382

Kaput, J. J., Carraher, D. W., & Blanton, M. L. (2008). *Algebra in the early grades.* Lawrence Erlbaum Associates.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Kieran, C., Pang, J., Schifter, D., & Ng, S. F. (2016). *Early algebra: Research into its nature, its learning, its teaching*. Springer (open access eBook). https://doi.org/10.1007/978-3-319-32258-2

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

Kindt, M., Abels, M, Meyer, M. R., & M. Pligge (1998). Comparing Quantities. In National Center for Research in Mathematical Sciences Education & Freudenthal Institute (Eds.), *Mathematics in Context: A connected curriculum for grades 5–8*. Encyclopedia Brittanica Educational Corporation.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312. https://www.jstor.org/stable/30034852

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guildford Press.

Mann, R. L. (2004). Balancing act: The truth behind the equals sign. *Teaching Children Mathematics*, *11*(2), 65–70.

McCaffrey, T., & Matthews, P. G. (2017). An emoji is worth a thousand variables. *Mathematics Teacher*, *111*(2), 96–102. https://doi.org/10.5951/mathteacher.111.2.0096

McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development, 76*(4), 883–899. https://doi.org/10.1111/j.1467-8624.2005.00884.x

Muthén, L.K. & Muthén, B.O. (1998–2017). *Mplus User's Guide* (Eighth Edition). Muthén & Muthén.

National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics.* NCTM.

Núñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics*, *39*(1–3), 45–65. https://doi.org/10.1023/A:1003759711966

Otten, M., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2019). The balance model for teaching linear equations: a systematic literature review. *International Journal of STEM Education*, *6*(1), 30–51. https://doi.org/10.1186/s40594-019-0183-2

Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, M., & Heinze, A. (2019). Developing algebraic reasoning in primary school using a hanging mobile as a learning supportive tool / El desarrollo del razonamiento algebraico en educación primaria utilizando una balanza como herramienta de apoyo. *Journal for the Study of Education and Development / Infancia y Aprendizaje*, *42*(3), 615–663. https://doi.org/10.1080/02103702.2019.1612137

Papadopoulos, I. (2019). Using mobile puzzles to exhibit certain algebraic habits of mind and demonstrate symbol-sense in primary school students. *The Journal of Mathematical Behavior, 53*, 210–227. https://doi.org/10.1016/j.jmathb.2018.07.001

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's standard progressive matrices and vocabulary scales*. Oxford Psychologists Press.

Smith, J. P., III, & Thompson, P. W. (2008). Quantitative reasoning and the development of algebraic reasoning. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 95–132). Lawrence Erlbaum Associates.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488. https://doi.org/10.3102/00028312033002455

Stephens, A. C., Fonger, N., Strachota, S., Isler, I., Blanton, M., Knuth, E., & Gardiner, A. (2017). A learning progression for elementary students' functional thinking. *Mathematical Thinking and Learning*, *19*(3), 143–166. https://doi.org/10.1080/10986065.2017.1328636

Suh, J., & Moyer, P. S. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching*, *26*(2), 155–173.

Thom, J. S. (2011). Nurturing mathematical reasoning. *Teaching Children Mathematics*, *18*(4), 234–243. https://doi.org/10.5951/teacchilmath.18.4.0234

Van Amerom, B. A. (2003). Focusing on informal strategies when linking arithmetic to early algebra. *Educational Studies in Mathematics*, *54*(1), 63–75. https://doi.org/10.1023/B:EDUC.0000005237.72281.bf

Van den Heuvel-Panhuizen, M. (2003). The didactical use of models in realistic mathematics education: An example from a longitudinal trajectory on percentage. *Educational Studies in Mathematics*, *54*(1), 9–35. https://doi.org/10.1023/B:EDUC.0000005212.03219.dc

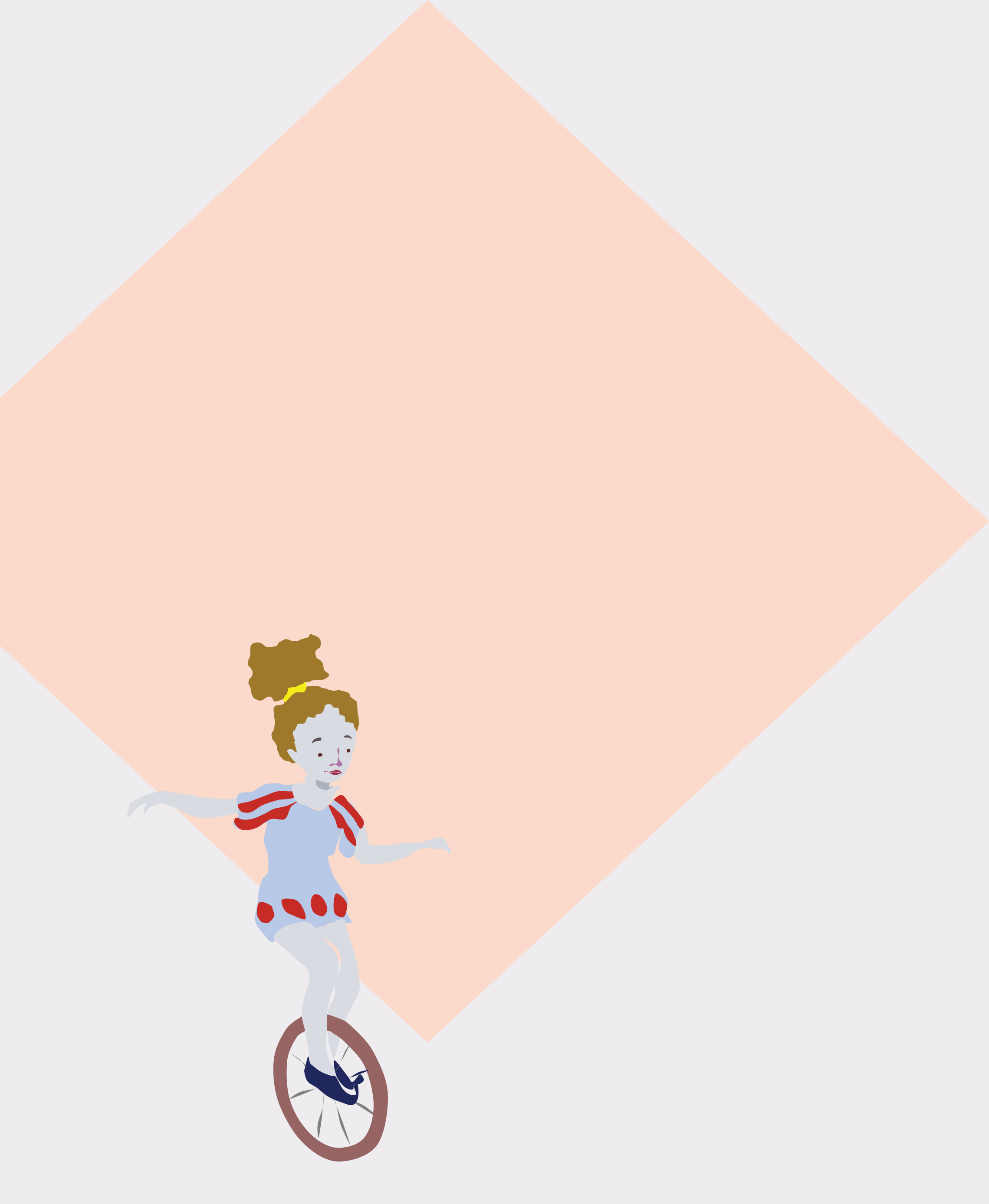
Van Reeuwijk, M. (1995, April). The role of realistic situations in developing tools for solving systems of equations. *Paper presented at the annual meeting of the American Educational Research Association,* San Francisco, CA.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics, 49*(3)*,* 341–359. https://doi.org/10.1023/A:1020229023965

Warren, E., & Cooper, T. J. (2005). Young children's ability to use the balance strategy to solve for unknowns. *Mathematics Education Research Journal, 17*(1), 58–72. https://doi.org/10.1007/BF03217409

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/BF03196322

Author contributions: All authors participated in designing the research. MO gave the lessons and collected the data. MO, MH, MV developed the assessment tasks and the coding scheme. MO, MV coded the data. MO, JB analyzed the data. Methods of data analysis were frequently discussed with all authors. MO prepared the first draft of the manuscript. All authors participated in revising the manuscript and/or provided feedback. All authors read and approved the final manuscript.



CHAPTER 5

Fifth-grade students' reasoning on linear equations and graphs of motion

Mara Otten, Carolien Duijzer, Marja van den Heuvel-Panhuizen, Michiel Veldhuis, Jan Boom, Michiel Doorman, and Paul Leseman

This chapter will be combined into a journal article with a parallel chapter having the focus on graphs of motion, which is published in the PhD thesis of Carolien Duijzer.

Fifth-grade students' reasoning on linear equations and graphs of motion

Abstract

Higher-order thinking (HOT) within the domains of algebra and graphs involves reasoning about covariation. Despite the conceptual overlap between these mathematical domains, little research has been conducted about whether stimulating reasoning within one domain might also be beneficial for reasoning in the other. In this study, we investigated the effect of a six-lesson teaching sequence about linear equation solving on the development of 132 fifth-grade students' reasoning about systems of informal linear equations and graphs of motion over the school year. The mathematical HOT related to reasoning about covariation across both domains was operationalized as *extracting, using, and combining sources of information about relationships*. Results from multi-group latent variable growth curve modelling analyses showed that students improved their algebraic reasoning, but not their graphical reasoning. These findings show that transfer of HOT from one mathematical domain to another, related mathematical domain cannot be taken for granted, suggesting that the HOT that students developed during the lessons was rather domain-specific.

Keywords: Linear equations, Motion graphs, Domain-specific mathematical higher-order thinking

1. Introduction

Lotte buys one pizza and one soda for €10. The next week, she buys three pizzas and two sodas for €27. What is the price of one pizza?[1]

In order to solve this problem, a 10-year old student should reason on the basis of the first relationship (i.e., the price of one pizza and one soda and the total price of €10) and the second relationship (i.e., the price of three pizzas and two sodas and the total price of €27), while keeping in mind how the price of the products and the total price relate to *each other*. This task is far from easy. It involves understanding of how to solve a system of two informal linear equations with two unknowns. For this, the student has to recognize the relationship between the unknowns, meaning that a change in the value of one has consequences for the value of the other. The student also has to understand the connection between the informal equations. This connection entails that both equations include the same unknowns representing the same value (i.e., the price of a pizza in the first equation is similar to the price of a pizza in the second equation), that the values for the unknowns need to be determined in such a way that *both* equations sum up to the correct price, and that both equations in this system of equations are mutually related (in this particular example, the first equation can also be been seen as part of the second equation). We consider 10-year old students' skills to understand and work with a system of informal linear equations, including manipulating the equations, making connections between them, and combining them in novel ways, to be important aspects of mathematical higher-order thinking (HOT) skills (e.g., Goldenberg et al., 2015). This type of task is considered to require HOT due to their non-algorithmic nature whereby students cannot simply recall information or apply routine solving procedures (e.g., Lewis & Smith, 1993; Murray, 2014).

Developing HOT-skills is of major importance in today's rapidly changing society, in which loads of information is available (Forster, 2004; Miri et al., 2007). Therefore, acquiring this HOT is considered to be a major educational goal (Zohar & Schwarter, 2005). There is general agreement that the foundation for HOT should be laid in primary school (National Council of Teachers of Mathematics [NCTM], 2000). Because doing mathematics involves solving problems, such as the problem described above, for which students have to examine ideas, interpret and connect information, and reason logically, mathematics is considered as a suitable domain

[1] Adapted example from one of the algebra tasks used in the current study

for developing HOT (e.g., Foong, 2000; Murray, 2011). Nevertheless, despite the growing attention for incorporating HOT in education, both at the international level (Organization for Economic Co-operation and Development [OECD], 2019) and within the Netherlands, where this study was situated (e.g., curriculum.nu, 2019; Thijs et al., 2014), opportunities for students to show and develop HOT are still almost absent in primary school mathematics textbooks (Kolovou et al., 2009; Van Zanten & Van den Heuvel-Panhuizen, 2018). In two previous studies, we designed, implemented, and evaluated a teaching sequence about linear equations, which stimulated students' HOT related to algebra (Otten et al., 2019, 2020, see *Chapter 3* and *4* of this thesis). In the current study, we focus on whether this teaching sequence also fosters students' HOT in another, related mathematical domain: graphing motion.

HOT in linear equation solving involves reasoning about complex mathematical concepts. An important example of such a mathematical concept is covariation (Thompson & Carlson, 2017). Covariation (or: covarying quantities) is present in in an equation such as $x + y = 10$: when the value of x changes (or: varies), the value of y must vary as well for the sum to remain equal to 10. This notion of covarying quantities is also important in other mathematical domains. One domain for which this particularly holds is the domain of graphing motion. For example, in order to understand a distance-time graph, students have to understand the relationship between the variables of distance and time represented on the graph's axes, make connections between them, and understand how changes in time relate to changes in distance (e.g., Leinhardt et al., 1990). Reasoning about covariation can thus be regarded as domain-specific mathematical HOT, used within the domain of algebra (e.g., reasoning about the relationship between unknowns in an equation or reasoning about the relationship between two equations in a system of equations) or within the domain of graphs (e.g., reasoning about the relationship between the variables on the x- and y-axis). However, at the same time, this covariational reasoning also calls upon processes that are rather similar across mathematical domains (i.e., calls on more general mathematical HOT components): covariation in terms of *extracting, using, and combining sources of information about relationships*.

The aim of the current study was to investigate whether the domain of linear equations offers opportunities to promote reasoning in the domain of graphing motion. We analyzed the effect of a teaching sequence about linear equations on students' reasoning about (systems of informal) linear equations and their reasoning about graphs of motion. The teaching sequence targets students' domain-specific mathematical HOT (e.g., reasoning about relationships between unknown values in equations and making connections between equations in a system of equations), while also touching upon more general components of mathematical HOT (e.g., extracting, using, and combining sources of information about relationships). The results of this study will contribute to the knowledge about whether and how mathematical HOT can be stimulated within and across mathematics domains.

1.1 Nature of higher-order thinking

There is general agreement that lower-order cognition or thinking skills can be distinguished from higher-order cognition or thinking skills (e.g., Lewis & Smith, 1993). The taxonomy of Bloom et al. (1956) is a hierarchy of cognitive skills which is often referred to in the context of HOT. Within this taxonomy, lower-order thinking is reflected by the bottom three levels (knowledge, basic comprehension, and application of basic knowledge) while the upper three levels (analysis, synthesis, and evaluation of information) are often used to operationalize HOT. Newmann (1991) defines lower-order thinking as demanding routine, more or less mechanistic application of previously acquired knowledge. Such application of lower-order thinking takes place when someone recalls information from memory, for example in mathematics, to work on closed problems that can be solved through arithmetic fact retrieval. In contrast, the application of HOT takes place when someone has to organize, interpret, analyze, or manipulate information, to work on a problem which cannot be solved through routine application of previously learned knowledge. According to Resnick (1987), HOT involves "elaborating [on] the given material, making inferences beyond what is explicitly presented, building adequate representations, analyzing and constructing relationships" (p. 45).

Most conceptualizations of HOT, such as the one of Bloom et al. (1956), reflect a domain-general view on HOT. This means that they consider HOT as being about applying general principles such as elaborating, reasoning, analyzing, and synthesizing to any academic domain (e.g., Leighton, 2004). Others argue that HOT cannot be separated from specific academic topics or domains (e.g., Alexander et al., 2011; Ericsson, 2003). This latter view is a domain-specific view on HOT, which entails that HOT stems from — and is entangled with — specific academic domains. In this respect, Alexander et al. (2011) emphasize the necessity of including the role of the domain into the description of HOT, proposing that HOT "exhibits distinctive qualities arising from the nature of the domain within which the task or activity is situated" (p. 53). Following this view on HOT, we can assume that solving (informal) systems of linear equations (for which students have to reason about the relationships between unknowns and about the mutual relationships between equations in order to combine their information) is domain-specific HOT.

1.2 Linear equation solving in primary school mathematics

In an equation, the expressions on both sides of the equal sign represent the same quantity. In that sense, both sides are *equal*, which makes the concept of equality crucial in linear equation solving (Bush & Karp, 2013; Kieran et al., 2016; Knuth et al., 2005; Li et al., 2008). Equality in an equation is often explicated by referring to the metaphor of two sides being "in balance" (Antle et al., 2013). Students need to perceive the equal sign as the symbol denoting an equivalent relationship between qualities (e.g., Behr et al., 1980). However, instead of such a relational understanding of the equal sign, students often have a more operational understanding, meaning that they perceive the equal sign as symbol directing them "to-do-something". This leads to the incorrect belief that the equal sign must always be followed by an answer (e.g., Behr et al., 1980; Falkner et al., 1999; McNeil & Alibali, 2005). A more relational understanding of the equal sign is associated with higher linear equation solving performances (Knuth et al., 2006).

The goal of solving a linear equation is to isolate the unknown so that its value can be determined. The equality of an equation should be maintained during this process (i.e., the equation should remain balanced). Solving an equation implies simplifying the equation by constructing a system of equivalent equations, while keeping the target of this isolating process, the unknown quantity, in mind (Kieran, 1997). The strategy of restructuring the equation, for example, changing the order of the terms in the expressions by using the commutative and associative properties or exchanging the expressions between the two sides of an equation, is used to prepare and simplify the expressions in an equation to facilitate its further solving. Isolation of an unknown can subsequently be achieved by doing transformations under the principle performing the same operations on both sides of the equation (e.g., Arcavi et al., 2016; Pirie & Martin, 1997).

A system of linear equations is a collection of at least two linear equations with the same set of unknowns. When solving such a system of equations, the information of multiple equations needs to be combined. Students need to reason about the relationships between the different unknowns and their pattern of covariation (i.e., how changes in the one result in changes in the other). When the value of one unknown varies, it implies that the value of the other unknown must vary as well to make sure that the validity of every equation in the system is satisfied. In order to solve a system of linear equations, the strategy of substitution is often used, whereby an expression in an equation is replaced by another expression of the same value or by another particular value.

1.3 Domain-specific HOT in linear equation solving and graphs of motion

Mathematical functions (e.g., a relationship between x and y) can be represented in different ways, for example algebraically (in the form of an equation) or graphically (in the form of a graph). In mathematics education, the introduction of algebraic and graphical representations of a function represents "one of the earliest points in mathematics at which a student uses one symbolic system to expand and understand another" (Leinhardt et al., 1990, p. 2). Students need to learn, among others, how to create equations to represent relationships between quantities, how to graphs these equations, and how to solve (systems of) equations algebraically and graphically (Common Core State Standards Initiative [CCSSI], 2010). The mathematical domains of algebra and graphs are often addressed together in mathematics education. In various research articles the relationship between these domains has been a topic of interest. In a study by Schliemann et al. (2012), in which third- to fifth-grade students were introduced to algebra, students explored relationships between variables through representing them as graphs and as equations and solving them algebraically. In another study, students solved a system of two informal linear equations by making use of algebraic strategies such as substitution, or by making use of the representation of a graph (Berks & Vlasnik, 2014).

In the current research, we were also interested in developing students' algebraic as well as students' graphical reasoning. However, instead of addressing both domains in the same intervention, we took another approach: we investigated whether a teaching sequence in the domain of linear equations would also have an effect on students' reasoning about graphs of motion. Solving (a system of) informal linear equations requires reasoning about how changes in one unknown result in changes in the other (i.e., reasoning about covariation; Thompson & Carlson, 2017). Such a system of linear equations can be seen as a complex and demanding problem for which there is no fixed solution procedure (Papadopoulos, 2019). Various skills are called upon, such as making comparisons between both linear equations and integrating the information found in both of them (in the terminology of Bloom et al., 1956: synthesizing); skills which can be considered HOT skills. Flexibility to switch between multiple representations is moreover necessary, such as pictorial or symbolic representations of the equations, or other ways of representing the data, such as a table or a graph.

Within the domain of motion graphs, reasoning about graphically represented change is essential. This reasoning about graphically represented change implies that a student should be able to reason about the values on the *x*-axis and *y*-axis of the graph, the respective quantities as well as their pattern of covariation. Graphical reasoning involves both graph interpretation and graph construction. Consider the following example: *Ollie and Eve are going to school. Eve leaves home a little earlier than Ollie. Halfway she waits for Ollie to catch up. They continue their journey together and arrive at the same time.*[2] To construct a distance-time graph of this particular motion situation, a student needs to have an understanding of how the relationship between time and distance can be represented as a line in the graph. When drawing a line-graph, students should be able to simultaneously coordinate the relationship between two changing variables with the graph's axes as a reference.

Covariation can thus be considered a core concept within both the domains of linear equations and motion graphs. This covariational reasoning can be regarded as domain-specific mathematical HOT: within algebra it involves reasoning about the relationship between unknowns or reasoning about the relationships between equations, while within graphs it involves reasoning about the relationship between the presented variables on both axes. Given this domain-specific nature of the concept of covariation, we cannot automatically assume that reasoning about covariation within one domain is similar to reasoning about covariation within another domain. However, we can describe this essentially domain-specific concept as also involving more general mathematical HOT skills, occurring within both mathematical domains: reasoning about extracting, using, and combining sources of information about relationships. For example, students can extract information about linear relationships from the system of equations, take into account the interrelatedness, and combine the information to find unknown values or new relationships. Likewise, students can extract information about quantities presented on a graphs' axes, take into account the interrelatedness, and combine the given quantities into something new. Given the importance of the concept of covariation across both domains, it is worthwhile to investigate whether stimulating reasoning

[2] Adapted example from one of the graphing tasks in the *Beyond Flatland* project (see *Chapter 1* of this thesis).

about covariation within one domain might also improve reasoning in the other domain.

Involving students in challenging domain-specific mathematics activities could open up possibilities for stimulating their HOT. For example, the domain of linear equations offers a fruitful starting point for such activities which are currently not included in the primary school mathematics curriculum. To this end, we developed a six-lesson teaching sequence in which fifth-graders explored, reasoned about, and solved (systems of) linear equations in informal contexts. In a previous study, we found that the activities in this teaching sequence stimulated students' use of algebraic strategies which resulted in higher levels of algebraic reasoning (Otten et al., 2019, see *Chapter 3* of this thesis). Moreover, students participating in this teaching sequence showed more improvement in their algebraic reasoning than students participating in an intervention on another mathematical topic, probability (Otten et al., 2020, see *Chapter 4* of this thesis).

