Supporting children’s science knowledge acquisition, science reasoning, and science self-concept through block play

Dissertation
Zur Erlangung des akademischen Grades eines Doktors der Philosophie
am Fachbereich 5: Erziehungswissenschaften
der Universität Koblenz-Landau

Vorgelegt am 2021-01-04

Von Anke Maria Weber

Erstgutachterin: Prof. Dr. Miriam Leuchter, Universität Koblenz-Landau
Zweitgutachterin: Prof. Dr. Julia Karbach, Universität Koblenz-Landau

Table of contents

Acknowledgements

I would like to thank my supervisor, Prof. Dr. Miriam Leuchter for her ongoing support and assistance during the time of my doctoral studies. She always supported me in the preparation of the three publications featured in the present dissertation and I regard her as being appreciative, open, interested, and constructive. I would also like to express my gratitude to her for giving me the opportunity to focus on my personal research interests. Moreover, I would like to thank Prof. Dr. Julia Karbach for her support during the time of my doctoral studies and her mentoring. She provided guidance whenever needed. Furthermore, I would like to thank Dr. Anna Sckopke for her support, advice, and mentoring. She always took the time to discuss questions concerning my dissertation and supported me in the development of plans for a future in research.

I wish to express my gratitude to my colleagues at the Institute for Children and Youth Education at the University of Koblenz-Landau. I especially thank my highly esteemed colleagues and friends Heide Sasse, Veronika Barkela, Verena Zucker, Sabrina Dämmer, and Timo Reuter for their continued professional and emotional support and for sharing many steps on this journey. Additionally, I also acknowledge my student research assistants Laura Löwe, Kristin Hoffmeister, Rebecca Ott, Hanna Fuhrmann, and Aliena Eichhorn. You did a great job!

Finally, I wish to express my heartfelt thanks to my family, spouse, and friends. I am especially thankful for their faith in me and for their support during this journey. They helped me take my mind off of work and made me laugh when I felt overwhelmed. It is a true gift to have them in my life.

Abstract

Previous research concerned with early science education revealed that guided play can support young children’s knowledge acquisition. However, the questions whether guided play maintains other important prerequisites such as children’s science self-concept and how guided play should be implemented remain unanswered. The present dissertation encompasses three research articles that investigated 5- to 6-year-old children’s science knowledge, science theories, and science self-concept in the stability domain and their relation to interindividual prerequisites. Moreover, the articles examined whether children’s science knowledge, science theories, and science self-concept can be supported by different play forms, i.e., guided play with material and verbal scaffolds, guided play with material scaffolds, and free play. The general introduction of the present dissertation first highlights children’s cognitive development, their science self-concept, and interindividual prerequisites, i.e., fluid and crystallised intelligence, mental rotation ability, and interest in block play. These prerequisites are applied to possible ways of supporting children during play. The first article focused on the measurement of 5-to-6-year-old children’s stability knowledge and its relation to interindividual prerequisites. Results suggested that children’s stability knowledge could be measured reliably and validly, and was related to their fluid and crystallised intelligence. The second article was concerned with the development of children’s intuitive stability theories over three points of measurement and the effects of guided and free play, children’s prior theories as well as their intelligence on these intuitive theories. Results implied that guided play with material and verbal scaffolds supported children’s stability theories more than the other two play forms, i.e., guided play with material scaffolds and free play. Moreover, consistency of children’s prior theories, their fluid and crystallised intelligence were related to children’s theory adaptation after the intervention. The third article focused on the effect of the playful interventions on children’s stability knowledge and science self-concept over three points of measurement. Furthermore, the reciprocal effects between knowledge acquisition and science self-concept were investigated. Results implied that guided play supported knowledge acquisition and maintained children’s science self-concept. Free play did not support children’s stability knowledge and decreased children’s science self-concept. No evidence for reciprocal effects between children’s stability knowledge and their science self-concept was found. Last, in a general discussion, the findings of the three articles are combined and reflected amidst children’s cognitive development. Summarising, the present dissertation shows that children’s science knowledge, science theories, and science self-concept can be supported through guided play that considers children’s cognitive development.