1.4 The current study

In the current study, we investigated the effect of receiving lessons in linear equations on reasoning about graphs of motion. In parallel, another study was carried out to investigate the possible transfer of receiving lessons on graphs of motion towards students' reasoning about linear equations (Duijzer, Otten, et al., 2020). Because reasoning about covariation is important in both mathematical domains, we assume that stimulating students' reasoning in one domain might affect their reasoning in the other domain. The following research question was formulated: *To what extent does a six-lesson teaching sequence on linear equation solving affect students' algebraic and graphical reasoning?*

Two sets of four tasks were used to assess primary school students' development of algebraic and graphical reasoning. On the four algebra tasks, students had to reason about two informal linear equations. On the four graphing tasks, students had to reason about two changing variables (presented on the x- and y-axis of a graph) or they had to construct a graph of a relationship between two changing variables. Understanding of, or reasoning about, the underlying relationship between the covarying variables is required to solve these mathematical problems in both domains. We expected that participating in the teaching sequence on linear equation solving would not only have a positive effect on students' ability to reason about linear equations, but also on students' ability to reason about graphs of motion. A possible effect of the teaching sequence on students' graphical reasoning will give us more insight in the extent to which stimulating students' domain-specific mathematical reasoning can also support students' reasoning within a different, yet related domain of mathematics

2. Method

2.1 Participants

Participants were 142 students of six classes from six different Dutch elementary schools. All schools and classes participated on a voluntary basis. Data from five students were excluded because there was no parental consent, and from five other students the data were excluded because these students were absent in more than half of the lessons. The final sample consisted of 132 students (45% boys), with an average age of 10 (age range 9-11).

2.2 Research design and procedure

The study was approved by the ethical committee of the Faculty of Social and Behavioural Sciences of Utrecht University. A staged comparison design with three different cohorts was used (see Table 1), allowing that the same teacher taught all lessons to all classes. The six-lesson teaching sequence on linear equation solving was provided to the students in different time periods during the school year (see Table 1): students of the two classes in Cohort 1 received the lessons in the beginning of the school year, students of the two classes in Cohort 2 in the middle of the school year, and students of the final two classes of Cohort 3 in the end of the school year. During the teaching sequence, students received one lesson of approximately 50 minutes per week over a period of six weeks. All lessons were taught by the first author of this paper, with the first two lessons assisted by a teaching assistant.

Table 1
Staged comparison design

Cohort		Phase						
			Nov.–Dec. 2016		Feb.–March 2017		May–June 2017	
1	(n = 44)	M1	*Lessons on linear equations*	M2		M3		M4
2	(n = 39)	M1		M2	*Lessons on linear equations*	M3		M4
3	(n = 49)	M1		M2		M3	*Lessons on linear equations*	M4

Note. M1-M4 reflect Measurements 1-4

2.3 Teaching sequence

The aim of the teaching sequence was to foster students’ algebraic reasoning. Students were provided with a learning environment in which they were able to invent, in an informal way, the strategies that are fundamental for linear equation solving (see also Otten et al., 2019, 2020, see *Chapter 3* and *4* of this thesis). Over the course of this teaching sequence, students could further develop these strategies and use them to solve (informal) linear equations in various contexts. The teaching sequence could be divided into four episodes based on the focus and content of the lessons. In the first episode, students were reasoning about one informal equation, represented as a balance model. Through exploration of various ways to maintain the balance of this model, which can be seen as representing equality in an equation, students could develop informal context-connected algebraic strategies which underlie conventional equation solving strategies: restructuring, isolation, and substitution. From Episode 2 onwards, students could use these strategies for solving systems of linear equations, first in the familiar informal context of the balance (Episode 2), then in new informal contexts (Episode 3), and finally students’ had to solve a system of two linear equations represented in formal algebraic symbols (Episode 4). Table 2 provides an overview of the teaching sequence, including the key activities for each of the episodes. Students’ reasoning about linear equation solving was constantly stimulated by asking them to hypothesize and evaluate their own ideas and to discuss and compare them with their peers in small groups and during classroom discussions.

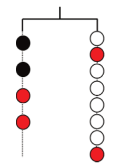
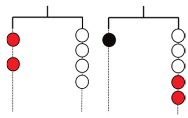
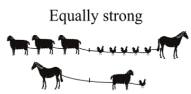
Two parallel versions of the teaching sequence were created, which differed regarding how the balance model was represented and used during the lessons: a static, pictorial representation of the balance model versus a pictorial representation combined with a dynamic physical balance model. In a previous study (Otten et al., 2020, see *Chapter 4* of this thesis), no significant differences in the development of students’ level of algebraic reasoning were found between these two versions of the teaching sequence. The differences between these two versions of the teaching sequence were therefore disregarded in the current study.

2.4 Measures

2.4.1 General reasoning ability

As a measure of students’ general reasoning ability an abbreviated version of Raven’s Standard Progressive Matrices (Raven SPM; Raven et al., 1996) was administered, containing 18 items with increasing difficulty (Bilker et al., 2012). Each item consists of a diagram with a missing part. Students have to reason which part is missing and select this missing part among six or eight possibilities. Answers can either be correct (1) or incorrect (0), resulting in possible scores between 0 and 18.

Table 2
Overview of the six-lesson teaching sequence on linear equation solving

Lesson	*Main topics* and activities	
1+2	*Reasoning about one informal linear equation:* - Maintain the equality of one informal linear equation (represented as balance model) - Find relationships between unknowns - Use informal notations - Construct informal equations	
3	*Reasoning about a system of two informal linear equations:* - Maintain the equality of both informal linear equations (represented as balance model) - Combine the equations to find relationships between unknowns - Use informal notations	
4	*Reasoning about a system of two informal linear equations in a new, informal context:* - Maintain the equality of both informal linear equations (represented in new informal contexts such as a tug-of-war situation) - Combine the equations to find relationships between unknowns - Gradually use more formal notations	
5+6	*Reasoning about a system of two formal linear equations:* - Maintain the equality of both linear equations (represented in formal symbols) - Combine the equations to find values of unknowns Gradually use more formal notations	$2N+4W+N=36$ $2W=3N$

2.4.2 General mathematics performance

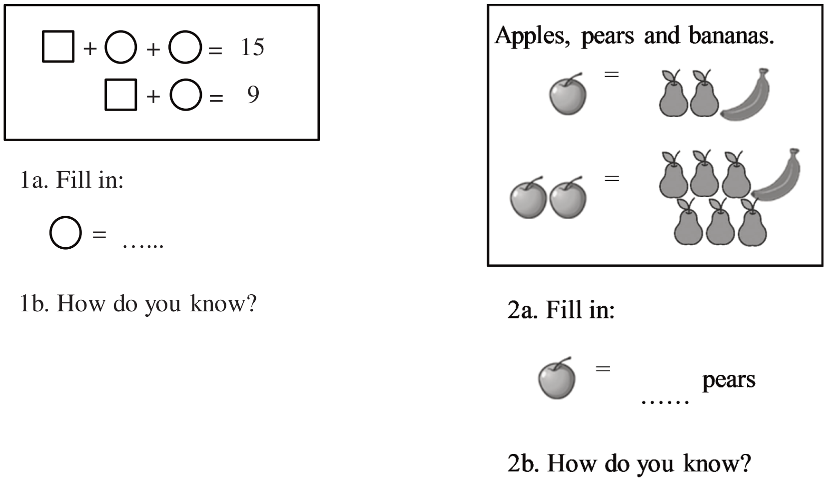
As a measure of students' general mathematics performance students' end-term Grade 4 scores on the mathematics test of the CITO Monitoring System were obtained, a Dutch standardized student monitoring test for different subjects and grade levels (Janssen et al., 2005).

2.4.3 Mathematical HOT

Mathematical HOT was measured by four paper-and-pencil algebra tasks, used as an indication of their algebraic reasoning, and four paper-and-pencil graphing tasks, used as an indication of their graphical reasoning. Examples of these tasks are provided below (Figures 1 and 2). All tasks were open-ended. By asking the question "*How do you know*?" students were explicitly invited to elaborate on their reasoning.

2.4.3.1 Algebra tasks

The set of algebra tasks consisted of four tasks in which students had to solve a system of linear equations. In two tasks students had to find values of unknowns; in two other tasks a relationship between unknowns was asked for. In order to find the correct solution to each of the tasks, students had to combine the information of two given equations. Figure 1 shows two examples of the algebra tasks. Task 1 (Figure 1, left panel) represents a system of two informal linear equations with two unknowns; the value of one unknown (the circle) needs to be determined. Both equations contain circles and squares — of which the values are unknown — on the one side of the equal sign, and values on the other side of the equal sign. The value of the circle needs to be determined, which can be accomplished by isolation of this unknown. This can be done by substitution of a part of the first equations by a value (i.e., substitution of one square and one circle in the first equation by the value of 9, on the basis of the second equation), by subtracting the second equation from the first one, or by the substitution of different values for both unknowns in both equations. Task 2 (Figure 1, right panel) represents a system of two informal linear equations with three unknowns. Both equations contain apples, pears, and bananas — of which the values are unknown — on both sides of the equal sign. The unknown relationship between apples and pears needs to be determined. This relationship can be deduced by isolating the apple. Isolating one apple can be accomplished by substituting unknowns in the second equation by other unknowns (on the basis of the first equation), by subtracting the first equation from the second one, by doubling the first equation and then comparing it with the second one, or by the substitution of different values for all unknowns in both equations. Thus, in order to solve these two tasks, consisting of a system of informal linear equations, students should take into account the relationships represented in both equations and combine this information in their reasoning. We consider this reasoning to reflect the mathematical HOT component extracting, using, and combining sources of information about relationships.



Example solution:
“Square + circle = 9. Then I look how many I need, which is 15. There is one additional circle. So you what I do: square + circle + circle = 15, 9 + circle = 15. 9 + 6 = 15. Ready”

Example solution:
“2 apples is twice as much. So 4 pears and 2 bananas. But 1 banana is missing and there are 2 additional pears. So 1 banana = 2 pears. So 2 pears makes four pears up there, the banana must go and two pears can be added”

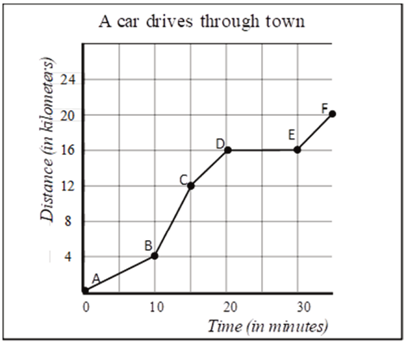
Figure 1. Two algebra tasks (translated from Dutch). The complete coding scheme, including examples of student responses for each task, can be found in Otten et al. (2020), see *Chapter 4* of this thesis.

2.4.3.2 Algebra tasks: combining sources of information

An example solution for the first algebra task is provided in Figure 1 (bottom left). Both linear equations are incorporated in the reasoning of this student. First, this student compares both equations to see on which aspects they differ. Subsequently she combines the information from both equations through substitution of two unknowns in the first equation by the value of 9 (on the basis of the second equation), so that the unknown circle is isolated and its value can be determined by means of a simple calculation (9 + circle = 15). Figure 1 (bottom right) provides an example solution for the second algebra task. After comparing both equations, this student concludes that the two apples in the second equations are twice as many apples as the number of apples depicted in the first equation. Therefore, the right part of the first equation is doubled by this student as well. Subsequently, this student compares this "newly" constructed equation with the second equation in which there are two extra pears and one banana less, resulting in the conclusion "1 banana = 2 pears". Lastly, this finding is used to substitute the banana in the first equation by two pears. The students in both these examples combine the information from both given equations, and thus show mathematical HOT in terms of extracting, using, and combining sources of information about relationships.

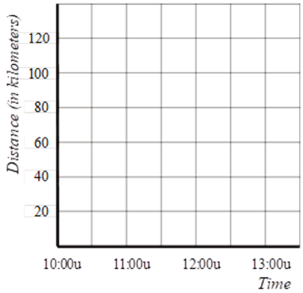
2.4.3.3 Graphing tasks

The set of graphing tasks consisted of four tasks in which students either needed to interpret a distance-time graph or to construct one. In all four tasks, students had to look at and combine the information from the *x*-axis and the *y*-axis in order to find the correct solution. Figure 2 provides two examples of these graphing tasks. Task 1 (Figure 2, left panel) shows a distance-time graph representing the movement of a car driving through town. Students should derive at which moment the car goes fastest. The speed of the car (i.e., the hidden quantity) can be deduced by a visual inspection of the slope of the line or by combining the values presented on the *x*-axis and *y*-axis (i.e., by looking at the distance traveled within a period of time and compare this with the other segments of the graph). Task 2 (Figure 2, right panel) shows an empty distance-time graph, combined with a description of a motion situation of a train ride. This motion situation consists of three separate segments in which the train travels at different speeds. These different speeds should be quantified and visualized in the graph, while taking into account the relative differences in speed between the three "parts" of the train ride. Thus, in order to find the correct answer for the graph interpretation task or to construct the distance-time graph students should take into account the meaning of the variables on both the *x*-axis and the *y*-axis and compare various segments within the graph.



1a. Between which points does the car go fastest?
1b. How do you know?

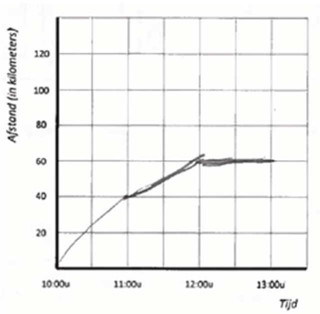
A train ride.
A train travels **twice as fast** between **10:00 and 11:00** o'clock than between **11:00 and 12:00** o'clock. The train stands still from **12:00 to 13:00** o'clock.



2a. Draw a graph that fits the description above.
2b. How do you know?

Example solution:
"He goes from 4 to 12 in 5 minutes, and nowhere else he goes faster than that"

Example solution:



"First you do two squares, and then 1. Because 1 is half of 2. And then he stands still for one hour, so straight"

Figure 2. Two graphing tasks (translated from Dutch). The complete coding scheme, including examples of student responses for each task, can be found in Duijzer, Van den Heuvel-Panhuizen et al., (2020).

2.4.3.4 Graphing tasks: combining sources of information

An example solution for the first graphing task is provided in Figure 2 (bottom left). The student states that the car goes from "4 to 12 in 5 minutes", referring to points B and C. This student thus connects the quantities represented on both axes (i.e., time and distance) to draw a conclusion about the speed of the car. Subsequently, a comparison with the other features of the graph is made by stating that "nowhere else he goes faster than that". Figure 2 (bottom right) provides an example solution for the graph construction task. This solution is a correct translation of the motion situation: the student connects the variables represented on both axes in the correct way and takes into account the relative differences between the segments of the graph (which is also indicated by the explanation: "First you do two squares, and then 1. Because 1 is half of 2"). The students in both these examples combine the information found on the *x*-axes and *y*-axes of the graph, and thus show mathematical HOT in terms of extracting, using, and combining sources of information about relationships.

2.5 Coding schemes for students' reasoning showing HOT

Separate (domain-specific) coding schemes were developed for both the algebra tasks (see also Otten et al., 2019, 2020, see *Chapter 3* and *4* of this thesis) and the graphing tasks (see also Duijzer et al., 2019; Duijzer, Van den Heuvel-Panhuizen, et al., 2020), which shared the common structure of qualifying reasoning in terms of increasing levels of complexity. For the algebra tasks, students' reasoning was coded on the basis of the number of used equations. A distinction was made between students who reasoned about none of the given linear equations that were given in the system of equations (i.e., showed no algebraic reasoning, Level R0), students who reasoned on the basis of only one of the two given equations (Level R1), and students who reasoned on the basis of both given equations by combining the information of both of them (Level R2). This latter, highest level of reasoning can be considered equivalent to the mathematical HOT skill of extracting, using, and combining sources of information about relationships. A similar distinction can be found in the coding scheme used to indicate students' graphical reasoning. Here, a distinction was made between students who took none of the variables into account in their reasoning (i.e., showed no apparent graphical reasoning; Level $R0_1$) or students who reasoned only on the basis of iconic or superficial characteristics of the graph (Level $R0_2$), students who reasoned on the basis of only one of the variables (distance, time, or speed; Level R1), and students who reasoned on the basis of multiple variables (distance, time, and/or speed; Level R2). Also this latter, highest level of reasoning can be considered equivalent to the mathematical HOT skill of extracting, using, and combining sources of information about relationships. Table 3 shows the alignment between both coding schemes and the overarching HOT.

Table 3
Alignment of the coding schemes of algebra and graphs in relation to the mathematical HOT extracting, using, and combining sources of information about relationships

Code	Algebra	Graphs	HOT
$R0_1$	No reasoning	No reasoning	No reasoning
$R0_2$		Iconic/superficial reasoning	
R1	Reasoning on the basis of one equation	Reasoning with a single variable	Reasoning taking into account one source of information
R2	Reasoning on the basis of two equations	Reasoning with multiple variables	Reasoning taking into account more than one source of information

2.6 Data analysis
2.6.1 Descriptive statistics
Proportions of each level of algebraic reasoning (levels R0, R1, R2) and graphical reasoning ($R0_{1,}$ $R0_2$, R1, R2) on the mathematical HOT test were calculated for each cohort on each of the four measurements.

2.6.2 Latent variable growth curve modelling
Students' development of algebraic reasoning and graphical reasoning over the four measurements was modelled by making use of latent variable growth curve modeling (LGM). LGM is suitable for modelling change over repeated measures (Bollen & Curran, 2006) and offers more possibilities for modelling longitudinal data compared to traditional statistical analyses (Willett & Bub, 2005). The core of a LGM analysis is the assumption that each person has underlying latent abilities which differ between persons and which can change within persons over time due to, for example, experience with certain tasks or participating in an intervention. These inter- and intra-individual differences can be modelled by using LGM.

A cohort sequential multi-group LGM (Duncan et al., 2006) was used in this study, with the three cohorts as groups. We built an integrated LGM to model both latent abilities (i.e., algebraic and graphical reasoning abilities) in *one* model. Such a LGM model allowed us to model students' development in latent algebraic reasoning and graphical reasoning over time and to investigate the relationship between these abilities. Both latent abilities were modelled as a combination of four partial effects. The *intercept* (students' baseline level of reasoning), the *slope* (a linear change over measurements), the *intervention effect* (the effect of the intervention, influencing only the measurements after the intervention), and the *weakening effect* (any decrease of the effect of the intervention, influencing only the delayed measurements after the intervention). This integrated LGM model thus contained two intercepts, two slope effects, two intervention effects, and two weakening effects, for algebraic reasoning and graphical reasoning ability respectively.

In addition, we had to take the different cohorts in our study into account. Because the three cohorts differed in the timing of the intervention, loadings for the intervention and weakening effects were different for all three cohorts. For example, for Cohort 1, in which the intervention took place between measurements M1 and M2, the intervention effect only had a possible influence on students' reasoning on measurements after the intervention (thus on M2, M3, and M4), and the weakening effect for this cohort only had a possible a possible influence on students' reasoning on the delayed measurements after the intervention (M3 and M4). Moreover, to take any possible differences between cohorts in students' baseline levels of reasoning (both algebraic and graphical) into account, differences between the intercepts of the different cohorts were allowed. All other parameters were constrained to be equal in all cohorts. Measures of general reasoning ability and general mathematics performance were added in a centered form as a predictor of the intercepts (both for algebraic and graphical reasoning). However, adding general mathematics performance resulted in problems with the model estimation, because of the high correlation between general reasoning ability and general mathematics performance. Therefore, we decided to only include general reasoning ability as a predictor in the LGM model.

In order to deal with our categorical data, item response theory (IRT) transformation was used to map the likelihood of displaying a certain level of reasoning (i.e., Level R0, R1, or R2 for algebraic reasoning and Level $R0_{1,}$ $R0_2$, R1, or R2 for graphical reasoning) onto students' latent algebraic and graphical reasoning abilities. IRT models estimate the probability of reasoning at a particular level as a function of the difficulty of the task and the latent ability of the student. In order for the model to be estimated, for each of the three cohorts, all levels of reasoning on all four tasks on all four measurements should appear at least once (i.e., there should be no "empty cells"). For graphing Task 2 this prerequisite was not met as empty cells were found for Level R2. We therefore decided to merge Level R2 with Level R1 for this task. Also for graphing Task 3 there were some

5

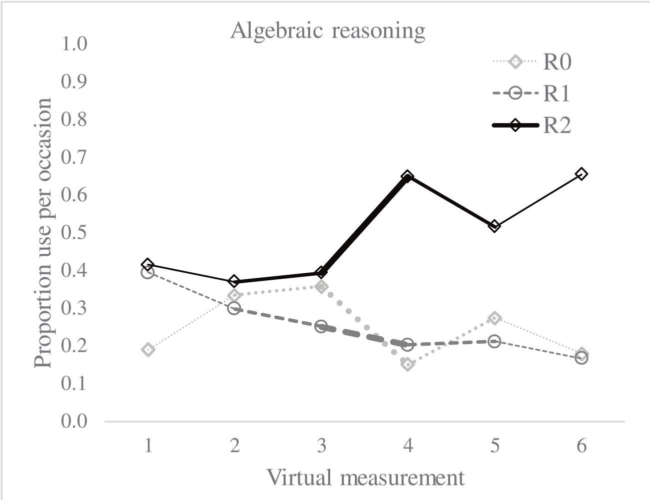
empty cells for Level $R0_1$. This level was therefore merged with the other R0 level, Level $R0_2$.

The model was estimated in Mplus 8 (Muthén & Muthén, 1998-2017) with a PROBIT link and Weighted Least Squares Means and Variances (WLS-MV) estimation. The advantage of using a PROBIT link is that parameter estimates are expressed in units representing standard deviations, which refer to a standard normal distribution with a mean of zero. Therefore, the parameter estimates can be directly interpreted as standard effect sizes. The overall model fit was evaluated in terms of the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI). Conventional recommendations as regards the criteria for an acceptable model fit imply an RMSEA below .08 and a CFI and a TLI above .90 (Little, 2013).

3. Results

3.1 Descriptives

There were no differences between Cohort 1, ($M = 10.57$, $SD = 2.51$), Cohort 2, ($M = 10.05$, $SD = 2.65$), and Cohort 3, ($M = 10.51$, $SD = 2.79$) in students' general reasoning ability, $F(2, 129) = 0.47$, $p = .629$, partial $\eta^2 = .007$.



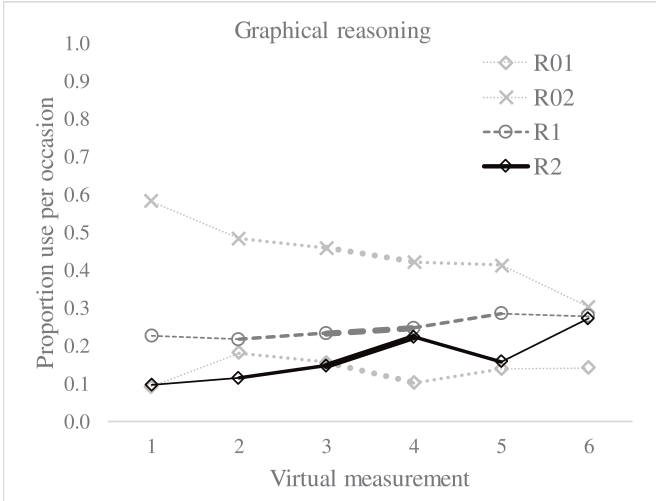


Figure 3. Proportions of level of algebraic reasoning (R0, R1, R2; left panel) and graphical reasoning ($R0_1$, $R0_2$, R1, R2; right panel) on each virtual measurement. The intervention took place between virtual measurements 3 and 4. Thin lines reflect the scores of students of one cohort, thicker lines of two cohorts, and the thickest lines are based on the scores of all three cohorts.

5

Proportions of each level of reasoning on the four algebraic reasoning problems and on the four graphical reasoning problems were calculated for each of the four measurement moments. To allow for direct (visual) comparison of the change in students' algebraic and graphical reasoning for all cohorts together (which differed in timing of the received intervention), six virtual measurements were created. This means that the students in Cohort 1 were depicted as having participated in virtual measurements 3-6, students in Cohort 2 as having participated in virtual measurements 2-5, and students in Cohort 3 as having participated in virtual measurements 1-4. In this way, virtual measurement 3 is identified as the measurement before the intervention (i.e., the pretest) and virtual measurement 4 as the measurement directly after the intervention (i.e., the direct posttest). The change from virtual measurement 3 to 4 thus reflects the change in reasoning due to the intervention, visualized for all cohorts together. Figure 3 shows the proportions of each level of algebraic reasoning (left panel) and graphical reasoning (right panel) on each of the six virtual measurements. When comparing students' algebraic reasoning on the pretest and on the direct posttest, reasoning on the basis of none of the equations (Level R0) or only one of the equations (Level R1) decreased (.21 and .05 decrease in proportion respectively), while reasoning on the basis of both given equations increased (.26 increase in proportion). For graphical reasoning, all proportions of reasoning remained more or less stable from pretest to direct posttest: reasoning while not incorporating any of the variables slightly decreased (.06 and .04 decrease in proportion for $R0_1$ and $R0_2$ respectively), reasoning on the basis of one variable (Level R1) slightly increased (.02 increase in proportion), and reasoning with multiple variables (Level R2) also increased a little (.08 increase in proportion).

3.2 Multi-group LGM

A multi-group LGM with the three cohorts as groups was fitted to students' algebraic reasoning and graphical reasoning scores to investigate whether our teaching sequence on linear equation solving affected students' development in algebraic reasoning and graphical reasoning. The model had an acceptable overall fit in terms of the RMSEA (.073, 90%-CI [.062 - .083]). However, the fit of the model in terms of the CFI (.772) and TLI (.795) was insufficient. Additional model options were explored. These explorations were guided both by the modification indices which Mplus provided, including allowing correlations between partial effects and between specific tasks on specific measurements. While these explorations did not result in clear improvements of the CFI and TLI fit indices, they did reveal robustness of the relevant parameter estimates. We suppose that the small sample size per cohort (44, 39, and 49 students for Cohorts 1-3, respectively) in combination with the strict assumptions of our model (necessary to test our hypotheses) made it difficult to obtain a better overall model fit. However, because commonly applied cut-off criteria for the model fit do not necessarily generalize to longitudinal or multi-group models (Little, 2013), and because the fit of our model was sufficient in terms of the RMSEA, which is the most reliable fit measure when it concerns categorical data (Little, 2013), we considered our model and corresponding parameter estimates informative and trustworthy.

Table 4

Parameter estimates of the final multi-group LGM model

	Algebra			Graphs		
Model parameter	*M*	*p*-value	*var*	*M*	*p*-value	*var*
Intercept						
Cohort 1	@0		0.49	@0		0.02
Cohort 2	-0.12	.431	0.49	-0.13	.007	0.02
Cohort 3	0.57	<.001	0.49	-0.07	.127	0.02
Slope	-0.03	.552	0.02	0.03	.017	0.05
Intervention	0.93	<.001	0.08	0.04	.097	0.01
Weakening	-0.10	.338	@0	-0.08	.013	@0
Predictor regression (β)						
General reasoning ability on Intercept	.32	<.001		.27	<.001	
Covariance (β)						
Intercept algebra with intercept graphs	.79	<.001				

Table 4 shows the parameter estimates of our model. The positive effect of our intervention on students' algebraic reasoning was strong and significant ($M = 0.93$, $p < .001$). However, the effect of our intervention on students' graphical reasoning was small and non-significant ($M = 0.04$, $p = .096$). Recall that, due to the use of the Probit model, these values can be interpreted as standard effect sizes.

Figure 4 visualizes the effect of the intervention on students' algebraic (left panel) and graphical (right panel) reasoning, on algebra Task 1 and graphing Task 1. The left panel of Figure 4 shows two standard normal distributions. The left distribution of the left panel represents students' latent *algebraic* reasoning ability on the measurement directly *before* the intervention. The total area under the curve is divided into three parts, referring to each of the three levels of reasoning, which are separated by the so called thresholds. The surface of each area (i.e., of each level of reasoning) is proportional to the likelihood of reasoning in accordance with that level. The second curve in the left panel of Figure 4 (the curve on the right side) shows students' algebraic reasoning ability directly *after* participating in the intervention. This curve has shifted to the right, which means that on average students' latent algebraic reasoning abilities are higher. Because the thresholds remain stable, the changes of reasoning in accordance with Level R0 decrease while the likelihood of reasoning in accordance with Level R2 increases. The right panel of Figure 4 shows also shows two standard normal distributions: one reflecting students' latent *graphical* reasoning ability on the measurement directly before the intervention (left distribution) and one reflecting this graphical reasoning ability directly after the intervention (right distribution). The fact that the difference between these two distributions is almost imperceptible shows that the effect of our intervention on students' graphical reasoning is negligible.

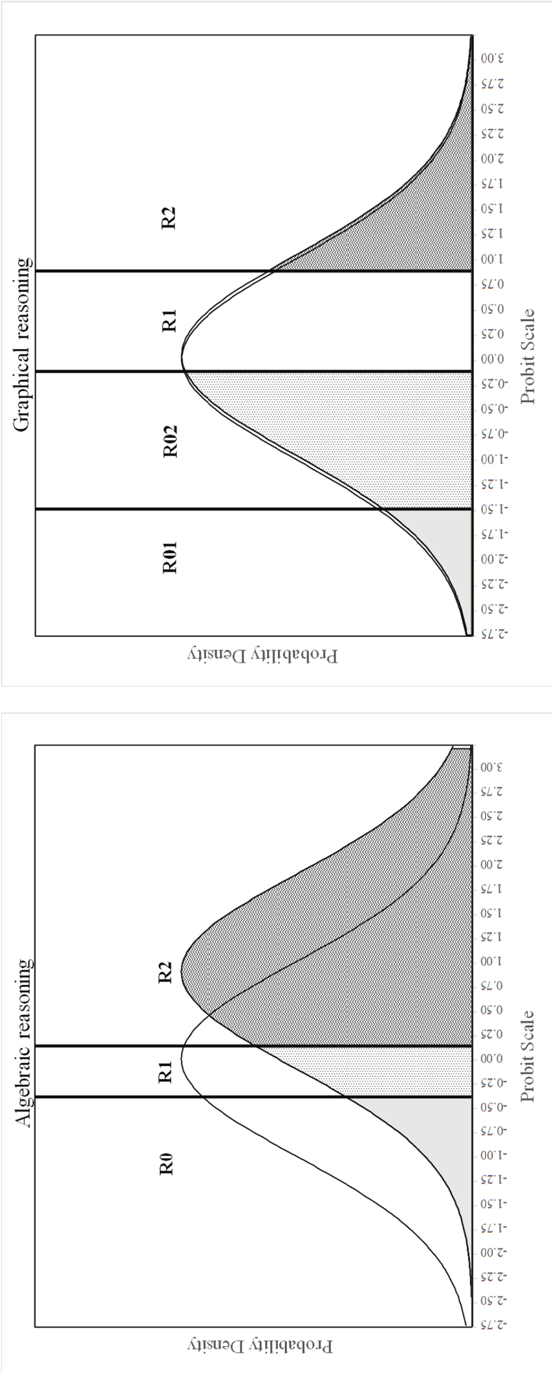


Figure 4. The effect of the intervention on students' algebraic reasoning (left panel) and their graphical reasoning (right panel).

5

4. Discussion

In this study, we focused on the development of primary school students' mathematical HOT across two distinct but related mathematical domains. We investigated whether a six-lesson teaching sequence about linear equation solving would affect not only students' algebraic reasoning but also students' reasoning about graphs of motion. Results from our multi-group latent variable growth curve modelling analyses showed that fifth-grade students' algebraic reasoning related to solving systems of linear equations improved after participating in the teaching sequence, which was a confirmation of the results of a previously conducted study (Otten et al., 2020, see *Chapter 4* of this thesis). Even though reasoning about covariation characterizes HOT in both domains, participating in this teaching sequence, however, did not seem to result in an improvement of students' graphical reasoning. The HOT that the students developed in the lessons on linear equation solving rather seemed to be domain-specific and linked to the mathematical domain of algebra (Alexander et al., 2011, Ericsson, 2003). Hence, our teaching sequence did not structurally seem to elicit elements of more general mathematical HOT, relevant to both mathematical domains (which in this study was operationalized as extracting, using, and combining sources of information about relationships).

It thus appears that the transfer of HOT from one mathematical domain to another related mathematical domain can not be taken for granted. In order to stimulate students' HOT across multiple mathematical domains, it seems necessary to explicate the conceptual link between both domains. This can, for example, be accomplished by intertwining the learning strands of algebra and graphs in mathematical activities, as is already common in secondary school. For example, students in the eighth grade must develop an understanding of the interrelations between various representations of a function (e.g., algebraically and graphically), and must be able to find a solution to a system of two linear equations with two unknowns by drawing a graph of both equations and then find their point of intersection (CCSSI, 2010). Schliemann et al. (2012) showed that the relation between algebra and graphs can also informally be stimulated in primary school. They reported about an early algebra intervention in which fourth-grade students worked on a problem involving the comparison of two linear functions. In this problem, two boys, Robin and Mike, each have some money: Mike has $8 in his hand and in addition some money in his wallet; Robin has exactly three times as much money as the money in Mike's wallet. Students were given the task to figure out the amount of money both boys had. The fourth-grade students participating in this study made use of various representations when solving this problem: they created a table with various possible values for the amount of money in the wallet and the total amount of money for both boys, they used algebraic expressions for representing Mike's ($N + 8$) and Robin's ($3N$) amounts of money, and they created graphs of the amounts of money of both students. We consider the work of these fourth-grade students as involving HOT, which provides an example of how primary school students' HOT across the domains of algebra and graphing can be stimulated. When algebra would be an explicit part of the mathematics curriculum in Dutch primary schools (which is currently not the case), the interrelations between both mathematical domains could be made more explicit.

There are some limitations to our study. First, our LGM model allowed us to deal with the multiple cohorts, multiple repeated measurements, and categorical outcome measures. Nevertheless we encountered some difficulties throughout the process of model estimation. Although the fit of our model was sufficient in terms of the RMSEA, which is the most reliable fit measure when it concerns categorical data (Little, 2013), the CFI and TLI values were low. While the commonly applied cut-off criteria are said to not generalize to longitudinal or multi-group models (Little, 2013), preferably these values would have been higher in order to draw reliable conclusions on the basis of the parameter estimates of the model. The obtained parameters should thus be interpreted carefully. However, an exploration of model options did reveal robustness of the effects found in this study. On the one hand, the effect of the teaching sequence on students' algebraic reasoning was consistently large and highly significant under different modelling assumptions. On the other hand, whereas the effect of the teaching sequence on students' graphical reasoning fluctuated around the significance level, the effect was consistently very small. Including more students in our study or making use of one large group of students instead of multiple cohorts would probably have resulted in a better model fit and more power of our study.

Second, next to covariation as an important concept to attend to, also other concepts played an important role within the algebra lessons. This inevitably resulted in only partial overlap between the lessons on linear equations and the graphing tasks. The main focus of the algebra lessons was to stimulate students' algebraic reasoning by providing them with a learning environment incorporating a balance model in which they could reason about the model being in balance (representing equality in an equation) and through which they could develop informal context-connected algebraic strategies which underlie conventional equation solving strategies such as restructuring, isolation, and substitution. In these lessons, reasoning about the concept of equality was crucial. Although the concept of covariation plays a major role within reasoning about equality, a more direct relationship between these mathematical concepts addressed in the lessons and their expression within the graphing tasks (e.g., Berks & Vlasnik, 2014) could potentially have resulted in a larger improvement in students' graphical reasoning.

Furthermore, within this study, we focused on the mathematical HOT component combining sources of information about relationships. This HOT component can be related to the synthesis component of Bloom's taxonomy (Bloom et al., 1956; in a revised version of this taxonomy this component was called "create", Anderson & Krathwohl, 2001), which can be defined as putting together parts in a new way, or synthesizing parts into something new and different to form a coherent functional whole (Collins, 2014). However, this is only one possible conceptualization of HOT; focusing on alternative HOT components would have been an option as well. We propose representational fluency as an alternative HOT component which is worthwhile to consider when investigating the domains of linear equation solving and graphs of motion in future research. Representational fluency, or representational flexibility, is the capability of using multiple representations (such as graphs, tables, algebraic expressions, and verbal statements) and the flexibility to switch between them (Heinze et al., 2009). Flexible and adaptive use of representations is considered a key aspect of learning mathematics.

And finally, our classification of students' reasoning was based only on their written responses. These writings do not always reflect students' understanding of a task. Students possibly did not always write down their entire reasoning, resulting in an underestimation of their level of reasoning. Also, we classified students' reasoning by means of a coding scheme consisting of three (algebra) or four (graphs) levels of reasoning to allow for an IRT-based LGM analysis of students' reasoning. This classification of students' reasoning into levels might have resulted in a loss of richness of individual students' answers.

Notwithstanding these limitations, this study provides some insights concerning the development of mathematical HOT across two different but related mathematical domains. We showed that participating in a teaching sequence on linear equation solving resulted in a strong improvement of fifth-grade students' algebraic reasoning, but not of their graphical reasoning. Hence, no transfer to the domain of graphing motion appeared to take place. Rather, the HOT that students developed during the lessons seemed to be domain-specific and linked to the mathematical domain of algebra. The results of this study are in line with the domain-specific view on HOT, indicating that HOT emerges within the domain in which the task is situated. Transfer of HOT from one mathematical domain to another can thus not be taken for granted, not even when it concerns two related mathematical domains (i.e., algebra and graphs) in which similar concepts are addressed (i.e., covariation).

References

Alexander, P. A., Dinsmore, D. L., Fox, E., Grossnickle, E. M., Loughlin, S. M., Maggioni, L., Parkinson, M. M., & Winters, F. I. (2011). Higher order thinking and knowledge: Domain-general and domain-specific trends and future directions. In G. Schraw, & D. R. Robinson (Eds.), *Assessment of higher order thinking skills* (pp. 47–88). Information Age.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives.* Addison Wesley.

Antle, A. N., Corness, G., & Bevans, A. (2013). Balancing justice: Comparing whole body and controller-based interaction for an abstract domain. *International Journal of Arts and Technology, 6*(4), 388–409. https://doi.org/10.1504/IJART.2013.058285

Arcavi, A., Drijvers, P., & Stacey, K. (2016). *The learning and teaching of algebra: Ideas, insights and activities*. Routledge.

Berks, D. R., & Vlasnik, A. N. (2014). Working the system. *Mathematics Teacher*, *107*(7), 542–546. https://doi.org/10.5951/mathteacher.107.7.0542

Behr, M., Erlwanger, S., & Nichols, E. (1980). How children view the equals sign. *Mathematics Teaching, 92*(1), 13–15.

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354–369. https://doi.org/10.1177/1073191112446655

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: Handbook I. Cognitive domain.* Addison Wesley.

Bollen, K.A., & Curran, P. J. (2006). *Latent curve models.* Wiley.

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

Collins, R. (2014). Skills for the 21st Century: teaching higher-order thinking. *Curriculum & Leadership Journal, 12*(14).

Common Core State Standards State Initiative [CCSSI] (2010). *Preparing America's students for college and career*. National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Curriculum.nu (2019, October). Leergebied rekenen & wiskunde [Learning domain mathematics]. Retrieved from https://www.curriculum.nu/voorstellen/rekenen-wiskunde/

Duijzer, C., Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, Boom, J., M., Doorman, M., & Leseman, P. (2020). Fifth grade students' reasoning on graphs of motion and linear equations. In C. Duijzer, *Moving towards understanding: Reasoning about graphs in primary mathematics education* [Doctoral dissertation]. Utrecht University.

Duijzer, C., Van den Heuvel-Panhuizen, M., Veldhuis, M., & Doorman, M. (2019). Supporting primary school students' reasoning about motion graphs through physical experiences. *ZDM Mathematics Education*, *51*(6), 899–913. https://doi.org/10.1007/s11858-019-01072-6

Duijzer, C., Van den Heuvel-Panhuizen, M., Veldhuis, M., Doorman, M., & Leseman, P. (2020). Moving towards understanding: Students interpret and construct motion graphs. *Mediterranean Journal for Research in Mathematics Education*, *17*, 25–51.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling*. Lawrence Erlbaum Associates.

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 31–83). Cambridge University Press.

Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children's understanding of equality: A foundation for algebra. *Teaching Children Mathematics, 6*(4), 232–236.

Foong, P. Y. (2000). Open-ended problems for higher-order thinking in mathematics. *Teaching and Learning, 20*(2), 49–57.

Forster, M. (2004). Higher order thinking skills. *Research Developments, 11*(1), 10–15.

Goldenberg, P. E., Mark, J., Kang, J., Fries, M., Carter, C. J., & Cordner, T. (2015). *Making sense of algebra: Developing students' mathematical habits of mind.* Heinemann.

Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *ZDM Mathematics Education*, *41*(5), 535–540. https://doi.org/10.1007/s11858-009-0214-4

Janssen, J., Scheltens, F., & Kraemer, J. M. (2005). *Leerling- en onderwijsvolgsysteem rekenen-wiskunde [Student monitoring system mathematics]*. Cito.

Kieran, C. (1997). Mathematical concepts at the secondary school level: The learning of algebra and functions. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 133–158). Psychology Press.

Kieran, C., Pang, J., Schifter, D., & Ng, S. F. (2016). *Early algebra: Research into its nature, its learning, its teaching*. Springer (open access eBook). https://doi.org/10.1007/978-3-319-32258-2

Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens, A. C. (2005). Middle school students' understanding of core algebraic concepts: Equivalence & variable. *ZDM Mathematics Education*, *37*(1), 68–76. https://doi.org/10.1007/BF02655899

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312. https://www.jstor.org/stable/30034852

Kolovou, A., Van Den Heuvel-Panhuizen, M., & Bakker, A. (2009). Non-routine problem solving tasks in primary school mathematics textbooks–A needle in a haystack. *Mediterranean Journal for Research in Mathematics Education, 8*(2), 31–68.

Leighton, J.P. (2004). Defining and describing reason. In J. Leighton & R. Sternberg (Eds.), *The nature of reasoning* (pp. 3–11). Cambridge University Press.

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, *60*(1), 1–64. https://doi.org/10.3102/00346543060001001

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice*, *32*(3), 131–137. https://doi.org/10.1080/00405849309543588

Li, X., Ding, M., Capraro, M. M., & Capraro, R. M. (2008). Sources of differences in children's understandings of mathematical equality: Comparative analysis of teacher guides and student texts in China and the United States. *Cognition and Instruction*, *26*(2), 195–217. https://doi.org/10.1080/07370000801980845

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guildford Press.

McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development, 76*(4), 883–899. https://doi.org/10.1111/j.1467-8624.2005.00884.x

Miri, B., David, B. C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education*, *37*(4), 353–369. https://doi.org/10.1007/s11165-006-9029-2

Murray, E.H. (2011). I*mplementing higher order thinking in middle school mathematics classrooms* [Unpublished doctoral dissertation]. University of Georgia.

Murray, J. W. (2014) Higher-order thinking and metacognition in the firstyear core-education classroom: A case study in the use of color-coded drafts. *Open Review of Educational Research*, *1*(1), 56–69. https://doi.org/10.1080/23265507.2014.964297

Muthén, L.K. & Muthén, B.O. (1998–2017). *Mplus User's Guide* (Eighth Edition). Muthén & Muthén.

National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics.* NCTM.

Newmann, F. (1991). Higher order thinking in the teaching of social studies: Connections between theory and practice. In D. Perkins, J. Segal, & J. Voss (Eds.), *Informal reasoning and education*. Lawrence Erlbaum Associates.

Organization for Economic Co-operation and Development [OECD] (2019). *Skills outlook 2019: Thriving in a digital world*. Paris: OECD Publishing. https://doi.org/10.1787/df80bc12-en

Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, M., & Heinze, A. (2019). Developing algebraic reasoning in primary school using a hanging mobile as a learning supportive tool / El desarrollo del razonamiento algebraico en educación primaria utilizando una balanza como herramienta de apoyo. *Journal for the Study of Education and Development / Infancia y Aprendizaje*, *42*(3), 615–663. https://doi.org/10.1080/02103702.2019.1612137

Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, M., Boom, J., & Heinze, A. (2020). Are physical experiences with the balance model beneficial for students' algebraic reasoning? An evaluation of two learning environments for linear equations. *Education Sciences, 10*(6), 163, https://doi.org/10.3390/educsci10060163

Papadopoulos, I. (2019). Using mobile puzzles to exhibit certain algebraic habits of mind and demonstrate symbol-sense in primary school students. *The Journal of Mathematical Behavior, 53*, 210–227. https://doi.org/10.1016/j.jmathb.2018.07.001

Pirie, S. E., & Martin, L. (1997). The equation, the whole equation and nothing but the equation! One approach to the teaching of linear equations. *Educational Studies in Mathematics*, *34*(2), 159–181. https://doi.org/10.1023/A:1003051829991

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's standard progressive matrices and vocabulary scales*. Oxford Psychologists Press.

Resnick, L. B. (1987). *Education and learning to think*. National Academies Press.

Schliemann, A. D., Carraher, D. W., & Brizuela, B. M. (2012). Algebra in elementary school. In L. Coulange & J. P. Drouhard (Eds.), Enseignement de l'algèbre élémentaire: Bilan et perspectives (pp. 107–122). Special Issue of *Recherches en Didactique des Mathématiques*. La Pensée Sauvage.

Thijs, A., Fisser, P., & Van der Hoeven, M. (2014). *21e eeuwse vaardigheden in het curriculum van het funderend onderwijs: een conceptueel kader* [21st century skills in the curriculum of the foundation of education: a conceptual framework]. SLO.