General Introduction

Early science education has received increased attention in recent years as researchers and educators assume that an early acquisition of science knowledge and theories will facilitate advanced science understanding later in life (Eshach & Fried, 2005; OECD, 2014; Tytler & Prain, 2010). However, education needs to always consider children's developmental constraints and it remains unclear how young children's science learning can be adequately supported (Trundle, 2015). In order to find means of fostering children's science knowledge and theories effectively, a first step is to consider and support their cognitive development and their interindividual prerequisites (Tytler & Prain, 2010).

Additionally, research has demonstrated that children's trust in their own academic abilities, i.e., a positive academic self-concept, is a pillar of their development and interacts with their academic achievement (Marsh et al., 2012; Murayama et al., 2013). Their academic self-concept influences how they interact with educational content. If they perceive themselves as able, they will also perceive the educational content as more fun and interact with it more and more often, thus, learning more about the content (Wigfield et al., 2009). Conversely, children, who know more about a content, might perceive themselves as competent in that content and therefore have a higher self-concept (Marsh & Craven, 2006). Even though the support of children's trust in their own science abilities is demanded (OECD, 2014; Pintrich, 2003), science self-concept is often neglected especially for young children (Arens et al., 2016). In order to successfully foster children's science knowledge and theories, their science self-concept should be promoted or at least stabilised, since education that undermines children's trust in their abilities will likely undermine long-term learning as well (Marsh et al., 2012). In order to foster children's science knowledge and theories as well as their science self-concept, education should be developmentally appropriate and allow positive experiences of competence and success (Copple & Bredenkamp, 2009; Marsh et al., 2012).

To meet these demands, education considers children's everyday activities in a domain, about which children already have acquired knowledge or theories and in which they feel competent. One such activity is construction play, e.g., block play (Borriello & Liben, 2018; Rubin et al., 1978), and one such domain is stability (Baillargeon et al., 2014; Bonawitz et al., 2012; Krist, 2010).

Stability is the field of physics that is concerned with forces acting on objects that are in a static equilibrium (Riley & Sturges, 1993). An object that forces act upon can either be at rest or in motion, e.g., a building block remaining stable on a supporting surface or tumbling over, as forces such as weight stabilise or destabilise the block. Very basically, an object will remain stable if its centre of gravity is supported and will tumble if it is unsupported by a surface. If a person was to estimate the stability of different objects, they might differentiate between symmetrical and asymmetrical objects. For a symmetrical object, the consideration of the object's middle is sufficient to correctly estimate its stability, as the object's geometrical centre and its centre of gravity correspond. Therefore, if the geometrical centre is supported, the object will remain in a state of static equilibrium. If the geometrical centre is not supported, the object will tumble. For asymmetrical objects however, the centre of gravity must be considered to estimate the object's stability correctly, because the centre of gravity and the geometrical centre do not correspond. Therefore, if the centre of gravity of an asymmetrical object is supported but the geometrical centre is not, the object will remain in a state of static equilibrium. If the centre of gravity is not supported but the geometrical centre is, the object will tumble. Studies have found that children between 5- and 6-years of age face problems with estimating asymmetrical objects' stabilities, because they overgeneralise the geometrical centre and ignore the importance of the centre of gravity indicated by the object's mass (Bonawitz et al., 2012; Krist, 2010). Thus, the stability domain might be suitable to investigate ways to support children's science learning during block play.

Concerning this support, both science learning as well as children's science self-concept might be fostered through an adult's scaffolding (Belland et al., 2013; van de Pol et al., 2010). Thus, (block) play and scaffolding might promote children's science knowledge and theories and their trust in their science abilities, while considering children's developmental constraints.

Following these demands, three studies were conducted to investigate (1) whether children's stability knowledge can be assessed and how it relates to possible interindividual prerequisites such as science motivational self-concept, science competence self-concept, intelligence, mental rotation ability, and interest in block play. (2) The second study was concerned with the support of children's stability theories through guided and free block play, and whether intelligence contributes to theory acquisition. (3) The third study focused on the relation between children's stability knowledge and their science self-concept, and if both can be supported through guided and free block play.

Concluding, effective science education should consider children's cognitive development and their science self-concept in a developmentally appropriate setting such as play (Borriello & Liben, 2018; Pintrich, 2003). An adult's scaffolds might support children's science learning as well as their science self-concept (Belland et al., 2013; van de Pol et al., 2010). This dissertation aims to contribute to the research on how to best support children's science learning and their science self-concept, while still considering their developmental constraints.

1 Interindividual differences and prerequisites

1.1 Developmental perspective on reasoning in early childhood

In order to investigate children’s knowledge and theory acquisition and how to support them in a developmentally appropriate way, children’s cognitive development needs to be considered (Trundle, 2015). Children’s ability to reason is at the heart of their cognitive development and is discussed in light of two different perspectives concerning general and specific reasoning processes (Chapter 1.1.1). Moreover, deduction, induction, and abduction are reasoning processes that young children apply to a certain extent (Chapter 1.1.2). Both general and specific reasoning processes as well as deduction, induction, and abduction interact with the cognitive processes embedded in children’s reasoning, such as generalisation, categorisation, the formation of concepts, and causality (Chapters 1.1.3 to 1.1.6; Kuhn, 2013). These processes are combined in theory theory conjoined with Bayesian inference and findings on children’s ability to coordinate theory and evidence (Chapter 1.1.7; Gopnik & Wellman, 2012; Koerber et al., 2005). These allow suggestions how to support children’s knowledge acquisition, e.g., their intuitive knowledge about stability, and their theory adaptation, e.g., their reasoning about stabilities that they can state explicitly. The literature on young children’s knowledge and theories about stability is presented in Chapter 1.1.8.

Reasoning is defined as consciously executed and purposeful thinking that has a goal, be it understanding, deciding, or acting (Kuhn, 2013). Reasoning might be triggered by at least three different questions targeting mental processes (how did you know?), the formal structure of a problem (how could you know?), and a justification (how do you know? Kuhn, 2013).

From a developmental psychological perspective, young children already possess the ability to reason to a certain extent, but their ability is still far from completely developed (Bjorklund, 2013; Kuhn, 2013). Moreover, while reasoning develops, there is no universal developmental trajectory that all children follow, because experiences play a crucial role in the development of reasoning. Even if children acquire a more sophisticated reasoning strategy, they may not transfer this strategy to other problems, but may still use less advanced strategies simultaneously. Their use of those strategies, however, often decreases, while their use of the more sophisticated strategies increases over time (Kuhn, 2013). The use of advanced strategies can be supported, e.g., by explaining their wrong ideas (for an overview, see Lombrozo, 2006), by choosing and conducting experiments (Kuhn & Ho, 1980), and by explanations (Sobel & Sommerville, 2009).

The development of children's reasoning is discussed in light of two different developmental psychological perspectives that have different assumptions about how children's cognitive development takes place. Both perspectives contribute to the explanation of children's cognitive development and provide indications on how this development might be promoted.

1.1.1 General and specific aspects of reasoning

Concerning the development of children's reasoning, two different perspectives of when and how children's reasoning develops have been suggested and both have implications for possible ways to support children (Rakison & Lawson, 2013). The first perspective is concerned with generalised learning mechanisms, often called domain-general processes, such as children's learning through statistical regularities (Halford & Andrews, 2006; Rakison & Lawson, 2013). This perspective postulates that there are certain learning mechanisms such as covariation that apply to all domains, e.g., reasoning about math, language, science, music, etc. The second perspective is concerned with specialised learning mechanisms, often called domain-specific processes, implying that reasoning is dependent on knowledge that guides expectations (Demetriou et al., 2002; Rakison & Lawson, 2013). This perspective postulates that specific knowledge or theories such as children's prior knowledge or theories about stability influence reasoning processes in a certain domain.

Concerning the general perspective, researchers claim that children's reasoning develops through general learning mechanisms such as habituation, association, or conditioning (Rakison & Lawson, 2013), and that children learn through perceptual similarities (Sloutsky & Fisher, 2004). The child is regarded as a *data analyst*, who learns by considering statistical probabilities and combining them with general learning mechanisms (Gelman, 2013). Moreover, researchers hypothesise that similar learning mechanisms are used at every stage of life and prior knowledge is integrated into the learning process (Rakison & Lawson, 2013). Knowledge is viewed as an accumulation of concrete patterns of covariation (Gopnik, 2013). Through these general mechanisms, children may detect regularities such as covariation patterns quickly, however, these mechanisms would lead to a rather slow development of reasoning (Rakison & Lawson, 2013). For example, if children often observe that blocks whose mass is supported by a supporting surface remain stable, children might infer that the mass is a deciding factor for stability by observing covariation. Moreover, this view is substantiated by findings that preschool children can transfer their reasoning skills to other contents, situations, and new materials (Brown & Kane, 1988).