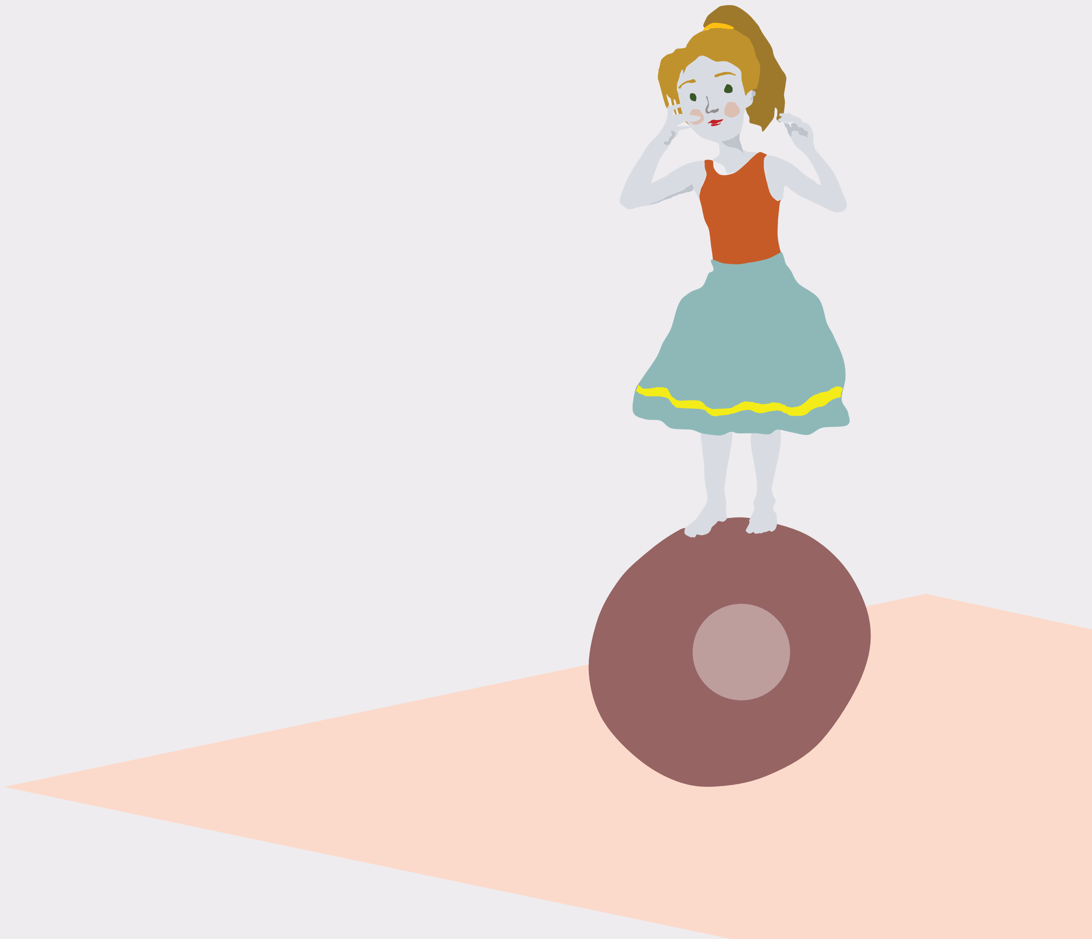
Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Handbook of research in mathematics education* (pp. 421–456). National Council of Teachers of Mathematics.

Van Zanten, M., & van den Heuvel-Panhuizen, M. (2018). Opportunity to learn problem solving in Dutch primary school mathematics textbooks. *ZDM Mathematics Education*, *50*(5), 827–838. https://doi.org/10.1007/s11858-018-0973-x

Willett, J. B., & Bub, K. L. (2005). Latent growth curve analysis. In D. Rindskopf (Ed.), *Encyclopedia of statistics in behavioral science* (2nd ed.). Wiley.

Zohar, A., & Schwartzer, N. (2005). Assessing teachers' pedagogical knowledge in the context of teaching higher-order thinking. *International Journal of Science Education*, *27*(13), 1595–1620. https://doi.org/10.1080/09500690500186592

Author contributions: All authors participated in designing the research. MO gave the lessons and collected the data. MO, CD, MH, MV, MD developed the assessment tasks and coding schemes. MO, CD, MV coded the data. MO, CD, JB analyzed the data. Methods of data analysis were frequently discussed with all authors. MO, CD prepared the first draft of the manuscript. All authors participated in revising the manuscript and/or provided feedback. All authors read and approved the final manuscript.



CHAPTER

6

Summary and discussion

Summary and discussion

The importance of including algebraic activities in the primary school mathematics curriculum is increasingly being emphasized (e.g., Kaput et al., 2008). Starting in the elementary grades with solving informal algebra problems that build on students' intuitive understanding and natural ways of thinking can provide students with a conceptual basis for the study of more formal algebra in the later grades (e.g., Kieran et al., 2016; Stephens et al., 2017). Although there is abundant evidence available that primary school students are capable of algebraic reasoning (Blanton et al., 2015; Brizuela & Schliemann, 2004; Kolovou et al., 2013; Papadopoulos, 2019; Van den Heuvel-Panhuizen et al., 2013), within the Dutch primary school mathematics curriculum algebra currently is virtually absent. This is not only a missed opportunity for creating a continuous learning strand from primary to secondary school, but also because stimulating young students' reasoning about advanced mathematical topics such as algebra within the primary school mathematics classroom has the potential to foster higher-order thinking (HOT). The main goal of this PhD project was to gain insight in whether, in what ways, and to what extent primary school students' early algebraic reasoning could be fostered.

The research presented in this thesis mainly focused on the evaluation of a six-lesson teaching sequence on early algebra, specifically on linear equations. A hanging mobile (see Figure 1), a physical balance model consisting of a horizontal beam with on each side a number of bags hanging on a chain, played a key role in this teaching sequence. When working with this hanging mobile, students' main task was to "discover everything which can be done, while keeping the hanging mobile straight". Students could manipulate the bags on the hanging mobile, and the beam tilted in response to students' actions. Keeping this beam of the mobile in a horizontal position represented maintaining equality in an equation. Understanding of the concept of equality is crucial when learning to solve linear equations (e.g., Bush & Karp, 2013; Kieran, 1981). While maintaining the balance of the model, students could intuitively apply and develop various informal algebraic strategies by changing the order of the bags on one side of the mobile, taking away bags with similar colors from both sides, or replacing on color of bags by bags from another color. These context-connected algebraic strategies underlie the conventional strategies for solving linear equations, such as restructuring, isolation, and substitution. In line with embodied cognition theory (e.g., Lakoff & Johnson, 1980; Núñez et al., 1999; Wilson, 2002), we assumed students' perceptual-motor experiences with the model to be beneficial for grounding the concept of equality and the strategies for maintaining equality in the bodily-based experience of maintaining balance.

Figure 1. The hanging mobile, a balance model which played a key role in our teaching sequence on linear equations.

Through working with the balance model, students were provided a basis for solving equations in various new contexts and eventually even for solving systems of linear equations. Solving systems of equations requires sophisticated algebraic skills, such as reasoning about the relationship between unknowns in equations and about the relationships between equations in the system of equations (i.e., reasoning about covariation; Thompson & Carlson, 2017), manipulating these relationships, combining them, and reasoning on the basis of newly gathered information (e.g., Goldenberg et al., 2015). These skills are also referred to as HOT skills (Lewis & Smith, 1993). Reasoning about early algebra problems might thus be a fruitful approach for stimulating the development of young students' HOT.

The first aim of this PhD thesis was to investigate the role of the balance model in teaching linear equation solving as reported in the international research literature (*Chapter 2*). The second aim was to investigate the potential of various representations of the balance model for supporting primary school students' understanding of linear equations. To this end, the above mentioned teaching sequence was developed and taught in various fifth-grade classes, and its effect on students' algebraic reasoning was evaluated (*Chapter 3* and *4*). The final aim of this thesis was to investigate whether stimulating primary school students' algebraic reasoning related to solving linear equations would also promote students' reasoning in a related mathematical domain which also draws on covariational reasoning (Leinhardt et al., 1990): graphing motion (*Chapter 5*). For this, the effect of our teaching sequence about linear equation on students' graphical reasoning was evaluated.

In this final chapter, I first summarize the findings of the studies described in this thesis. Then, I will discuss the implications of the findings, discuss the limitations of the studies in this thesis, and propose suggestions for further research.

1. Research overview and main findings

1.1 The role of the balance model in teaching linear equation solving

The first study that was conducted for this PhD thesis was a systematic literature review investigating the balance model. This review study is described in *Chapter 2*. The purpose of this study was to create an overview of the use of the balance model for teaching linear equation solving. Ninety-three peer-reviewed research journals were searched through in order to find research articles in which the balance model was used for teaching linear equation solving. This resulted in the final selection of 34 articles, which were analyzed in terms of the rationales for using the model, the appearances of the model, the situations of using the model, and the effect of the model on students' learning.

This review revealed a scattered picture of the use of the balance model for teaching linear equations. In terms of the rationales, the balance model was most often used for enhancing students' understanding of the concept of equality in an equation and the strategies that can be applied while maintaining this equality (such as: taking away similar things on both sides of the equation). Also, previous physical experiences related to maintaining balance or concurrent physical experiences with a concrete balance model were common arguments for using it. The restricted possibilities for representing equations containing negative values and subtractions were considered a limitation. Three types of appearances of the balance model could be distinguished: physical models (i.e., concrete models), virtual models (i.e., models in a digital environment), and drawn models (i.e., models presented on paper). The situations in which the balance model was used varied considerably as regards the age and the prior experience of the students involved (from kindergarten to Grade 9), the duration of the intervention (from one activity to multiple years), the type of equation problems that were taught (e.g., equations with or without negative values), and the type of provided instruction (e.g., classroom instruction by a teacher or exploring the model individually). Lastly, there were large variations in the reported effectiveness of using the balance model for teaching linear equation solving.

Notwithstanding the found kaleidoscopic image of the use of the balance model, some trends could be identified. Physical and virtual balance models were most often used for teaching students without prior algebra experience. These models were generally chosen because of their added value for understanding equality in an equation and for the physical experiences gained with the model. The equations taught with these models mostly only contained positive values and addition. Importantly, these balance models in general seemed to have a positive effect on enhancing students' understanding of linear equations. Meanwhile, drawn balance models were more frequently used for teaching somewhat older students with prior algebra experience. Additional features, such as manipulatives, were often added to these models to allow for representing a broader range of equation types (such as equations containing negative values and/or subtraction). The reasons for choosing these models were less clearly defined and the effects on students' learning varied considerably, with a mixture of positive and negative effects being reported. However, it is important to emphasize that the characteristics of the studies included in this review greatly differed. The abovementioned trends should therefore be interpreted with caution. In-depth knowledge about when which type of balance model could best be used effectively is still lacking. Hence, there still remains a lot to discover about this didactical model in order to be optimally used for teaching linear equation solving.

1.2 The potential of the balance model for supporting primary school students' understanding of linear equations

In the studies described in *Chapter 3* and *4*, we addressed this knowledge gap by evaluating the effects of an intervention consisting of a six-lesson teaching sequence with a balance model. The developed teaching sequence was aimed at stimulating primary school students' algebraic reasoning, with the focus on linear equations, as an approach to foster their mathematical HOT. Two parallel versions of the teaching sequence were created. In one version, a physical balance model in the form of a hanging mobile (see Figure 1) was used, combined with pictorial representations. In the other version, only pictorial representations of this model were used

1.2.1 Students' development over the six-lesson teaching sequence

The study described in *Chapter 3* focused on the development of students' algebraic reasoning over the teaching sequence with the physical balance model. To this end, we examined the reasoning of 65 fifth-grade students without prior experience in algebra. We analyzed their responses on lesson-specific assessment tasks, which were administered at the end of three consecutive episodes of the teaching sequence in which students solved systems of two linear equations. Solving systems of two equations requires students to take into account both linear equations and to combine their information, which can, for example, be accomplished by making use of the algebraic substitution strategy. When coding students' reasoning, we decided to focus on the number of equations that students used in their reasoning: none of the given equations (Level R0), only one of the given equations (Level R1), or both given equations within the system (Level R2). There appeared to be an intricate relation between students' use of algebraic strategies and the number of equations they referred to in their reasoning: the higher the level of reasoning, the more strategies and the more advanced strategies were used. Distinguishing these three levels of algebraic reasoning thus not only reflected the straightforward number of equations to which students referred in their reasoning, but also the depth of their reasoning by the use of the algebraic strategies.

Results showed that students clearly improved. After participating in the teaching sequence, 77% of the students could, for example, find the values of unknowns M and L in the system of equations $M + 3L = 25$, $2M = 4L$. Over the course of the teaching sequence, the percentage of students that reasoned without making use of the given equations decreased from 57% to 20%, while the percentage of students that combined the information of both given equations in their reasoning increased from 17% to 52%. Students also improved in their ability to explain their reasoning and to write down the algebraic strategies they applied. Lastly, also at individual level, the reasoning of most students improved over time without any decline.

The relationship between students' experiences in the learning environment with the physical balance model and their use of algebraic strategies was investigated by means of an analysis of videos of classroom interactions, students' written work during the lessons, and students' responses to the lesson-specific assessment tasks.

We found multiple indications for a relationship between students' experiences with the physical model and their use of algebraic strategies and forms of notations. Students, for example, used the context-connected algebraic strategies which they acquired when working with the balance model, such as restructuring, isolation, and substitution, to solve problems in other contexts. Moreover, some students incorporated a representation of the model in their reasoning when solving systems of two symbolic linear equations.

The environment with the physical balance model thus laid a basis for developing algebraic reasoning. In line with embodied cognition theory, we assumed students' perceptual-motor experiences with the physical balance model to contribute to the development of this reasoning. The design of this study, though, did now allow for systematic investigation of the specific contribution of the physical experiences.

1.2.2 Students' development over the school year

The quasi-experimental study described in *Chapter* 4 allowed us to further investigate the effects of our teaching sequence on the long-term and to investigate the specific contribution of physical experiences with a balance model. In total, 212 fifth-grade students, without prior algebra experience, participated in the study. These students were divided over the two intervention conditions and a control condition. Students of three classes received the six-lesson teaching sequence using only pictorial representations of the balance model, students of three classes received the exact same lessons with the addition of a physical balance model (see also *Chapter 3*), and students of the three classes forming the control condition received lessons on another mathematics topic. A staged comparison design was used: students' algebraic reasoning related to solving systems of two informal linear equations was assessed four times over the school year, and students participated in the teaching sequence in between two of these measurements in three successive cohorts, one class per cohort, for each condition. The effect of the intervention and the differential effect of both intervention conditions on the level of algebraic reasoning were analyzed by means of multi-group latent variable growth curve modelling. Qualitative analyses were conducted to further examine possible differences between conditions.

Results showed that students reached higher levels of reasoning when solving systems of informal linear equations, after having participated in the teaching sequence ($d = 0.73$). Students showed more often reasoning in which they combined the information of both given equations in their reasoning, instead of reasoning on the basis of only one equation or none of the given equations. This highest level of reasoning was displayed more frequently after the intervention (65%) than before (39%). These results emphasized the effectiveness of our teaching sequence with the balance model on primary school students' reasoning about systems of informal linear equations.

Participating in the teaching sequence with both a physical and a pictorial balance model, however, did not result in a significantly larger improvement in the level of algebraic reasoning (frequency of the highest level of reasoning increased 33%) than participating in a teaching sequence with only a pictorial balance model (18% increase in the highest level of reasoning). Yet, students who worked with the physical balance model, compared to students who worked only with a pictorial representation of the model, more often used representations of the model in their reasoning (17% vs. 1%) and more frequently used advanced algebraic strategies such as substitution or elimination (60% vs. 40%).

Taken together, this study showed the effectiveness of our teaching sequence with the balance model on primary school students' reasoning about systems of equations. While no significant differences were found between using a balance model on paper or a physical balance model on the development of students' level of reasoning, having a closer look on this reasoning revealed that students who worked with the physical model more often used representations of the model or advanced algebraic strategies when solving systems of equations. This suggests that different representations of the balance model might play a different role in individual learning processes.

1.3 Promoting students' reasoning across mathematical domains

In the studies described in *Chapter 3* and *4* we proved the effect of our teaching sequence on reasoning about systems of linear equations. Reasoning about and solving such systems asks for mathematical HOT. This HOT is, among others, characterized by reasoning about covariation, because students have to reason about the relationship between unknowns in an equation and about the mutual relationship between the equations. Reasoning about covariation is also relevant to other mathematical domains, such as the domain of graphs. This begs the question whether stimulating reasoning in the domain of linear equations might also be beneficial for reasoning about graphs. In the final study, described in *Chapter 5*, we investigated the effect of our teaching sequence about linear equations on 132 fifth-grade students' reasoning about both linear equations and graphs of motion, Students' algebraic and graphical reasoning were assessed four times over the school year. During each of these assessments, students solved four algebra tasks consisting of a system of two informal linear equations and four graphing tasks in which students had to interpret or construct a time-distance graph. Understanding the underlying relationship between covarying variables was required to solve the problems in both domains.

Students' reasoning on both sets of tasks was categorized in various levels. An integrated multi-group latent variable growth curve model was used in which the latent algebraic and graphical reasoning abilities were modelled together. In this way, we could investigate the effect of our intervention on both reasoning abilities. Our results, which due to the suboptimal fit of our model need to be interpreted with caution, showed an improvement of fifth-grade students' algebraic reasoning after participating in the teaching sequence, which was a confirmation of the results of the study described in *Chapter 4*. However, participating in the teaching sequence did not result in an improvement of these students' graphical reasoning. These findings seem to indicate that the HOT students developed in the lessons on linear equation solving did not transfer to the domain of graphing motion, and was rather domain-specific. We thus may conclude that the transfer of HOT from one mathematical domain to another related mathematical domain cannot be taken for granted.

2. Conclusions and implications

From the perspective of eliciting HOT into the mathematics classroom, students' reasoning related to solving systems of equations with multiple unknowns is particularly interesting. Solving these systems requires reasoning about how changes in one unknown result in changes in the other (i.e., covariational reasoning), making comparisons between both linear equations, and integrating the information found in both of them. Participating in our teaching sequence resulted in an improvement of students' ability to reason about systems of equations (*Chapter 4*), confirming the potential of our approach for fostering students' mathematical HOT. In our final study (*Chapter 5*), we investigated whether this HOT transferred to the related mathematical domain of graphing motion. We did not find indications for this transfer. Hence, the HOT students developed in the lessons on linear equation solving seems to be rather domain-specific, and linked to the mathematical domain of *algebra*.

2.1 Stimulating algebraic reasoning in primary school

When most people hear the word “algebra”, the first association which comes to mind is “symbols” such as *x* and *y*. When I was giving lectures and workshops during my PhD project, the perspective on algebra, as being inextricably linked to symbols, also came to the fore multiple times. This was, for example, evidenced by a secondary school teacher, commenting on my lecture: “It’s amazing that these 10-year olds can reason about and solve these problems, but… where exactly is the algebra?” The research described in this PhD thesis shows ample evidence that primary school students, before having received lessons in algebra, are very well able to engage in informal algebraic reasoning (*Chapter 3* to *5*). Moreover, when gradually introducing algebraic symbols, this informal algebraic reasoning can evolve into more formal algebraic thinking (*Chapter 3*). Our research also shows that such early algebra activities can be taught in real classroom settings. However, as Carraher et al. (2008) aptly described it, this “early algebra is not the same as algebra early” (p. 235). In other words, it does not mean that we teach algebra in its current form but then to younger students. Early algebra builds on rich problem contexts and using natural language to express ideas and to reason about algebraic concepts is considered an important starting point, because it allows students to make sense of the algebraic concepts (Kaput et al., 2008). An example of the use of such natural language to express ideas about the concept of equality (in this case to describe the symmetry property of equality; $a = b \Leftrightarrow b = a$) was provided by Eva (*Chapter 3*), who reasoned while working with the physical balance model: “We can do exactly the same, but then on the other side. So, these bags to the other side, and these ones also all to the other side”.

Searching for possibilities for engaging young students in algebraic reasoning thus requires adopting a broad view of algebraic reasoning, perceiving algebraic reasoning as including, but certainly not restricted to, reasoning with algebraic symbols (see also Kaput et al., 2008). In this way, a longitudinal learning strand throughout K-12 mathematics can be created, starting in the elementary grades with solving such informal algebraic problems that build on students’ natural ways of thinking and understanding of mathematical patterns and relationships and evolving into more formal algebra in later grades (e.g., Carraher et al., 2008, Kieran et al., 2016; NCTM, 2000). In order to prevent abrupt and isolated introduction of algebra in secondary school, and to add more depth into the K-8 curriculum by introducing activities requiring mathematical HOT, I highly recommend incorporating such early algebra activities into the Dutch primary school mathematics curriculum.

Incorporating algebra into the primary school mathematics curriculum, however, can bring along multiple challenges. This thesis is entitled “Algebraic reasoning in primary school: A balancing act”. This “balancing act” can be taken very literally (i.e., as a reference to maintaining balance or the balance of a model), but, at the same time, it can also be interpreted as a metaphor for the challenges we face when incorporating algebraic reasoning into the primary school curriculum. In the remainder of this final chapter, I elaborate on four of these “balancing acts”. The first balancing act is closely linked to the results of the research described in this thesis regarding how to foster primary school students’ algebraic reasoning. Balancing acts 2 to 4 widen the perspective and focus on implications of our research for educational practice with respect to curriculum, teacher education and dealing with heterogenity.

2.1.1 Balancing act 1: How to stimulate young students’ algebraic reasoning?

In the design of the teaching sequence aimed to foster students’ algebraic reasoning about linear equations, we made several decisions based on both a mathe-didactical analysis of solving linear equations and insights from the psychology of learning. In the lessons making up the teaching sequence these insights were combined. What conclusions can be drawn about how to stimulate students’ algebraic reasoning?

The first two main decisions about the design of the teaching sequence resulted from a mathe-didactical analysis of solving linear equations. The first decision was using the balance model for enhancing students’ understanding of equality and strategies for maintaining equality. Starting with a meaningful context and using models to bring students to higher levels of mathematical understanding, are key principles of the domain-specific instruction theory of Realistic Mathematics Education (RME; e.g., Freudenthal, 1973, 1991; Treffers, 1978, 1987; Van den Heuvel-Panhuizen, 2001; Van den Heuvel-Panhuizen & Drijvers, 2020). Students have an intuitive understanding of what “balance” means, as a result from many bodily experiences in everyday life, such as walking without falling, riding a bicycle, or holding objects of similar or different weights (Alessandroni, 2018). Also, most students have experiences playing with objects in which balancing plays a key role, such as a teeter-totter. These experiences together make that students can imagine what the problem is about when the model of a balance is used (i.e., the model serves as a meaningful context to students). Students can build on their intuitive sense of balance when exploring possible ways to maintain the balance of the model. When students worked with this model during our lessons, this directly resulted in more or less spontaneous ideas of how to approach the problem

(*Chapter 3*). Over the lessons, the *model of* the balance evolved into a *model for* solving systems of linear equations in other informal contexts (Streefland, 1985; 2003; see also *Chapter 3* and *4*) and for solving systems of equations at a formal level (*Chapter 3*). This was directly evidenced by students' use of the representation of the model when solving such problems, or indirectly by students' applying algebraic strategies which could be related to their experiences with the model. Our studies thus clearly demonstrate the suitability of the balance model for fostering students' understanding of linear equations. However, as we will further elaborate on later in this balancing act, the possible differential effect of using different representations of this model (a representation of the model on paper or a concrete model with students can gain physical experiences) on students' reasoning, warrants further research.

The second decision, which was linked to our approach of using the balance model to foster students' understanding of linear equations, came down to first starting with reasoning about relations between unknowns. Algebra is often treated as an elaboration of arithmetic, going from reasoning with numbers to reasoning with unknowns (e.g., Filloy & Rojano, 1989). Our approach was to directly start with evoking students' reasoning with unknowns. When learning algebra, students must go beyond the mastery of certain computational skills and instead develop new ways of thinking, such as the ability to analyze relationships, generalize, notice structure, solve problems, and make justifications and predictions (Cai & Knuth, 2011). This implies that students have to make a "shift from thinking about relations among particular numbers and measures toward thinking about relations among sets of numbers and measures, from computing numerical answers to describing and presenting relationships among variables" (Carraher et al., 2008, p. 266). By requiring students to reason about relationships between unknowns, we stimulated them to look for logic and coherence and to search for structure, instead of to solve the problems by performing calculations. This prompted students to come up with context-connected strategies, which can be seen as first steps towards fundamental linear equation solving strategies such as restructuring, isolation, and substitution (*Chapter 3*). Students further developed these strategies over the course of the teaching sequence and used them to solve other (systems of) linear equations (*Chapter 3* and *4*).

The third decision was to make use of a physical balance model with which students could gain perceptual-motor experiences (*Chapter 3* and *4*). This decision was inspired by embodied cognition theory, stating that higher-order cognition, such as mathematical cognition, benefits from such perceptual-motor experiences of our body in interaction with the world around us (e.g., Núñez et al., 1999). Students who worked with the physical balance model appeared to improve more from pretest to posttest than students who worked with the model on paper. However, no significant contribution of students' perceptual-motor experiences with the balance model on their level of reasoning about linear equations was found (*Chapter 4*). Different explanations for this finding are possible. First, the balance model could be such a strong didactical model, which, independent of its representations, is very accessible for students to make sense of the problem situation. Related to this, maybe both representations of the model actually result in similar neural activity, either directly through perceptual-motor experiences, or indirectly through mental simulation. Finally, students working with different representations of the balance model mainly took place in the first three lessons; this might have been too short to actually affect students' reasoning differently.

While the differential effect of the representation of the balance model was not significant for students' *level* of reasoning, zooming in on students' reasoning did reveal differences as regards students' *use* of the model. Students who had worked with a physical balance model more often used representations of the model or advanced algebraic strategies when solving systems of informal linear equations *(Chapter 4)*. An alternative explanation for the lack of significant differential effects on students' level of reasoning might therefore be that our coding scheme was unable to capture subtle differences between conditions. Hence, further research is required to investigate the exact contribution of embodied experiences with a physical balance model on students' reasoning. Live registration of students' reasoning might be an alternative research method for capturing students' reasoning.

So taken together, the use of a balance model with which students had to reason about relationships between unknowns, and which prompted them to develop context-connected equation solving strategies, has probably been crucial within our approach to stimulate students' algebraic reasoning. At the same time, there is no clear-cut conclusion about the contribution of physical experiences. Our approach consisted of a six-lesson teaching sequence for Grade 5. These lessons were an add-on to students' normal lessons, without making connections to their regular mathematics curriculum. This brings me to the second balancing act.

2.1.2 Balancing act 2: How to add algebraic reasoning to an already bulging mathematics curriculum?

Incorporating algebra in primary school does not simply imply: add this topic to the current curriculum, or "make the curriculum bigger". Instead, it means that activities need to be integrated and that existing topics need to be treated more deeply (e.g., Goldenberg et al., 2003; Kaput, 2008). As our teaching sequence proved to be viable for stimulating students' algebraic reasoning about linear equations (*Chapter 3* to *5*), it seems promising to incorporate the main ideas of our teaching sequence into the mathematics curriculum and to connect them with activities which already take place in mathematics education. This means that, from very young ages on, students have to develop understanding of mathematical structures and relationships, such as understanding of relationships between quantities, of properties of operations (e.g., commutative and associative properties) and of the concept of equality (e.g., symmetry property of equality). For this, students need many experiences with recognizing, defining, creating, and maintaining equality. This can be actualized by various activities with balance scales (e.g., Taylor-Cox, 2003). Most kindergartens already use physical balance scales in the classroom. These scales are used for all kinds of activities, such as for "doing groceries" or during cooking activities. Posing the right kind of question can guide students in their process of discovery and bring the focus onto discovering and reasoning about equality. Students are, for example, asked to predict what will happen with a balance scale when a number of objects, either identical or different, are placed on each side of the balance scale. Or what will happen when all objects from the right side of the scale switch with the objects on the left side. During these activities, students can hold various objects, one in each hand, to experience and feel the weights. Students can also spread their arms and enact a balance scale (e.g., Mann, 2004). In this way, students can develop basic understanding of equality, rooted with meaning in natural language.

Students can also be asked what would be necessary in order to bring an unbalanced scale, with, for example, six identical objects on the one side and four of these objects on the other side, into balance. Such activities can, in a later stage, be connected to solving addition problems such as $6 = 4 + \ldots$. In this way, algebra and arithmetic get intertwined, and algebra activities can provide rich, meaningful contexts in which students can also develop computational skills (e.g., Warren et al., 2009). During such activities, though, it is important to keep the focus on reasoning, instead of merely on performing calculations.

In higher grades, these activities can be further extended to applying algebraic strategies for solving other types of equations. Following the ideas of RME a process of concreteness fading then might be necessary, starting with the concrete model providing a meaningful context and then, via representations on paper, fading into abstract symbols (e.g., Fyfe et al., 2014). In this way, the model of the balance will become a model for solving new equations. Activities can also be intertwined with other learning strands. Students can, for example, learn to use other representational tools such as tables and graphs, for solving problems. Such activities might also be beneficial for stimulating students' HOT in related mathematical domains (see *Chapter* 5).

The current primary school mathematics curriculum thus already contains activities which can be used for stimulating students' algebraic reasoning. Yet, full advantage of the potential of this type of activities to actually foster this reasoning is currently not taken. Further research is necessary to investigate the feasibility of connecting these activities to early algebra experiences. In addition, putting these ideas regarding long-term sustained algebra experiences into practice, demands much from primary school teachers.

2.1.3 Balancing act 3: How to prepare primary school teachers for teaching early algebra?

If primary school students are to have early algebraic reasoning experiences, then (prospective) elementary teachers have to be prepared to guide students during such activities. This has implications for the teacher education program and professional development programs. Research shows that prospective teachers are often unaware of students' common misconceptions related to algebra, for example about the meaning of the equal sign (e.g., Stephens, 2006). They also tend to have a rather narrow view of algebra: most teachers "equate[d] algebra with the manipulation of symbols" (p. 33), and very few characterize other forms of reasoning, such as relational reasoning, as algebraic (Stephens, 2008). Moreover, prospective teachers experience difficulties regarding how to make algebraic reasoning accessible to primary school students (Hohensee, 2017). The findings of these studies are in line with the experiences I had when teaching early algebra to prospective teachers during this PhD project. When prospective teachers engaged their own students in early algebra activities with a hanging mobile, the teachers were inclined to stimulate their students to assign values to the unknowns and to make calculations (i.e., focus on operations), instead of stimulating their reasoning about the relationships between the unknowns (i.e., focus on relations). In addition, most of these prospective teachers underestimated their students' capabilities to engage in early algebra activities (see also Dobrynina & Tsankova, 2005) and commented that they considered such activities as being suitable only for high-achieving students (see also Zohar et al., 2001). Also, the majority of the teachers had severe doubts about their own capacity to reason algebraically. Such low self-efficacy can have a negative influence on teachers' receptiveness to implement reform-oriented instructional practices and their persistence in its implementation (e.g., Guskey, 1988), and on their teaching performances (e.g., Klassen & Tze, 2014).

Teachers' unawareness of students' common misconceptions, a narrow conception of algebra, and low self-efficacy beliefs might be troublesome when teaching early algebra. However, as Hohensee (2017) showed, engaging prospective teachers in early algebra activities can help them with developing new insights on how to teach algebra to elementary students in a way which is different from the algebra they (probably) have been taught themselves. Similarly, in-service teachers might need professional development training in order to prepare them for teaching (e.g., Wilkie, 2013). Jacobs et al. (2007) reported about the effectiveness of a professional development course for elementary school teachers, focusing on relational thinking and the equal sign. Teachers participating in this course, compared to non-participating teachers, were better able to generate possible strategies used by students for solving equivalence problems like 8 + 4 = __ + 5. They also were better able to use a strategy reflecting relational thinking themselves, and also the students of these teachers showed better understanding of the equal sign and showed more relational thinking compared to students of non-participating teachers.

Integrating early algebra in primary school thus also requires integrating early algebra activities in teacher education and professional development programs, so that teachers are adequately prepared for teaching such lessons. Such activities will assist teachers in gaining insight in the nature of early algebra, in the importance of integrating such activities in the primary school mathematics curriculum, and in students' capabilities to reason algebraically. This latter insight is related to the final balancing act.

2.1.4 Balancing act 4: How to make early algebra accessible to *all* students?

Extra-curricular materials in the Dutch primary mathematics classroom occasionally contain tasks which can be considered as early algebra tasks, and which can foster students' HOT. These materials are typically meant for the high-achieving students. Teachers often believe that tasks requiring HOT are only appropriate for high-achieving students (Zohar et al., 2001). As a result, in practice, mainly only high-achieving students are provided opportunities for developing HOT. Zohar and Dori (2003) investigated the relation between student achievement and HOT in the science classroom. They showed that both low- and high-achieving students benefitted from participating in tasks involving HOT. In one of their studies, the gain for low achievers was even significantly higher than the gain for high achievers. These researchers strongly recommended involving students of all academic levels in HOT. Also Peltenburg et al. (2012) showed that primary school students attending special education can develop mathematical HOT. Although these studies did not focus on the learning domain of algebra, their results are promising.

When integrating early algebra activities in primary school, the goal should thus be to provide *all* students opportunities to engage in such activities and to develop HOT. We assume that our approach of stimulating students' algebraic reasoning from early ages on, by means of activities which build on the intuitive understanding of balance which is a common ground for all students, is a viable approach to realize this goal.

3. Limitations of this PhD research

The first limitation of our research relates to the way in which we measured students' algebraic reasoning. Our classification of students' level of reasoning (*Chapter 3* to *5*) was fully based on their written responses. Students were asked to elaborate on their reasoning and the applied algebraic strategies by answering the question "*How do you know?*" When a student did not provide an answer to this question, their answer was by default classified in the lowest level of reasoning. However, one cannot be sure that the absence of a written explanation reflected a low understanding. A lack of motivation, for example, could also have caused the absence of a response. As a consequence, we cannot be sure that students' written answers provided an accurate reflection of their complete understanding (e.g., Fagginger-Auer et al., 2015). Based on our video and audio data of classroom interactions, presented in *Chapter 3*, we are inclined to think that focusing on students' written explanations might have led to an underestimation of their understanding. Some students were very well able to provide a verbal explanation of their reasoning during the lessons, while these students did not write down explanations with the same level of sophistication on the assessment tasks. In addition, students' development in reasoning about linear equations over the school year was measured by only four tasks. Although this limited number of tasks contributed to the ecological validity of our study, including more tasks might have resulted in a more elaborate picture of students' understanding. Also, within our research we only focused on one topic of early algebra (see *Chapter 1*): linear equations. Further research is required to investigate Dutch primary school students' ability to reason about other (early) algebra topics, such as functional thinking (e.g., Kaput, 2008).

The second limitation related to our way of analyzing students' reasoning. The data collection of this thesis mainly took place in one school year. Students' algebraic reasoning was measured four times over the school year. In between two of these measurements, students participated in the teaching sequence on linear equations. This staged comparison design with different cohorts allowed that I taught all lessons in all classrooms myself, which was crucial for being able to draw reliable conclusions without having to account for the effects of having different teachers. For the analysis of students' development in algebraic reasoning over the school year (*Chapter 4* and *5*), we used an advanced multi-group latent variable growth curve modelling technique. This statistical analysis matched our research design very well, because it allowed us to model the longitudinal change in algebraic reasoning, while taking into account the different cohorts of our study. However, there were also drawbacks to this analysis. Ideally, this analysis is conducted with larger samples (Little, 2013) so that smaller effects are detected more easily. Including more students in our research or making use of one large cohort of students instead of multiple cohorts would probably have resulted in a better model fit and larger power of our study. This was, however, not feasible from a practical point of view. Moreover, the analysis did not allow for a very detailed coding scheme, because, in order for the model to be estimated, for each of the cohorts all levels of reasoning on all tasks on all measurements had to appear at least once (i.e., no empty cells were allowed). This played a role in deciding on a relatively straightforward coding scheme for students' algebraic reasoning. However, the decisive argument for using this coding scheme was that focusing on the number of equations the students involved in their reasoning turned out to be the most reliable way of coding (with a high inter-rater reliability), whereas students' algebraic strategies were often difficult to discern from their written responses. Because there appeared to be an intricate relation between the use of algebraic strategies and the number of used equations, coding students' reasoning in this way was our best option. Nevertheless, in the end we did lose some of the richness in students' reasoning. This makes it difficult to draw clear-cut conclusions about the results of

Chapter 4, where we did not find a significant difference between our intervention conditions with our advanced statistical analysis on students' *levels* of reasoning, while we did find some differences when zooming in on students' reasoning by focusing on students' *use* of the balance model after participating in the lessons (reflected by either their use of a representation of the model or their use of advanced algebraic strategies).

Third, the goal of the studies (*Chapter 3* to *5*) described in this PhD thesis was to investigate the effects of the teaching sequence on linear equations in *real classrooms*. While making use of this real classroom setting resulted in a high ecological validity, there are also some methodological drawbacks to this study design. The study taking place in this natural classroom setting and the assessments of the study being spread out over almost the entire school year, made it impossible to control for some confounding factors (e.g., people entering the classroom during an assessment; extra-curricular activities over the course of the school year). In addition, a convenience sample of schools participated in the study, which means that no random selection took place. Also, although we did control for initial differences between classes, there was no random assignment of students over conditions.

And lastly, the limited duration of the research project did not allow us to follow these students when receiving algebra lessons in secondary school. Although we were able to follow the development of these students over the course of one school year, we do not know whether our intervention provided a firm conceptual basis on which these students could still rely when participating in algebra classes and study more formal algebra. Yet, our findings that most students were able to solve systems of formal equations after six lessons (*Chapter 3*), with seemingly little problems with often-reported difficulties related to solving linear equations, such as difficulties with understanding the equal sign (e.g., Behr et al., 1980), difficulties with understanding algebraic symbols (e.g., Koedinger & Nathan, 2004), or difficulties with operating on unknowns (e.g., Herscovics & Linchevski, 1994), seems promising. These results warrant further investigation on the long-term.

4. Concluding remarks

In the recent years, discussions have been going on about what Dutch education should encompass within this 21st century. In the recently released proposal on mathematics education (curriculum.nu, 2019), the importance of creating longitudinal learning strands which start in primary education and continue in the higher grades, among others for the mathematical domain of algebra, is endorsed. This is in line with the international consensus that a foundation for algebra should be laid in primary school and research showing that algebra can be integrated in primary school (e.g., Kaput et al., 2008; Kieran et al., 2016). Within the proposal on the Dutch future mathematics curriculum (curriculum.nu, 2019), it is moreover emphasized that attention should be paid to primary school students' development of higher-order ways of mathematical thinking. Also the Dutch Association for the Development of Mathematics Education (NVORWO, 2020) recently launched recommendations to put effort in supporting students' development of HOT within the primary school mathematics classroom. Yet, the question remains on how stimulating such HOT within the mathematics classroom can be put into practice.

In this PhD thesis, we provided evidence on a particular way to foster students' HOT within the primary school mathematics classroom. The domain-specific mathematical HOT targeted in these studies was operationalized as algebraic reasoning about linear equations. We showed that partaking in activities with a balance model, targeting students' reasoning about relationships between unknowns, resulted in students developing context-connected algebraic strategies, which in a later stage could be applied to solve (systems of) equations in other contexts and even for solving systems of formal equations. This PhD thesis was part of the *Beyond Flatland* project (see *Chapter 1*). The aim of this research project was to explore the possibilities for enriching the Dutch "flat" primary school mathematics curriculum as an approach to stimulate primary school students' HOT. Providing primary school students opportunities for engaging in algebraic reasoning is a worthwhile approach to consider and to further explore.

References

Alessandroni, N. (2018). Varieties of embodiment in cognitive science. *Theory & Psychology, 28*(2), 227–248. https://doi.org/10.1177/0959354317745589

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J. S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education*, *46*(1), 39–87. https://doi.org/10.5951/jresematheduc.46.1.0039

Brizuela, B., & Schliemann, A. (2004). Ten-year-old students solving linear equations. *For the Learning of Mathematics, 24*(2), 33–40.

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior*, *32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

Cai, J., & Knuth, E. (2011). A global dialogue about early algebraization from multiple perspectives. In J. Cai, & E. Knuth (Eds.), *Early algebraization* (pp. vii–x). Springer.

Carraher, D. W., Schliemann, A. D., & Schwartz, J. (2008). Early algebra is not the same as algebra early. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 235–272). Lawrence Erlbaum Associates.

Curriculum.nu (2019, October). Leergebied rekenen & wiskunde [Learning domain mathematics]. Retrieved from https://www.curriculum.nu/voorstellen/rekenen-wiskunde/

Dobrynina, G., & Tsankova, J. (2005). Algebraic reasoning of young students and preservice elementary teachers. In G. M. Lloyd, M. Wilson, J. L. M. Wilkins, & S. L. Behm (Eds.), *Proceedings of the 27th annual meeting of the North American chapter of the international group for the Psychology of Mathematics Education*. PMENA. Retrieved from https://www.pmena.org/pmenaproceedings/PMENA%2027%202005%20Proceedings.pdf

Fagginger Auer, M. F., Hickendorff, M., & van Putten, C. M. (2015). Strategiegebruik bij rekenen afleiden uit het schriftelijk werk van basisschoolleerlingen [Derive strategies in arithmetic from primary school students' written work]. *Pedagogische Studiën*, *92*(1), 9–23.

Filloy, E., & Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics*, *9*(2), 19–25.

Freudenthal, H. (1973). *Mathematics as an educational task*. Reidel Publishing Company.

Freudenthal, H. (1991). *Revisiting mathematics education.* Kluwer Academic Publishers.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*(1), 9–25. https://doi.org/10.1007/s10648-014-9249-3

Goldenberg, P. E., Mark, J., Kang, J., Fries, M., Carter, C. J., & Cordner, T. (2015). *Making sense of algebra: Developing students' mathematical habits of mind.* Heinemann.

Goldenberg, E. P., Shteingold, N., & Feurzeig, N. (2003). Mathematical habits of mind for young children. In F.K. Lester & R.I. Charles (Eds.), *Teaching mathematics through problem solving: Prekindergarten-grade* (pp. 15–29). National Council of Teachers of Mathematics.

Guskey, T. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, *4*(1), 63–69. https://doi.org/10.1016/0742-051X(88)90025-X

Herscovics, N., & Linchevski, L. (1994). A cognitive gap between arithmetic and algebra. *Educational Studies in Mathematics*, *27*(1), 59–78. https://doi.org/10.1007/BF01284528

Hohensee, C. (2017). Preparing elementary prospective teachers to teach early algebra. *Journal of Mathematics Teacher Education*, *20*(3), 231–257. https://doi.org/10.1007/s10857-015-9324-9

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, *38*(3), 258–288. https://www.jstor.org/stable/30034868

Kaput, J. J. (2008). What is algebra? What is algebraic reasoning? In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 5–18). Lawrence Erlbaum Associates.

Kaput, J. J., Carraher, D. W., & Blanton, M. L. (2008). *Algebra in the early grades.* Lawrence Erlbaum Associates.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Kieran, C., Pang, J., Schifter, D., & Ng, S. F. (2016). *Early algebra: Research into its nature, its learning, its teaching.* Springer (open access eBook). https://doi.org/10.1007/978-3-319-32258-2

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review,* 12, 59–76. https://doi.org/10.1016/j.edurev.2014.06.001

Kolovou, A., Van den Heuvel-Panhuizen, M., & Köller, O. (2013). An intervention including an online game to improve Grade 6 students' performance in early algebra. *Journal for Research in Mathematics Education, 44*(3), 510–549. https://doi.org/10.5951/jresematheduc.44.3.0510

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, *13*(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1

Lakoff, G., & Johnson, M. (1980). Conceptual metaphor in everyday language. *The Journal of Philosophy, 77*(8), 453–486. https://doi.org/10.2307/2025464

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Researc*h, *60*(1), 1–64. https://doi.org/10.3102/00346543060001001

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice*, *32*(3), 131–137. https://doi.org/10.1080/00405849309543588

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guildford Press.

Mann, R. L. (2004). Balancing act: The truth behind the equals sign. *Teaching Children Mathematics*, *11*(2), 65–70.

National Council of Teachers of Mathematics [NCTM] (2000). *Principles and standards for school mathematics.* NCTM.

Nederlandse Vereniging voor de Ontwikkeling van Reken-Wiskunde Onderwijs [NVORWO] (2020, January). *Position paper – Reactie NVORWO op eindproduct Curriculum.nu* [Position paper – Reply NVORWO on final product Curriculum.nu]. Retrieved from https://www.nvorwo.nl/proposition-paper-reactie-nvorwo-op-eindproduct-curriculum-nu/

Núñez, R. E., Edwards, L. D., & Matos, J. F. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics*, *39*(1–3), 45–65. https://doi.org/10.1023/A:1003759711966

Papadopoulos, I. (2019). Using mobile puzzles to exhibit certain algebraic habits of mind and demonstrate symbol-sense in primary school students. *The Journal of Mathematical Behavior, 53*, 210–227. https://doi.org/10.1016/j.jmathb.2018.07.001

Peltenburg, M., Van den Heuvel-Panhuizen, M., & Robitzsch, A. (2013). Special education students' strategies in solving elementary combinatorics problems. *In A. M. Lindmeier, & A. Heinze (Eds.), Proceedings of the 37th conference of the international group for the Psychology of Mathematics Education* (Vol. 4, pp. 9–16). PME.

Stephens, A. C. (2006). Equivalence and relational thinking: Preservice elementary teachers' awareness of opportunities and misconceptions. *Journal of Mathematics Teacher Education, 9*(3), 249–278. https://doi.org/10.1007/s10857-006-9000-1

Stephens, A. C. (2008). What "counts" as algebra in the eyes of preservice elementary teachers? *The Journal of Mathematical Behavior, 27*(1), 33–47. https://doi.org/10.1016/j.jmathb.2007.12.002

Stephens, A. C., Fonger, N., Strachota, S., Isler, I., Blanton, M., Knuth, E., & Gardiner, A. (2017). A learning progression for elementary students' functional thinking. *Mathematical Thinking and Learning*, *19*(3), 143–166. https://doi.org/10.1080/10986065.2017.1328636

Streefland, L. (1985). Wiskunde als activiteit en de realiteit als bron [Mathematics as an activity and the reality as a source]. *Tijdschrift voor Nederlands Wiskundeonderwijs (Nieuwe Wiskrant)*, *5*(1), 60–67. Retrieved from http://www.fisme.science.uu.nl/wiskrant/artikelen/artikelen00-10/051/0501september_streefland.pdf

Streefland, L. (2003). Learning from history for teaching in the future. *Educational Studies in Mathematics, 54*(1), 37–62. https://doi.org/10.1023/B:EDUC.0000005236.36499.fc

Taylor-Cox, J. (2003). Algebra in the early years. *Young Children, 58*(1), 14–21.

Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Handbook of research in mathematics education* (pp. 421–456). National Council of Teachers of Mathematics.

Treffers, A. (1978). *Wiskobas doelgericht* [Wiskobas goal-directed]. Instituut voor Ontwikkeling van het Wiskunde Onderwijs.

Treffers, A (1987). *Three dimensions. A model of goal and theory description in mathematics instruction – The Wiskobas project*. Reidel Publishing Company.

Van den Heuvel-Panhuizen, M. (2001). Realistic mathematics education in the Netherlands. In J. Anghileri (Ed.), *Principles and practices in arithmetic teaching: Innovative approaches for the primary classroom* (pp. 49–63). Open University Press.

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2020). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 713–717). Springer. https://doi.org/10.1007/978-3-030-15789-0_170

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

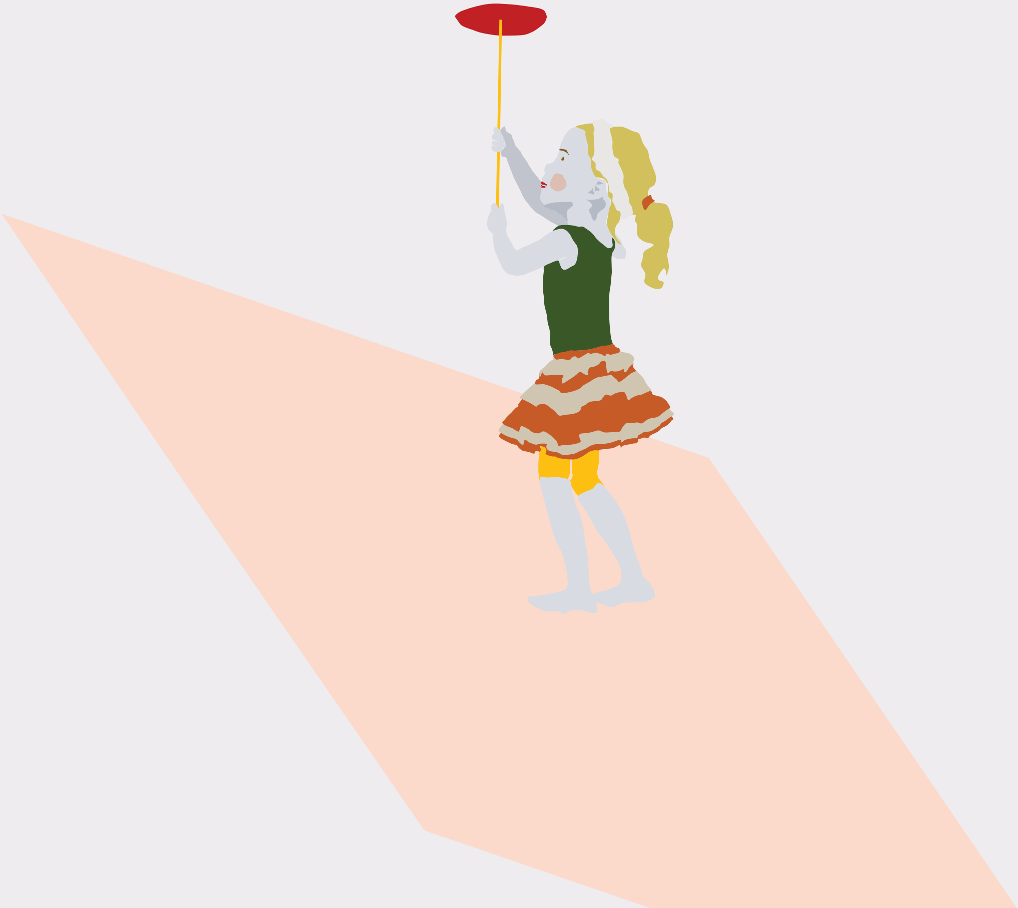
Warren, E., Mollinson, A., & Oestrich, K. (2009). Equivalence and equations in early years classrooms. *Australian Primary Mathematics Classroom*, *14*(1), 10–15.

Wilkie, K. J. (2014). Upper primary school teachers' mathematical knowledge for teaching functional thinking in algebra. *Journal of Mathematics Teacher Education*, *17*(5), 397–428. https://doi.org/10.1007/s10857-013-9251-6

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/BF03196322

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low achieving students and higher order thinking. *Teaching and Teachers' Education, 17*(4), 469–485. https://doi.org/10.1016/S0742-051X(01)00007-5

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *Journal of the Learning Sciences*, *12*(2), 145–181. https://doi.org/10.1207/S15327809JLS1202_1



Samenvatting (summary in Dutch)

Dankwoord (acknowledgements)

About the author

List of publications related to this thesis

ICO dissertation series

Appendices

Samenvatting (summary in Dutch)

Algebra is in Nederland nog geen onderdeel van het reken-wiskundecurriculum van het basisonderwijs. Dit is een gemiste kans, want er is veel bewijs dat jonge kinderen in staat zijn om algebraïsch te redeneren, zoals het redeneren over relaties tussen onbekenden. Door leerlingen al op de basisschool activiteiten aan te bieden waarin algebraïsch redeneren een rol speelt, kan de basis gelegd worden voor begrip van de formele algebra die leerlingen in latere klassen onderwezen krijgen. Algebraïsch redeneren vraagt om bepaalde hogere-orde denkvaardigheden (HOV), zoals het redeneren over relaties, het leggen van verbanden en het oplossen van problemen. Het opnemen van algebra in het reken-wiskundecurriculum van het basisonderwijs biedt mogelijk aanknopingspunten voor het stimuleren van HOV van basisschoolleerlingen.

Het doel van dit promotieonderzoek was inzicht te verkrijgen in of, hoe en in hoeverre het algebraïsch redeneren van basisschoolleerlingen gestimuleerd kan worden. Dit onderzoek maakte deel uit van het *Beyond Flatland* project, dat zich in bredere zin richtte op het verrijken van het “platte” reken-wiskundecurriculum op de basisschool met wiskundige activiteiten die het hogere-orde denken van leerlingen stimuleren. *Beyond Flatland* bestond uit drie deelprojecten, die elk gericht waren op een ander wiskundig domein: grafieken, kans en algebra. Dit proefschrift gaat over het algebra-deelproject.

De ontwikkeling van een lessenserie over algebra op de basisschool en de evaluatie van het effect hiervan op het redeneren van leerlingen stonden centraal in de in dit proefschrift beschreven onderzoeken. In deze serie van zes lessen lag de focus op het redeneren over, en het oplossen van, informele lineaire vergelijkingen. Dit zijn vergelijkingen die zijn weergegeven in een informele context. Een *hangmobiel* (zie Figuur 1), een fysiek balansmodel met een horizontale balk waar aan weerszijden twee kettingen met gekleurde balletjes van verschillende gewichten hangen, stond centraal in deze lessenserie. De belangrijkste taak voor de leerlingen was om “te ontdekken wat je allemaal kan doen, terwijl je zorgt dat de hangmobiel recht blijft”. Leerlingen konden verschillende handelingen uitvoeren, wat resulteerde in het al dan niet uit evenwicht brengen van de horizontale balk. Het in balans, en dus horizontaal, houden van deze balk staat model voor het behouden van de gelijkheid in een vergelijking. Begrip van dit gelijkheidsconcept is cruciaal voor het leren oplossen van vergelijkingen. Tijdens het in balans houden van het model, konden leerlingen op een intuïtieve manier informele algebraïsche strategieën toepassen en ontwikkelen. Zo konden ze bijvoorbeeld de volgorde van de balletjes aan één kant van het model veranderen, balletjes met gelijke kleuren aan beide kanten weghalen of één kleur balletjes vervangen door balletjes van een andere kleur. Deze contextgebonden algebraïsche strategieën liggen ten grondslag aan conventionele strategieën voor het oplossen van lineaire vergelijkingen, zoals herstructureren, isoleren en substitueren. De verwachting was dat de fysieke ervaringen met het model zouden bijdragen aan het ontwikkelen van begrip van het concept gelijkheid en strategieën om deze gelijkheid te behouden. Het idee dat fysieke ervaringen waardevol zijn voor het begrijpen van abstracte ideeën, zoals wiskundige concepten, komt terug in theorieën over *embodied cognition* (vertaling: belichaamde cognitie).

Figuur 1. De hangmobiel, het balansmodel dat centraal stond in onze lessenserie over lineaire vergelijkingen.

Door leerlingen met het balansmodel te laten werken, beoogden we een basis mee te geven voor het oplossen van vergelijkingen in verschillende contexten, en uiteindelijk zelfs voor het oplossen van systemen van vergelijkingen. Het oplossen van systemen van vergelijkingen vraagt om geavanceerde algebraïsche vaardigheden, zoals het redeneren over relaties tussen onbekenden in vergelijkingen en over de onderlinge relaties tussen vergelijkingen (oftewel, het redeneren over covariantie), het manipuleren van deze relaties, het combineren ervan en het redeneren op basis van deze nieuwverworven informatie. Dit soort vaardigheden wordt ook wel gezien als een belangrijk aspect van wiskundige HOV. Het redeneren over informele vergelijkingen, waarbij wordt voortgebouwd op de intuïtieve ervaringen van leerlingen met balans (variërend van leren lopen tot het spelen op een wip) en de kennis die ze al hebben, lijkt dus een kansrijke aanpak voor het ontwikkelen van de HOV van basisschoolleerlingen.

Het eerste doel van dit proefschrift was om de rol van het balansmodel bij het lesgeven over lineaire vergelijkingen in kaart te brengen (*Hoofdstuk 2*). Hiervoor hebben we een literatuurstudie uitgevoerd. Het tweede doel was om de potentie van verschillende representaties van het balansmodel voor het bevorderen van het begrip van lineaire vergelijkingen bij basisschoolleerlingen te onderzoeken (*Hoofdstuk 3* en *4*). Hiervoor hebben we de eerdergenoemde lessenserie ontwikkeld en onderwezen in groep 7 van verschillende basisscholen en hebben we het effect van deze lessen op de ontwikkeling van het algebraïsch redeneren van de leerlingen bekeken. Het laatste doel van dit proefschrift was om na te gaan in hoeverre het stimuleren van het redeneren van basisschoolleerlingen over vergelijkingen mogelijk ook bevorderlijk is voor hun redeneren binnen een ander wiskundig domein waarin redeneren over covariantie eveneens een rol speelt: grafieken (*Hoofdstuk 5*). Daarom hebben we onderzocht wat het effect was van de lessenserie over lineaire vergelijkingen op het redeneren over grafieken van beweging.

Hoofdstuk 2 beschrijft een systematisch literatuuronderzoek naar het balansmodel. We zochten in 93 internationale *peer-reviewed* tijdschriften naar artikelen waarin het balansmodel werd gebruikt voor het lesgeven over lineaire vergelijkingen. Dit resulteerde in een selectie van 34 artikelen, die we analyseerden wat betreft de rationale(s) die werd(en) aangedragen voor het gebruik van het model, de verschijningsvorm van het model, de situatie waarin het model werd gebruikt en het effect van het gebruik van het model op de leeruitkomsten van de leerlingen. Ons doel was om een overzicht te creëren van het gebruik van het balansmodel bij het lesgeven over vergelijkingen, dat mogelijk bruikbaar zou kunnen zijn voor leerkrachten, onderzoekers en ontwikkelaars van instructiemateriaal voor het maken van onderbouwde keuzes om dit didactische model al dan niet te gebruiken.

De resultaten van deze literatuurstudie lieten een zeer gevarieerd beeld zien van het gebruik van het balansmodel. Het model werd het meest gebruikt om het begrip van gelijkheid in een vergelijking te bevorderen en van strategieën om deze gelijkheid te behouden (zoals hetzelfde weghalen aan beide kanten van de vergelijking).

Daarnaast werden eerdere fysieke ervaringen met balans of huidige fysieke ervaringen met een concreet balansmodel genoemd als argumenten voor het gebruik van het model. Als beperking van het model werden de geringe mogelijkheden om vergelijkingen met negatieve getallen of aftrekkingen weer te geven aangedragen. Drie verschijningsvormen van het model kwamen naar voren: fysieke modellen, virtuele modellen en getekende modellen. De situaties waarin het model werd gebruikt liepen uiteen wat betreft de leeftijd van de leerlingen (van de kleuterschool tot de derde klas van de middelbare school) en hun eerdere ervaringen met algebra, de duur van de interventie (van één activiteit tot lessen gedurende een aantal jaar), het type vergelijkingen dat werd onderwezen (bijvoorbeeld vergelijkingen die wel of geen negatieve waarden bevatten) en het type instructie dat werd gegeven (bijvoorbeeld klassikale instructie door een leerkracht of het individueel exploreren van het model). Ten slotte liepen ook de gerapporteerde leeruitkomsten van de studies erg uiteen, variërend van positief tot voornamelijk negatief.

Binnen dit zeer uiteenlopende beeld waren echter wel enige trends zichtbaar. Fysieke en virtuele balansmodellen werden het meest gebruikt voor het lesgeven aan leerlingen zonder eerdere ervaring met algebra. Argumenten die hierbij vaak werden aangedragen, waren de geschiktheid van het model voor het vergroten van het begrip van gelijkheid en de fysieke ervaringen met het model. De modellen werden meestal gebruikt voor het onderwijzen van vergelijkingen met positieve getallen en optellingen. Het gebruik van dit type modellen leek gemiddeld genomen een positief effect te hebben op het leren over vergelijkingen. Getekende modellen, daarentegen, werden het meest gebruikt voor het onderwijzen van leerlingen die gemiddeld al iets ouder waren en al enige ervaring hadden met algebra. Aan deze modellen werden regelmatig extra kenmerken toegevoegd die het mogelijk maakten om een bredere range van vergelijkingen te representeren, zoals vergelijkingen met negatieve waarden. De rationales voor het gebruik van dit type model waren vaak minder duidelijk gedefinieerd. En als laatste liepen de gerapporteerde effecten van het gebruik van dit type model op het leren over vergelijkingen behoorlijk uiteen. Het is echter belangrijk om te benadrukken dat de kenmerken van geïncludeerde studies van dit literatuuronderzoek erg varieerden. De hierboven beschreven trends moeten daarom voorzichtig geïnterpreteerd worden. Precieze kennis ontbreekt over welk type balansmodel in welke situatie leidt tot de beste leeruitkomsten, en verder onderzoek is nodig om dit model optimaal in te kunnen zetten voor het leren oplossen van vergelijkingen.

De studies beschreven in *Hoofdstuk 3* en *4* doen iets aan dit hiaat in de onderzoeksliteratuur over het balansmodel. Voor deze studies ontwikkelden we een interventie bestaande uit een reeks van zes lessen met een balansmodel. Het doel van de lessenserie was het stimuleren van het algebraïsch redeneren van basisschoolleerlingen over vergelijkingen. Twee parallelle versies van de lessenserie werden ontwikkeld. In de ene versie werd een fysiek balansmodel in de vorm van een hangmobiel (zie Figuur 1) gebruikt in combinatie met een representatie van het model op papier; in de andere versie werd alleen een representatie van het model op papier gebruikt.

De in *Hoofdstuk 3* beschreven studie richtte zich op de ontwikkeling van het algebraïsch redeneren van 65 leerlingen, zonder eerdere ervaring met algebra, gedurende de lessen met het fysieke balansmodel. Hiervoor analyseerden we het redeneren van de leerlingen over systemen van twee lineaire vergelijkingen. Om een dergelijk systeem van vergelijkingen op te lossen, moet de leerling de informatie uit beide vergelijkingen combineren. Voor de analyse van het redeneren van de leerlingen hebben we gekeken naar het aantal vergelijkingen dat werd meegenomen in het redeneren: geen enkele vergelijking (Niveau R0), maar één van beide vergelijkingen (Niveau R1) of beide vergelijkingen (Niveau R2).

Uit de resultaten bleek dat leerlingen duidelijk vooruitgingen gedurende de lessen. Na het volgen van de lessen kon bijvoorbeeld 77% van de leerlingen de waarden van de onbekenden M en L vinden in het systeem van vergelijkingen $M + 3L = 25$ en $2M = 4L$. Gedurende de lessenserie nam het percentage leerlingen dat redeneerde zonder gebruik te maken van de gegeven vergelijkingen (Niveau R0) af, van 57% naar 20%, terwijl het percentage leerlingen dat beide vergelijkingen combineerde in hun redeneren (Niveau R2) steeg van 17% naar 52%. Daarnaast werden leerlingen beter in het uitleggen van hun redeneren en in het opschrijven van de gebruikte algebraïsche strategieën. Ook op individueel niveau verbeterde het redeneren van de meeste leerlingen gedurende de lessenserie.

We waren eveneens geïnteresseerd in de relatie tussen de ervaringen van de leerlingen in de leeromgeving met het fysieke balansmodel en hun gebruik van algebraïsche strategieën. Dit werd onderzocht door een analyse van video's van interacties in de klas gedurende de lessen, het werk van leerlingen tijdens de lessen en de antwoorden van de leerlingen op de les-specifieke taken die na afloop van elke les werden afgenomen. Er werden verschillende indicaties gevonden voor een relatie tussen de ervaringen van leerlingen met het fysieke balansmodel en hun gebruik van strategieën en notaties. Leerlingen gebruikten bijvoorbeeld de context-gebonden strategieën die ze hadden ontwikkeld tijdens het werken met het balansmodel, zoals herstructureren, isoleren en substitueren, om problemen op te lossen in andere contexten. Daarnaast maakten sommige leerlingen gebruik van een representatie van het model voor het oplossen van een systeem van twee symbolische vergelijkingen.

De leeromgeving met het fysieke balansmodel bleek dus een goede context voor het ontwikkelen van algebraïsch redeneren. Onze hypothese was dat de fysieke ervaringen met het concrete balansmodel hier een belangrijk aandeel in hadden. Het design van onze studie was echter niet geschikt om de specifieke bijdrage van deze fysieke ervaringen op een systematische manier te onderzoeken.

Hoofdstuk 4 rapporteert over een quasi-experimentele studie waarin we zowel de effecten van onze lessenserie op de langere termijn, als de specifieke bijdrage van fysieke ervaringen met het balansmodel onderzochten. Aan dit onderzoek namen 212 leerlingen uit groep 7 deel, die niet eerder les hadden gehad in algebra. De leerlingen werden verdeeld over twee interventiecondities en een controleconditie. Drie klassen kregen les met een representatie van het balansmodel op papier, drie klassen kregen exact dezelfde lessen maar dan met de toevoeging van een fysiek balansmodel (zie ook *Hoofdstuk 3*), en drie klassen in de controleconditie kregen les over een ander wiskunde onderwerp, namelijk over kans. In deze studie gebruikten we een zogenoemd *staged-comparison design*. Dit hield in dat het redeneren van leerlingen over systemen van twee informele lineaire vergelijkingen op vier momenten gedurende het schooljaar gemeten werd. Tussen twee van deze meetmomenten kregen leerlingen de lessen, in drie opeenvolgende cohorten, met één klas per cohort voor elke conditie. Het effect van de interventie op het algebraïsch redeneren en de mogelijke verschillen tussen condities, werden zowel kwantitatief (met een latent groeicurvemodel) als kwalitatief onderzocht.

Deelname aan de lessen leidde tot een hoger niveau van redeneren over systemen van twee informele lineaire vergelijkingen ($d = 0.73$; dit betekent een middelmatig tot groot effect). De leerlingen werden beter in het combineren van de informatie van beide vergelijkingen in hun redeneren: dit hoogste niveau van redeneren kwam meer voor na afloop van de interventie (65%) dan ervoor (39%). Het kwam tegelijkertijd juist minder voor dat leerlingen geen enkele van de vergelijkingen of maar één van beide vergelijkingen gebruikten in hun redeneren. Hoewel leerlingen die les hadden gehad met zowel een fysiek balansmodel als een representatie van het model op papier een grotere vooruitgang in redeneren (een toename van 33% in het hoogste niveau van redeneren) lieten zien dan leerlingen die alleen les hadden gehad met een representatie van het model op papier (toename van 18% in het hoogste niveau van redeneren), was dit verschil niet significant ($p = .136$). Inzoomen op het redeneren van de leerlingen liet daarentegen wel zien dat de leerlingen die ook met het fysieke model hadden gewerkt vaker een representatie van het model gebruiken (17% vs. 1%) en vaker gebruik maakten van geavanceerde algebraïsche strategieën zoals substitutie of eliminatie (60% vs. 40%).

In de vorige twee studies toonden we het effect van onze lessenserie op het redeneren over systemen van vergelijkingen aan. Het redeneren over en het oplossen van dergelijke systemen van vergelijkingen vraagt om wiskundige HOV, zoals het vermogen van de leerlingen om te redeneren over covariantie, oftewel, redeneren over de samenhang tussen bepaalde variabelen. Dit redeneren over covariantie is ook relevant binnen andere wiskundige domeinen, zoals grafieken van beweging. Dit roept de vraag op of het stimuleren van het redeneren over vergelijkingen mogelijk ook leidt tot een verbetering in het redeneren over grafieken. In ons laatste onderzoek, beschreven in *Hoofdstuk 5*, hebben we het effect van onze lessenserie over vergelijkingen op zowel het redeneren over vergelijkingen als het redeneren over grafieken van beweging van 132 leerlingen uit groep 7 onderzocht. Het algebraïsch redeneren en het grafisch redeneren van de leerlingen werd gedurende het schooljaar vier keer gemeten. Tijdens elk van deze meetmomenten werkten leerlingen aan vier algebrataken waarin ze een systeem van twee informele vergelijkingen oplosten en aan vier grafiektaken waarin ze een tijd-afstand grafiek interpreteerden of construeerden.

Om het effect van onze interventie op het redeneren over zowel algebra als grafieken te onderzoeken, ontwikkelden we een latent groeimodel waarin de latente algebraïsche en de latente grafische redeneervaardigheid samen werden gemodelleerd. De resultaten, die voorzichtig geïnterpreteerd moeten worden vanwege een suboptimale fit van het model, lieten een verbetering zien in het algebraïsch redeneren van leerlingen na het deelnemen aan de lessen over vergelijkingen. Dit was een bevestiging van onze eerdere bevindingen (*Hoofdstuk 4*). Deelname aan deze lessen leidde echter niet tot een verbetering van het grafisch redeneren van deze leerlingen. Op basis hiervan kunnen we voorzichtig de conclusie trekken dat het ontwikkelen van HOV gedurende de lessen over vergelijkingen niet resulteerde in het ontwikkelen van HOV binnen het domein van grafieken. De overdracht van HOV van het ene wiskundige domein naar een (gerelateerd) ander wiskundig domein lijkt dus niet vanzelfsprekend.