Three arguments underline the general perspective. (1) Rakison and Lupyan (2008) suggested that statistical regularities facilitate learning, because children observe these regularities and reason accordingly. For example, children may observe that blocks which are supported at their mass, will remain stable and, thus, reason that the mass must be important. (2) General learning mechanisms are parsimonious, as they suggest that simple perceptual correlations are sufficient for the development of reasoning (Quinn & Eimas, 2000; Sloutsky & Fisher, 2004). If a child observes that symmetrical objects whose centre is unsupported, tumble over, the child might infer that the centre is of importance for stability. (3) They can explain and predict how learning and reasoning change with age (Rakison & Lawson, 2013). For example, young children are more likely to learn from perceptual correlations, however, with increasing age they include more complex regularities such as the mass into their reasoning (Namy & Gentner, 2002).

Regarding the specific perspective on children's reasoning, the child is regarded as a *theorist*, who has specific intuitive theories about the world (Gelman, 2013; Gopnik & Wellman, 2012). For example, an intuitive theory might be concerned with the question why objects remain stable or tumble. Specific processes would allow children to include non-obvious information into their reasoning (Rakison & Lawson, 2013), e.g., a child acts as an agent when they knock over a block tower. The possible inclusion of non-perceptual information is contrary to the assumption of the general perspective that children develop their reasoning mainly from perceptual covariation. Moreover, knowledge is viewed as abstract, structured and hierarchical (Gopnik, 2013). Evidence in support of this view has revealed that people solve reasoning problems more easily if they are presented in an everyday context, because reasoning is highly content and context specific (Evans, 2013; Harris & Nuntez, 1996).

Three arguments support the specific perspective in cognitive development: (1) The input children are confronted with is very complex, so that specific learning mechanisms are needed (Gelman, 1990; Mandler, 1992). For example, children are not only confronted with the mass when they reason about the blocks' stabilities, but with many different features. These features include shape, size, colour, material, centre, etc. It is likely that children have prior theories about which of the features might determine stability and which might not, thus, facilitating their reasoning by limiting the possibilities of the problem space. (2) The sole perceptual association of two cues is insufficient for the development of complex schemata and concepts (Mandler, 2003), because if two features co-occur it does not necessarily mean that they are related. For example, a child might observe that blocks of a certain colour remain stable. If the child has a prior theory about a factor contributing to stability, e.g., the centre or the mass as important factors, they might infer that the colour cannot determine stability. (3) A specific process, in contrast to general processes, allows the identification of the relevant features from the many possible features (Rakison & Lawson, 2013), and thus solves the problem concerned with the complexity of the input. For example, a child might have an intuitive theory that a block's shape (symmetrical or asymmetrical) might have an effect on its stability, but that its colour is unimportant.

Kuhn (2013) has tried to reconcile both views, claiming that both processes interact, as reasoning always takes place in a context, i.e., is specific, but transfer is also possible, i.e., reasoning is general. Moreover, some researchers claim that children use both general and specific processes flexibly in their cognitive development, and that both processes interact and support each other (Hayes & Thompson, 2007; Opfer & Bulloch, 2007). Both the general and specific perspective contribute to the explanation of the development of young children's reasoning, and thus have implications for the implementation of interventions that foster children's knowledge and theory acquisition. General as well as specific processes may facilitate children's knowledge and theory acquisition, because children draw inferences from covariation that they evaluate in light of their prior knowledge or theories.

1.1.2 Deduction, induction, and abduction

Reasoning comprises three different underlying processes, deduction, induction, and abduction, that are fundamental for young children's cognitive development. All three forms of reasoning are involved in the acquisition of science knowledge (Leuchter & Hardy, 2020). Deduction refers to top-down processes, induction to bottom-up processes, and abduction to the inference of rules and regularities. Deduction, induction, and abduction cannot be separated or hierarchically ordered, because the three processes are intertwined (Lawson, 2010). Goswami (2014) suggests that age related differences in deduction and induction stem partly from differences in knowledge between children and adults. Moreover, Kuhn (2013) underlines this view by claiming that development in reasoning and in knowledge go hand in hand.