In *Hoofdstuk 6* worden de resultaten van de verschillende onderzoeken samengevat en worden de implicaties van de resultaten besproken. Op basis van de resultaten van de afzonderlijke studies kunnen we een aantal algemene conclusies trekken. Ons onderzoek laat duidelijk zien dat het mogelijk is om het algebraïsch redeneren van leerlingen op de basisschool te stimuleren. We komen dan ook met de sterke aanbeveling om algebra op te nemen in het reken-wiskundecurriculum van de basisschool. Door al op de basisschool te starten met het aanbieden van activiteiten die het (informeel) algebraïsch redeneren van leerlingen bevorderen, kan een doorlopende leerlijn gecreëerd worden van de basisschool tot en met de middelbare school. Het belang van zulke doorlopende leerlijnen in het reken-wiskundeonderwijs wordt steeds meer onderschreven op zowel nationaal als internationaal niveau. Daarnaast kan het stimuleren van het algebraïsch redeneren gezien worden als een manier om het reken-wiskundecurriculum op de basisschool te verrijken. Dit algebraïsch redeneren kan gezien worden als een domein-specifieke operationalisatie van wiskundige HOV binnen het basisonderwijs.

Het opnemen van algebra in het basisonderwijs brengt echter wel bepaalde uitdagingen met zich mee. Of, om in de woorden van de titel van dit proefschrift te spreken: op een aantal punten is het nog wel zoeken naar balans. De eerste uitdaging is om nog meer zicht te krijgen op hoe het algebraïsch redeneren van jonge leerlingen het best gestimuleerd kan worden. Voor het stimuleren van redeneren over vergelijkingen blijken activiteiten met een balansmodel waarmee leerlingen redeneren over relaties tussen onbekenden waardevol te zijn. Hierdoor ontwikkelden leerlingen in ons onderzoek verschillende algebraïsche strategieën, die ze in een later stadium konden gebruiken voor het oplossen van (systemen van) vergelijkingen in nieuwe contexten en zelfs voor het oplossen van systemen van formele vergelijkingen. Daarentegen kunnen we op basis van ons onderzoek geen eenduidige conclusies trekken over de toegevoegde waarde van de fysieke ervaringen met een concreet balansmodel. Dit moet dan ook verder onderzocht worden. De tweede uitdaging is dat het huidige (reken-wiskunde)curriculum op de basisschool al overvol is. We moeten ervoor waken dat het opnemen van algebra in dit curriculum er niet toe leidt dat het curriculum alleen maar voller wordt. Algebra moet dus niet simpelweg toegevoegd worden aan het curriculum; het moet erin *geïntegreerd* worden. Door kleine aanpassingen door te voeren in de activiteiten die momenteel al plaatsvinden op de bassischool, kunnen deze worden ingezet om het algebraïsch redeneren van leerlingen te stimuleren. De derde uitdaging is het voorbereiden van basisschoolleerkrachten op het onderwijzen van algebra. Hierbij is het van belang dat ze meer inzicht krijgen in de aard van dergelijke activiteiten, in het belang van het integreren van dit soort activiteiten in het basisonderwijs en in het vermogen van basisschoolleerlingen om algebraïsch te redeneren. Dit vereist aanpassingen in de lerarenopleiding en het vraagt om nascholingsactiviteiten. Ten slotte is de laatste uitdaging om ervoor te zorgen dat *alle* leerlingen deel kunnen nemen aan activiteiten die het algebraïsch redeneren stimuleren, en niet alleen een select groepje best-presterende leerlingen. Een manier om dit te verwezenlijken, is door (alle) leerlingen al op jonge leeftijd te laten participeren in activiteiten die voortbouwen op hun intuïtieve ervaringen met balans, en door dit als basis te nemen voor het stimuleren van hun algebraïsch redeneren.

Dankwoord (acknowledgements)

Tijdens mijn promotie heb ik verschillende keren nagedacht over dit moment: het schrijven van dit dankwoord. Ik kon me moeilijk indenken hoe het zou voelen om de allerlaatste woorden van dit proefschrift op papier te zetten, de laatste pagina's te gebruiken om mensen te bedanken die mij, op wat voor manier dan ook, hebben geholpen bij de totstandbrenging hiervan. Het voelt tegelijkertijd gek en fantastisch dat ik nu daadwerkelijk op dit punt ben aangekomen.

Allereerst wil ik graag mijn begeleiders bedanken. Marja, je was voor mij een promotor met een kritische blik en een bewonderenswaardige hoeveelheid kennis van de reken-wiskunde didactiek. Ik heb veel geleerd tijdens onze samenwerking, waarvoor dank. Michiel, je blijft me keer op keer verbazen met je geheugen en je scherpte. Je grote betrokkenheid, bij het project maar ook zeker persoonlijk, heeft mij erg geholpen tijdens deze promotie. Vooral de carpoolgesprekken en de gesprekken in de vroege uurtjes op kantoor heb ik als ontzettend waardevol ervaren. Aiso, thank you for the nice discussions we have had at IPN and the UU, and for your valuable comments on my articles, especially during the last stage of my PhD.

Ook wil ik alle andere leden van het projectteam bedanken. Jan, hoeveel uren hebben we niet achter jouw computer doorgebracht… Hoewel het me regelmatig duizelde tijdens onze besprekingen, kijk ik er met plezier op terug. Paul en Michiel D., ik vond het fijn om af en toe even met jullie te kunnen sparren over zowel promotie als over niet-promotie gerelateerde zaken. Anke, thank you for your input in the beginning of the project during the meetings in Kiel and Utrecht.

Dit proefschrift was er nooit geweest zonder de medewerking van een groot aantal basisscholen. Leerlingen en leerkrachten van De Molenwiek (Jikke, Martien, Azadeh, Sandra, Mandy en Mara) Liduina (Mariëlle en Anniek), Piet Hein (Wendy en Floortje), Dorpsakker (Nathaniël en Petra), Tijo van Eeghen (Sofie en Marieke), Sparrenbos (Sharon en Marjolein), Princes Beatrix (Lisette en Laurens), Piramide (Marjolein) en Durv! (Wijmke), ontzettend bedankt! Mijn dank gaat eveneens uit naar onderzoeksassistenten en stagiaires die hebben geholpen met de dataverzameling en het coderen van de data en naar mijn nichtje Lala en neefje Mac voor het pilotten van de eerste activiteiten en jullie interessante redeneringen.

Carolien en Femke, wat voelt het goed om jullie hier als paranimfen achter mij te hebben staan. Lieve Carolien, twee toch wel behoorlijk verschillende mensen in een heel vergelijkbare situatie, kan leiden tot heel mooie dingen. Van het lunchen met lofbroodjes tot het stappen in New York tot het samen werken aan een artikel of aan de

omslag van dit proefschrift; we hebben heel wat beleefd de afgelopen jaren. Jij was er altijd. Dankjewel! Lieve Fem, niemand anders had hier kunnen staan. Vanaf ons 5^{e} jaar zijn we zo ongeveer onafscheidelijk geweest. Hoewel we tegenwoordig niet meer eerst een uur bellen om vervolgens te besluiten om 160 meter naar elkaar toe te lopen, is het gevoel nog altijd hetzelfde: het is fijn om te weten dat je er altijd zal zijn.

De afgelopen jaren heb ik met verschillende mensen de kamer mogen delen. Lieve Roos, tegen jouw enthousiasme en positiviteit kunnen er maar weinig op. Het is mooi om je nu zo gelukkig te zien. Dear Xiaoyan, as a colleague you already were a big inspiration for me, and today, you still are. I miss you! Ali, thank you for your kindness and for telling me many beautiful stories about your background. Ilona, met je ervaring in het onderzoek en je kalme uitstraling heb je voor mij heel wat dingen in perspectief weten te plaatsen. Dankjewel daarvoor. Ook wil ik mijn andere collega's graag bedanken. Van de afdeling Orthopedagogiek voor de vele lunchmomenten, praatjes bij de koffie automaat en afdelingsuitjes. En van het FI voor de interessante gesprekken over reken-wiskunde onderwijs.

Daarnaast heb ik gedurende dit promotietraject ontzettend veel familie en vrienden om me heen gehad die hebben bijgedragen door middel van een bemoedigend woordje of juist door het verzetten van mijn gedachten. Lieve vrienden, de avonden en weekendjes met (of eigenlijk vaker zonder) Ligretto, de uren op het voetbalveld – één van de enige plekken waar ik mijn hoofd écht helemaal leeg kreeg – met een biertje na afloop, de zondagen met Formule 1 en pizza, de spontane vrijmibo's en de vele etentjes, waren voor mij ontzettend waardevol. Lieve (schoon)familie, Mieke en Lon, keer op keer stonden jullie klaar met advies of met krantenartikelen over het reilen en zeilen op universiteiten. Tom en Daan, ook jullie vroegen regelmatig, elk op jullie eigen manier, aan mij hoe het op mijn werk was (en, ok, ook waarom ik dit ook alweer precies deed). Lieve pap en mam, de interesse in het werken met en het leren van kinderen en de voorliefde voor rekenen en wiskunde heb ik met de paplepel ingegoten gekregen. Dit, samen met jullie onvoorwaardelijke steun, maakt dat ik hier nu sta. Dank jullie wel.

Lieve Gijs, er is niemand die dit traject van zo dichtbij heeft meegemaakt als jij. Ondanks dat je mijn keuzes lang niet altijd begreep, heb je altijd volledig achter me gestaan. Met name in de laatste fase heb jij alles in het werk gesteld om mij dit traject af te kunnen laten ronden. Dit vereiste ook van jou een lange adem. Ik ben je ongelooflijk dankbaar en zo gelukkig met jou aan mijn zijde. En dan als laatste, onze lieve kleine Saar. Zo klein als je bent, zo veel leer ik nu al van jou. Wat breng jij een hoop moois in ons leven. Ik kan niet wachten om je bij te staan tijdens je verdere ontdekking van deze wereld.

About the author

Mara Otten was born on December 14, 1990, in Haarlem, the Netherlands. She obtained her secondary school degree (Atheneum) at the Schoter Scholengemeenschap in Haarlem in 2008. In 2009, she enrolled in a bachelor program in Psychology at the VU University in Amsterdam, graduating with an honors bachelor's degree in Neuropsychology in 2012. She continued with a Research Master in Cognitive Neuropsychology at the VU University. After finishing her research internship at the KU Leuven (Belgium), in which she wrote her thesis about an fMRI study on phonological processing and arithmetic fact retrieval in primary school children, Mara graduated with honors in 2014. In 2015, after a year of teaching various psychology,= and statistics courses to bachelor students at the VU University and supervising third year bachelor students Clinical Neuropsychology at the University of Amsterdam with writing their thesis, she started her PhD research in mathematics education at Utrecht University. This PhD research was carried out under the supervisor of Prof. dr. Marja van den Heuvel-Panhuizen and Dr. Michiel Veldhuis of Utrecht University and Prof. dr. Aiso Heinze of IPN Leibniz Institute for Science and Mathematics Education, Kiel (Germany). This research was part of the larger Beyond Flatland project, which was carried out with a grant from The Netherlands Initiative for Education Research. The Beyond Flatland research project focused on enriching the Dutch primary school mathematics curriculum by incorporating higher-order mathematical activities in the primary school classroom; Mara's PhD project particularly focused on stimulating primary school students' algebraic reasoning. Over the course of the project, she published several articles in international peer-reviewed journals and in national scholarly journals, and she presented the research findings at national and international conferences. Mara's PhD research fulfilled all requirements of the Interuniversity Center of Educational Sciences (ICO) Research School in the Netherlands. Mara currently works as a mathematics teacher educator at the iPabo University of Applied Sciences in Amsterdam/Alkmaar.

List of publications related to this thesis

Scholary publications:

Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, M., Boom, J., & Heinze, A. (2020). Are physical experiences with the balance model beneficial for students' algebraic reasoning? An evaluation of two learning environments for linear equations. *Education Sciences, 10*(6), 163. https://doi.org/10.3390/educsci10060163

Otten, M., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2019). The balance model for teaching linear equations: A systematic literature review. *International Journal of STEM Education*, *6*(1), 30–51. https://doi.org/10.1186/s40594-019-0183-2

Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, M., & Heinze, A. (2019). Developing algebraic reasoning in primary school using a hanging mobile as a learning supportive tool / El desarrollo del razonamiento algebraico en educación primaria utilizando una balanza como herramienta de apoyo. *Journal for the Study of Education and Development / Infancia y Aprendizaje*, *42*(3), 615–663. https://doi.org/10.1080/02103702.2019.1612137

Conference proceedings:

Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, M., & Heinze, A. (2019). Fifth-grade students solving linear equations supported by physical experiences. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education* (pp. 646–653). Freudenthal Group & Freudenthal Institute, Utrecht University and ERME.

Professional publications:

Otten, M. (2018). Het fizier gericht op... algebraïsch redeneren [Fi-visor on… algebraic reasoning]. *Euclides*, *93*(6), 32–34

Otten, M., Van den Heuvel-Panhuizen, M., Veldhuis, M., Heinze, A., & Goldenberg, P. (2017). Eliciting algebraic reasoning with hanging mobiles. *Australian Primary Mathematics Classroom, 22*(3), 14–19.

Otten, M., Van den Heuvel-Panhuizen, M., & Veldhuis, M. (2016). Lala en de hangmobiel: Redeneren met onbekenden [Lala and the hanging mobile: Reasoning with unknowns]. *Volgens Bartjens, 35*(4), 8–11.

+

ICO dissertation series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, Radboud University Nijmegen, University of Amsterdam, University of Antwerp, University of Ghent, University of Groningen, University of Twente, Utrecht University, Vrije Universiteit Amsterdam, and Wageningen University, and formerly Tilburg University (until 2002). The University of Groningen, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series.

430. Ackermans, K. (20-12-2019) *Designing Video-Enhanced Rubrics to Master Complex Skills*. Heerlen: Open University of the Netherlands
429. Post, T. (11-12-2019) *Fostering inquiry-based pedagogy in primary school: a longitudinal study into the effects of a two-year school improvement project*. Enschede: University of Twente
428. Harmsen, R. (10-12-2019) *Let's talk about stress. Beginning secondary school teachers' stress in the context of induction programmes*. Groningen: University of Groningen
427. Hopster-Den Otter, D. (28-11-2019) *Formative assessment design: A balancing act.* Enschede: University of Twente
426. De Bruijn, A.G.M. (14-11-2019) *The brain in motion: Effects of different types of physical activity on primary school children's academic achievement and brain activation*. Groningen: University of Groningen
425. Van der Zanden, P. (05-11-2019) *First-year student success at university: Domains, predictors, and preperation in secondary education*. Nijmegen: Radboud University Nijmegen
424. Van Ginkel, S.O. (23-10-2019) *Fostering oral presentation competence in higher education*. Wageningen: Wageningen University
423. Jansen, R.S. (11-10-2019) *Dealing with autonomy: Self-regulated learning in open online education*. Utrecht: Utrecht University
422. Wu, Q. (02-10-2019) *Making Construct-Irrelevant Variance Relevant: Modelling item position effects and response behaviors on multiple-choice tests*. Leuven: KU Leuven
421. Sun, X. (13-09-2019) *Teacher-student interpersonal relationships in Chinese secondary education classrooms.* Utrecht: Utrecht University
420. Liu, M. (13-09-2019) *Exploring culture-related values in Chinese student teachers' professional self-understanding and teaching experiences*. Utrecht: Utrecht University
419. Henderikx, M.A. (06-09-2019) *Mind the Gap: Unravelling learner success and behaviour in Massive Open Online Courses*. Heerlen: Open University of the Netherlands
418. Van Kampen, E. (05-09-2019) *What's CLIL about bilingual education?* Leiden: Leiden University
417. Vossen, T.E. (04-09-2019) *Research and design in STEM education*. Leiden: Leiden University
416. Cuyvers, K. (28-08-2019) *Unravelling medical specialists self -regulated learning in the clinical environment*. Antwerp: University of Antwerp
415. Agricola, B.T. (21-06-2019) *Who's in control? Finding balance in student-teacher interactions*. Utrecht: Utrecht University

414. Van Oeveren, C.D.P. (12-06-2019) *ITHAKA gaf je de reis.* Amsterdam: Vrije Universiteit Amsterdam
413. Donszelman, S. (06-06-2019) *Doeltaal-leertaal didactiek, professionalisering en leereffecten*. Amsterdam: Vrije Universiteit Amsterdam
412. Woldman, N. (24-05-2019) *Competence development of temporary agency workers.* Wageningen: Wageningen University
411. Filius, R.M. (23-05-2019) *Peer feedback to promote deep learning in online education. Unravelling the process*. Utrecht: Utrecht University
410. Degrande, T. (08-05-2019) *To add or to multiply? An investigation of children's preference for additive or multiplicative relations*. Leuven: KU Leuven.
409. Schrijvers, M.S.T. (03-05-2019) *The story, the self, the other. Developing insight into human nature in the literature classroom.* Amsterdam: University of Amsterdam
408. Wanders, F. H. K. (03-05-2019). *The contribution of schools to societal participation of young adults: The role of teachers, parents, and friends in stimulating societal interest and societal involvement during adolescence*. Amsterdam: University of Amsterdam
407. Jonker, H.M. (09-04-2019) *Teachers' perceptions of the collaborative design and implementation of flexibility in a blended curriculum.* Amsterdam: University of Amsterdam
406. Schep, M. (06-03-2019) *Guidance for guiding. Professionalization of guides in museums of art and history*. Amsterdam: University of Amsterdam
405. Van der Meulen, A.N. (28-02-2019) *Social cognition of children and young adults in context*. Amsterdam: Vrije Universiteit Amsterdam
404. Dobbelaer, M.J. (22-02-2019) *The quality and qualities of classroom observation systems*. Enschede: University of Twente
403. Wijnen, M. (01-02-2019) *Introduction of problem-based learning at the Erasmus School of Law: Influences on study processes and outcomes*. Rotterdam: Erasmus University Rotterdam
402. Gast, I. (13-12-2018) Team-*Based Professional Development – Possibilities and challenges of collaborative curriculum design in higher education*. Enschede: University of Twente.
401. Fix, G.M. (20-12-2018) *The football stadium as classroom. Exploring a program for at-risk students in secondary vocational education*. Enschede: University of Twente.
400. Kippers, W.B. (28-11-2018) *Formative data use in schools. Unraveling the process*. Enschede: University of Twente.
399. Nederhand M.L. (22-11-2018) *Improving Calibration Accuracy Through Performance* Feedback. Rotterdam: Erasmus University Rotterdam.
398. De Vetten, A.J. (21-11-2018) *From sample to population.* Amsterdam: Vrije Universiteit Amsterdam.
397. Gresnigt, R. (08-11-2018) *Integrated curricula: An approach to strengthen Science & Technology in primary education.* Eindhoven: Eindhoven University of Technology.
396. Walhout, J.H. (26-10-2018) *Learning to organize digital information* Heerlen: Open University of the Netherlands.
395. Van Riesen, S.A.N. (26-10-2018) *Inquiring the effect of the experiment design tool: whose boat does it float?* Enschede: University of Twente.
394. Van der Wilt, F.M. (08-10-18) *Being rejected*. Amsterdam: Vrije Universiteit Amsterdam.
393. Van Laer, S. (4-10-2018) *Supporting learners in control: investigating self-regulated learning in blended learning environments.* Leuven: KU Leuven.
392. Boelens, R. (27-09-2018) *Studying blended learning designs for hands-on adult learners*. Ghent: Ghent University.

Appendices

Appendix 2.1 (Chapter 2)

Search queries used in Scopus and Eric

Search query used in Scopus:

((TITLE-ABS-KEY ("equation*" OR "equal sign" OR "equals sign" OR "equality" OR "equivalence") OR TITLE (balanc*)) AND TITLE-ABS-KEY (("algebra*" OR "mathematic*" OR ("equation*" and "unknown*") OR ("equation*" and "balance*") OR ("equation*" and "solv*"))) AND (SRCTITLE ("Arithmetic teacher" OR "Australian Mathematics Teacher" OR "Australian Primary Mathematics Classroom" OR "Australian Senior Mathematics Journal" OR "Canadian Journal of Science, Mathematics and Technology Education" OR "Educational Studies in Mathematics" OR "For the learning of mathematics" OR "International Journal for Mathematics Teaching and Learning" OR "International Journal of Mathematical Education in Science and Technology" OR "International Journal of Science and Mathematics Education" OR "International Journal on Mathematics Education" OR "Investigations in Mathematics Learning" OR "Journal for Research in Mathematics Education" OR "Journal of Computers in Mathematics and Science Teaching" OR "Journal of Mathematical Behavior" OR "Journal of Mathematics Teacher Education" OR "Mathematical Thinking and Learning" OR "Mathematics Education Research Journal" OR "Mathematics education" OR "Mathematics Educator" OR "Mathematics enthusiast" OR "Mathematics Teacher" OR "Mathematics Teaching in the Middle School" OR "Research in Mathematics Education" OR "Review of Educational Research" OR "Review of Research in Education" OR "School Science and Mathematics" OR "Teaching children mathematics" OR "ZDM International Journal on Mathematics Education" OR "American Educational Research Journal" OR "British educational research journal" OR "Child development" OR "Cognition and instruction" OR "Contemporary educational psychology" OR "Digital Experiences in Mathematics Education" OR "Educational Researcher" OR "Elementary school journal" OR "Instructional science" OR "International Journal of Educational Research" OR "Journal of Educational Psychology" OR "Journal of educational research" OR "Journal of pedagogy" OR "Journal of the Learning Sciences" OR "Learning and Instruction" OR "Pedagogies" OR "Technology, pedagogy and education" OR "Teaching and teacher education" OR "Teacher college record" OR "British Journal of Developmental Psychology" OR "Developmental psychology" OR "Developmental science" OR "European Journal of Developmental Psychology" OR "Frontiers in psychology" OR "Journal of applied developmental psychology" OR "Journal of Cognition and Development" OR "Journal of educational and developmental psychology" OR "British journal of educational psychology" OR "Educational and Psychological Measurement" OR "Educational psychologist" OR "Educational psychology" OR "Journal of Educational Measurement" OR "Journal of Educational Psychology" OR "Journal of school psychology" OR "Journal of the learning sciences" OR "American Journal of Education" OR "Asia-Pacific Education Researcher" OR "British Journal of Educational Technology" OR "Early Childhood Education Journal" OR "European Journal of Education" OR "International Journal of STEM Education" OR "Scandinavian Journal of Educational Research"

OR "British Journal of Special Education" OR "European Journal of Special Needs Education" OR "International Journal of Special Education" OR "Journal of Research in Special Educational Needs" OR "Journal of Special Education" OR "Journal of Special Education Research" OR "Research Based Journal in Special Education" OR "Journal für Mathematik-Didaktik " OR "Recherches en Didactique des Mathématiques" OR "Australasian Journal of Educational Technology" OR "Canadian Journal of Learning and Technology" OR "Computers & Education" OR " Design and Technology Education: an International Journal" OR "Educational Technology" OR "Educational Technology Research and Development" OR "International Journal of Educational Technology" OR "Interactive Learning Environments" OR "International Journal of Technology and Design Education" OR "Journal of Interactive Technology and Pedagogy" OR "Journal of Research on Technology in Education" OR "Journal of Teaching and Learning with Technology" OR "Journal of Technology and Teacher Education" OR "Journal of Technology Education")))

Search query used in Eric (searched via "advanced search")

(("equation*" or "equal sign" or "equals sign" or "equality" or "equivalence").ab. or "balanc*".ti.) AND (("algebra*" or "mathematic*" or ("equation*" and "unknown*") or ("equation*" and "balanc*") or ("equation*" and "solv*")).ab.) AND ("Arithmetic teacher" or "Australian Mathematics Teacher" or "Australian Primary Mathematics Classroom" or "Australian Senior Mathematics Journal" or "Canadian Journal of Science, Mathematics and Technology Education" or "Educational Studies in Mathematics" or "For the learning of mathematics" or "International Journal for Mathematics Teaching and Learning" or "International Journal of Mathematical Education in Science and Technology" or "International Journal of Science and Mathematics Education" or "International Journal on Mathematics Education" or "Investigations in Mathematics Learning" or "Journal for Research in Mathematics Education" or "Journal of Computers in Mathematics and Science Teaching" or "Journal of Mathematical Behavior" or "Journal of Mathematics Teacher Education" or "Mathematical Thinking and Learning" or "Mathematics Education Research Journal" or "Mathematics education" or "Mathematics Educator" or "Mathematics enthusiast" or "Mathematics Teacher" or "Mathematics Teaching in the Middle School" or "Research in Mathematics Education" or "Review of Educational Research" or "Review of Research in Education" or "School Science and Mathematics" or "Teaching children mathematics" or "ZDM International Journal on Mathematics Education" or "American Educational Research Journal" or "British educational research journal" or "Child development" or "Cognition and instruction" or "Contemporary educational psychology" or "Digital Experiences in Mathematics Education" or "Educational Researcher" or "Elementary school journal" or "Instructional science" or "International Journal of Educational Research" or "Journal of Educational Psychology" or "Journal of educational research" or "Journal of pedagogy" or "Journal of the Learning Sciences" or "Learning and Instruction" or "Pedagogies" or "Technology, pedagogy and education" or "Teaching and teacher education" or "Teacher college record" or "British Journal of Developmental Psychology" or "Developmental psychology" or "Developmental science" or "European Journal of Developmental Psychology" or "Frontiers in psychology" or "Journal of applied developmental psychology" or "Journal of Cognition and Development" or "Journal of educational and developmental psychology" or "British journal of educational psychology" or "Educational and Psychological Measurement" or

"Educational psychologist" or "Educational psychology" or "Journal of Educational Measurement" or "Journal of Educational Psychology" or "Journal of school psychology" or "Journal of the learning sciences" or "American Journal of Education" or "Asia-Pacific Education Researcher" or "British Journal of Educational Technology" or "Early Childhood Education Journal" or "European Journal of Education" or "International Journal of STEM Education" or "Scandinavian Journal of Educational Research" or "British Journal of Special Education" or "European Journal of Special Needs Education" or "International Journal of Special Education" or "Journal of Research in Special Educational Needs" or "Journal of Special Education" or "Journal of Special Education Research" or "Research Based Journal in Special Education" or "Journal für Mathematik-Didaktik" or "Recherches en Didactique des Mathématiques" or "Australasian Journal of Educational Technology" or "Canadian Journal of Learning and Technology" or "Computers & Education" or "Design and Technology Education: an International Journal" or "Educational Technology" or "Educational Technology Research and Development" or "International Journal of Educational Technology" or "Interactive Learning Environments" or "International Journal of Technology and Design Education" or "Journal of Interactive Technology and Pedagogy" or "Journal of Research on Technology in Education" or "Journal of Teaching and Learning with Technology" or "Journal of Technology and Teacher Education" or "Journal of Technology Education").jn.

Appendix 3.1 (Chapter 3)

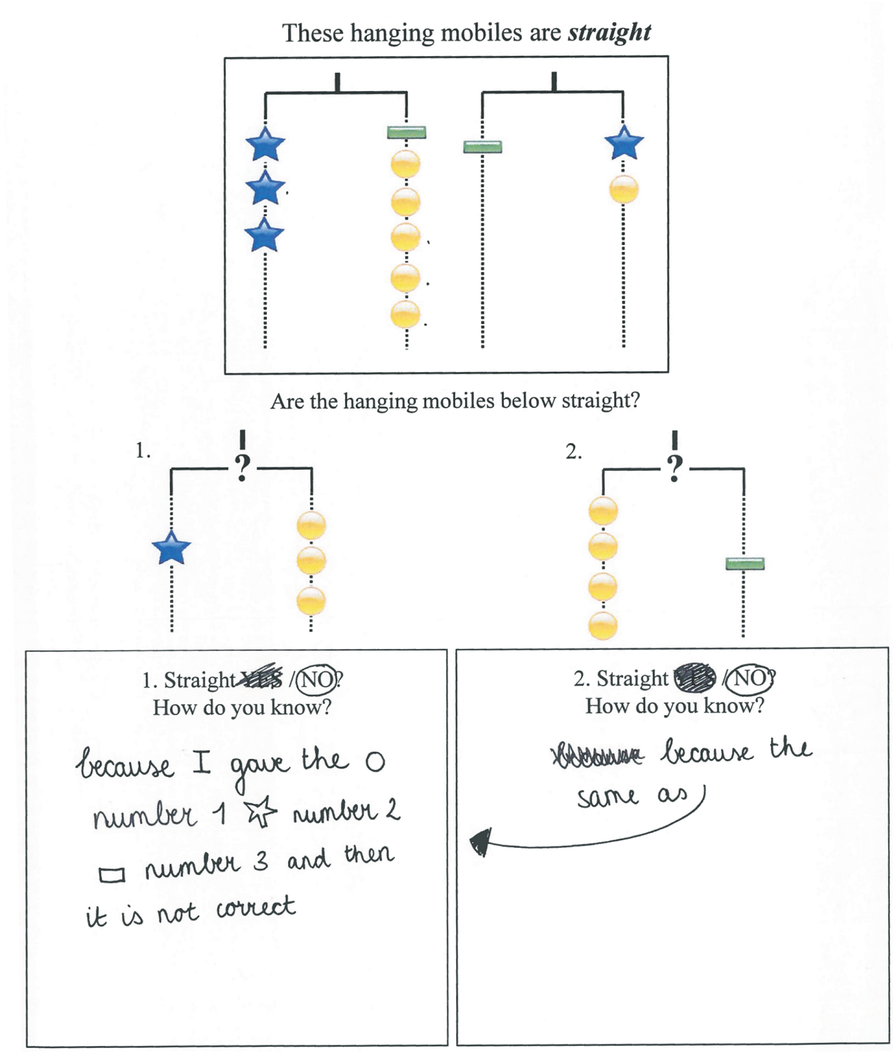


Figure A1. Julia's solution of the assessment task at the end of Episode 2 (text translated from Dutch). *Note*. Reasoning on the basis of one of the two given equations (Level R1), without showing strategies.

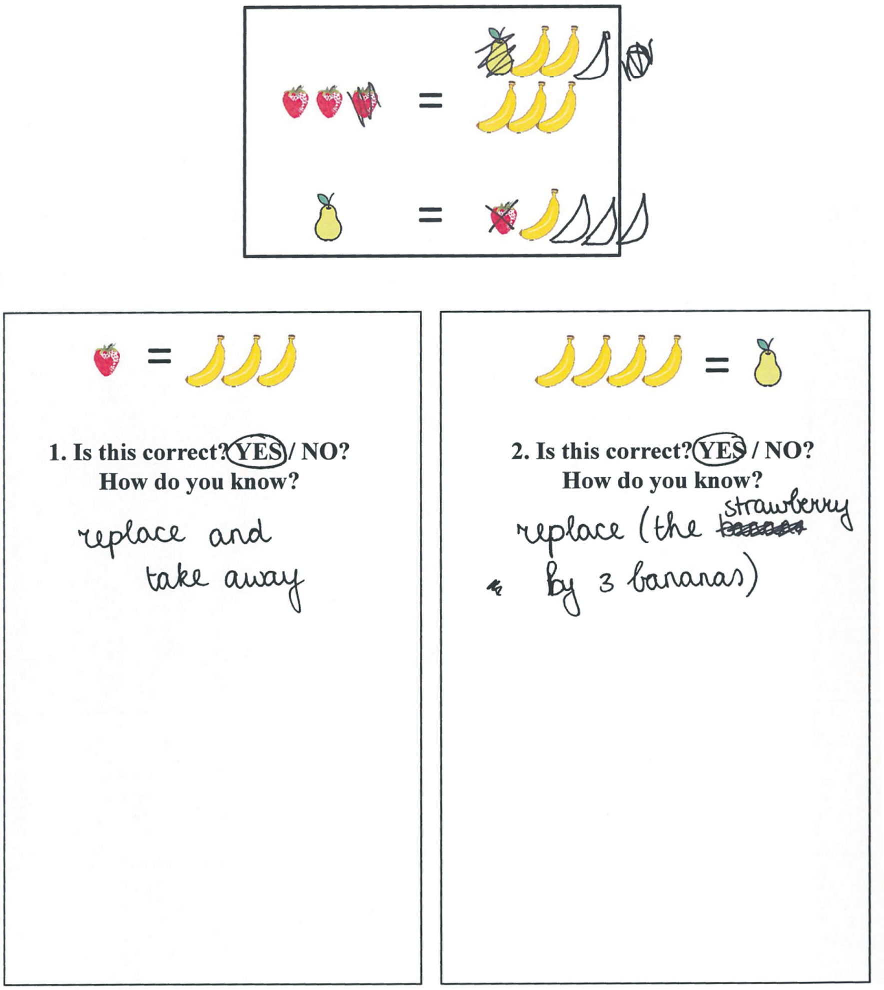


Figure A2. Julia's solution of the assessment task at the end of Episode 3 (text translated from Dutch). *Note*. Reasoning on the basis of both given equations by combining the information of both of them (Level R2), with showing strategies.

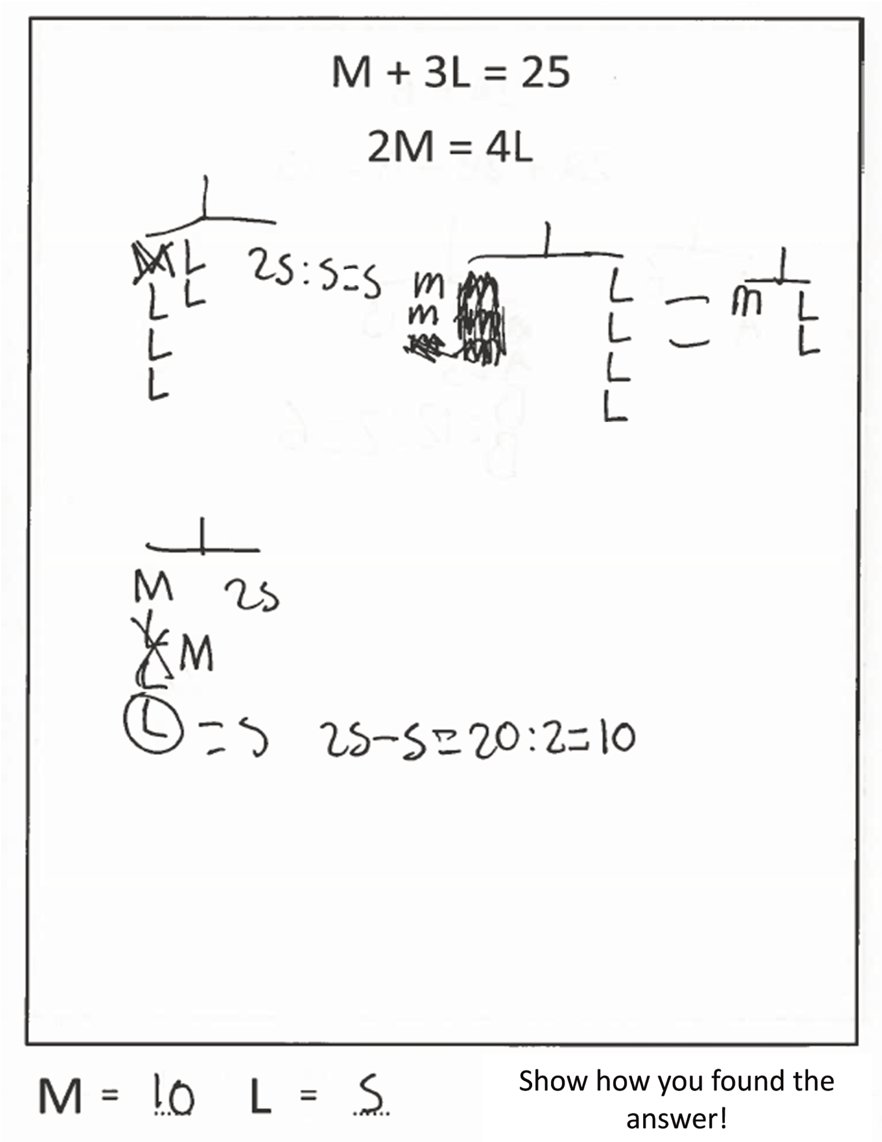


Figure A3. Julia's solution of the assessment task at the end of Episode 4 (text translated from Dutch). *Note*. Reasoning on the basis of both given equations by combining the information of both of them (Level R2), with showing strategies.

Appendix 3.2 (Chapter 3)

Coding scheme with examples of student responses; text in between square brackets is added as a clarification.

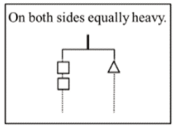
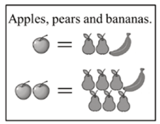
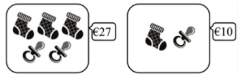
Level of reasoning	Description	Subtypes	Assessment task Episode 2	Assessment task Episode 3	Assessment task Episode 4
R0	Student does not use any of the given equations	R0_empty	-	-	- [No response]
		R0_don"t know	- "?"	-	- "?"
		R0_refer to or repeat example	- "Because a rectangle is just as much as a circle and a star" - "Look at the examples"	- "Look at the example" - "Because otherwise it is not correct"	- "$M + M = L + L + L + L$" - [Student converts equations into hanging mobiles without further reasoning]
		R0_general description	- "We have learned this!" - "By making a guess"	- "I think so" - "By making the right calculation"	- "By guessing with numbers" - "Guess and see"
		R0_unclear answer	- "Because I discovered that the content of one star equals three"	- "Because there are no four bananas" [question 2]	- [Student converts *both* equations into *one* hanging mobile and crosses out some unknowns]
		R0_remaining	- "One star equals three yellow circles" [question 1]; "1 rectangle = 4 circles" [question 2]	-"1x strawberry = 3x banana" [question 1]; "because 4 bananas are worth one pear!" [question 2] - "one strawberry should be two pears" [question 1]	-

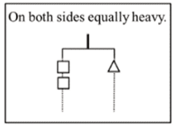
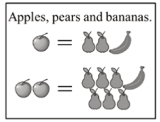
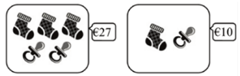
Level of reasoning	Description	Subtypes	Assessment task Episode 2	Assessment task Episode 3	Assessment task Episode 4
R1	Student reasons on the basis of only one of the two given equations	R1_without showing strategy	- [Student assigns values to unknowns on the basis of one provided equation] - “Three stars then equal 9 circles and one rectangle equals 4 circles and then it is correct” [refers to first example equation] [question 2] “Because the green one equals a star and a yellow, and a star equals three yellow” [question 2]	- [Student assigns values to unknowns on the basis of one provided equation] - “Because then you need 9 bananas in the first example” [question 1]	- [Student fills in values for both unknowns in one equation, and checks whether it is correct] - “$1M = 2L$” - “10 + 10 = 20, 5 + 5 + 5 + 5 = 20, 20 and 20 is the same” - “10 + 5 + 5 + 5 = 25”
		R1_with showing strategy	- “Above you see 3 stars and 1 green and 5 yellow. If you take half of that you do not get this” [1 star equals 3 circles] [question 1] “One star = 3 circles. You can discover this by taking away two stars and two circles and a rectangle in the left upper one” [question 1] - [Student takes away one star and three circles in the left example equation] “1 rectangle [in the left example equation] equals 4 circles, then add two circles, makes six circles. Two stars equal six circles, so one star equals 3 circles.” [question 1]	- “You can exchange the p” [pear in equation 1] - “I looked at the first example and I crossed one strawberry and three bananas” [question 1] - “One pear = 1 strawberry and a banana. One strawberry equals 3 bananas. 3 + 1 = 4.” [question 2]	- “$2M = 4L$, dividing by 2, $1M = 2L$” - “if you look at $2M$ and $4L$, then you can take half of $2M$ to get $1M$ $2L$.”

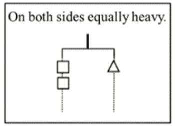
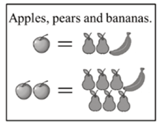
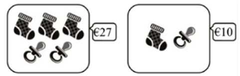
Level of reasoning	Description	Subtypes	Assessment task Episode 2	Assessment task Episode 3	Assessment task Episode 4
R2	Student reasons on the basis of both given equations by combining the information of both of them	R2_without showing strategy	- "Circle = 10, star = 30, rectangle = 40, 30 + 30 + 30 = 90, 40 + 10 + 10 + 10 + 10 + 10 = 90. Rectangle = 40, 30 + 10 = also 40"	- "With values" [assigned the values pear = 4, strawberry = 3, banana = 1, on the basis of both provided equations]	-"10 + 5 + 5 + 5 = 25, 10 + 10 = 5 + 5 + 5 + 5" - "$5L = 25$" - "$L + L + L + L + L = 25$, $5 \times 5 = 25$, $2 \times 5 = 10$"
		R2_with showing strategy	- [Student substitutes rectangle in the left equation by a star and a circle, crosses out a star on both sides, and explains:] "two stars equal 6 circles, so one 1star equals 3 circles" - [Student adds the second equation to the first one]	- "Because you can replace the pear by a banana and a strawberry, then you take away the strawberry and the strawberry, and then you get this [strawberry equals 3 bananas]" [question 1] - "Replace the pear and take away 1s [strawberry] from both sides, then 6:2 = 3, so 6b [banana] = 2s, take half means 3b = 1s." [question 1]	- [substitutes the M in the equation $M + 3L = 25$ by $2L$] - "$2M = 4L$, :2 makes $1M = 2L$, then $5L = 25$" - "$M + 3L = 25$, $M = 2L$, $5L = 25$, $L = 25/5 = 5$, $3L = 15$, $15 + \ldots = 25$, $M = 10$" - "I replaced the M by $2L$, so I got $5L$. $25/5 = 5$, so $L = 5$. Then $3 \times 5 = 15$. $25 - 15 = 10$. So you have 10 left and one M left so $M = 10$"

Appendix 4.1 (Chapter 4)

Coding scheme with examples of student responses for each problem and each level; text in between square brackets is added as a clarification.

Level of reasoning	Descrip-tion	Subtypes	Problem 1 □ + ○ + ○ = 15 □ + ○ = 9	Problem 2 	Problem 3 	Problem 4 
R0	Student does not use one any of the given equations	R0_empty	- [no response]	Idem	Idem	Idem
		R0_don't know	- "I don't understand it" - "I don't know" - "?"	Idem	Idem	Idem
		R0_just know	- "That's what I think" - "I just see it right away"	Idem	Idem	Idem
		R0_repeat given equation(s) or question(s)	-	- "3 squares = 2 triangles" [question a] - "they use the same figures as in the example"	- "that's what is written above" - "because the example shows apple = pear + pear + banana" - "look at the picture"	-
		R0_repeat answer(s)	-	-	- "one apple = 2 pears"	- "sock = 7, pacifier = 3" - "because the sock costs 5 euros and the pacifier costs 2 euros"
		R0_general description	- "just look at the problem for a while" - "I made a guess" - "I just tried something"	Idem	Idem	Idem

Level of reasoning	Descrip-tion	Subtypes	Problem 1	Problem 2	Problem 3	Problem 4
			$\square + \bigcirc + \bigcirc = 15$ $\square + \bigcirc = 9$			
R1	Student reasons on the basis of only one of the two given equations	R1_without showing strategy	- “3 + 6 = 9” - “5 + 5 + 5 = 15”	- “2 squares are 1 triangle, and that’s impossible here” - “one square is missing” [question a]	- ”pear = 1, banana = 2, apple = 4, 4=1+1+2” - “one banana = 2 pears, so one apple equals 2 + 2 = 4 pears”	- “5 + 5 = 10” - “3 socks = 21, 2 pacifiers = 6” - “2 × 9 = 18, 3 × 3 = 9, 18 + 9 = 27”
		R1_with showing strategy	- “take half of 9” - “divide 15 by 3”	- “it is the same as the example, only on both side there is one extra figure so it’s still equal”	- “you can see it in the upper one that one apple is two pears” - “2 apples are six pears so if you take half you know one apple is three pears”	- “take half of 10” - “try different combinations to make 10”

Level of reasoning	Descrip-tion	Subtypes	Problem 1 □ + ○ + ○ = 15 □ + ○ = 9	Problem 2 	Problem 3 	Problem 4 
R2	Student reasons on the basis of both given equations by combining the information of both of them	R2_without showing strategy	- “3 + 6 = 9, 3 + 6 + 6 = 15” - “when you add 3 and 6 it is 9, and 3 plus 6 plus 6 is fifteen” - “if you fill in 3 and 6 in both, it fits”	- “you need one additional square [question a]; 4 squares = 2 triangles and circle = circle” - “1 triangle = 2 squares and the circles are the same”	- ”pear = 1, banana = 2, apple = 4, 4=1+1+2, 8=1+1+1+1+1+1+2” - “if a banana = 2 pears, then one apple = 4 pears and 2 apples = 8 pears”	- “7 + 3 = 10, 7 + 7 + 7 + 3 +3 = 27” - “3 socks = 21, 2 pacifiers = 6, 7 + 3 = 10” - “I tried these values in both examples and this was OK”
		R2_with showing strategy	- “square + circle = 9, than you need 6 more to have 15. So circle must be 6” - “15 - 9 = 6” - “square + circle = 9, so 9 + circle = 15” - “I started with 4 and 5, that did not fit in both calculations, so then I tried 3 and 6. That worked well”	- “1 triangle = 2 squares, so 2 triangles = 4 squares, 1 circle = 1 circle” - “the circles are equal so you can remove them. Then there are 2 triangles and 4 squares left, divided by 2 equals 1 triangle and 2 squares. So that’s correct” - “the two circles are equally heavy and if you double the example you get 4 squares = 2 triangles”	- “if you subtract 2 pears and 1 banana from the second one, four pears remain. So one apple is four pears” - “if you double the apples the other part must also be doubled. So that must be four pears and two bananas. But there are six pears and one banana. So one banana = two pears. So an apple = 2 + 2 = 4 pears” - “two pears and 1 banana can be replaced by one apple, so then the other apple equals 4 pears”	- “27-10-10=7, so pacifier = 7, sock must be 3.” - “take away 10 two times, then 7 remains. 3 x 7 = 21, so sock = 3” - “One sock and one pacifier = 10, so 10 + 10 = 20, so pacifier = 7” - “I first tried 9 & 1, then 8 & 2, then 7 & 3. 7 & 3 worked for both questions”