Deduction. Deduction refers to top-down processes in reasoning and is therefore concerned with inferring features or characteristics from a theory or concept to singular cases. Deduction can take the form of conditional reasoning (e.g., *If the mass is supported, the block will remain stable*.), and transitive reasoning (e.g., *All blocks whose mass is supported remain stable. The block's mass is supported. Thus, the block remains stable*). Moreover, deduction allows logical conclusions with only one solution and is possible without having real world knowledge, making even counterfactual deductions (*All yellow blocks tumble. This block is yellow. Therefore, it tumbles.*) logically valid (Goswami, 2014). According to Markovits and Thompson

(2008), 6-year-olds are able to use deduction to a certain extent. In their study, children were confronted with conditional inferences (*if A, then B*), followed by two statements concerned with this condition (one correct statement: *You see A, is B certain?* and one incorrect statement: *You see B, is A certain?*). Most children agreed to the correct statement and rejected the incorrect statement. In another study, Chantal and Markovits (2017) found that even 2- to 5-year-old children use deduction if they have been prompted to think about alternative premises. Children were either prompted to think about ideas to solve three different problems such as generating ideas for presents or received an inhibition task. Afterwards, they received conditional inferences similar to those applied in the study by Markovits and Thompson (2008). Children who were prompted to think about alternatives were more likely to apply correct deductive inferences. Concerning children’s reasoning about stabilities, children could be prompted to think about what might happen if blocks on a supporting surface had been placed on the same supporting surface differently. Would the mass still be supported and the blocks remain stable or would they tumble?

Induction. Induction refers to bottom-up processes in reasoning and is therefore concerned with inferring theories and categories from singular cases and rules. Induction allows many possible solutions and is strongly influenced by prior knowledge or beliefs (Johnson-Laird & Khemlani, 2017), and does not necessarily lead to correct conclusions (Leuchter & Hardy, 2020). According to Hayes and Heit (2013), inductive inference may be based on (1) *premise similarity*, (2) *premise typicality*, (3) *premise diversity*, (4) *premise monotonicity*, and (5) *causal relations*. (1) *Premise similarity* refers to the generalisation of a novel property from two or more similar cases to the category as a whole. For example, if a child plays with blocks and observes that two asymmetrical blocks tumble, they might infer that all asymmetrical blocks tumble. (2) *Premise typicality* implies that inductions are more likely to be made from a typical example of a category than from an atypical example. Concerning stability of blocks, children might view symmetrical rectangular blocks as a typical block and thus might infer that if symmetrical blocks tumble all blocks tumble. In a study on premise typicality, Gelman and Coley (1990) told children that a robin was a bird and lived in a nest. Afterwards, children received pictures of birds that were either typical (e.g., a chickadee) or atypical (e.g., an ostrich). The researchers found that 2-year-old children use typicality for induction, because children were more likely to conclude that the typical birds lived in a nest than the atypical birds. (3) *Premise diversity* suggests that inductions are made from two different examples (Feeney & Heit, 2011). For example, if an asymmetrical and a symmetrical block tumble, all blocks are inferred to tumble. However, Lawson and Fisher (2011) found that children under the age of 8 fail to realise the value of diversity for induction. In a series of studies, children were confronted with evidence about small and large samples of mammals that differed in diversity, i.e., homogeneous versus diverse. Afterwards, children were asked to make inferences about mammals, vertebrates, and invertebrates. The children drew more inferences from large and homogenous samples than from diverse ones. Consequently, in order to support children's knowledge and theory acquisition, examples should share similarities to help young children infer a rule or a category. (4) *Premise monotonicity* describes an induction based on many different examples. For example, if many blocks of different shapes and colours tumble, it might be inferred that all blocks no matter their shape and colour tumble. (5) *Causal relations* play an important role in induction, because the underlying reasons are considered. For example, children might infer that the blocks' masses are unsupported and therefore the mass is the deciding feature. Hayes and Thompson (2007) and Opfer and Bulloch (2007) discovered that 5-year-old children use causal relations for induction. Hayes and Thompson (2007) confronted children with causal information about fantasy animals such as an animal having large eyes so it can see at night and non-causal information such as an animal having white wings. Afterwards, children were asked, which characteristics a target animal of the same category was likely to have. They found that children were more likely to make inferences about fantasy animals from causally related than causally unrelated features. However, young children were more likely to pay attention to irrelevant surface features than 8-year-old children or adults. Opfer and Bulloch (2007) came to similar results. If children had causal information about a species, they were more likely to use this information than perceptual similarity. However, if they had no causal information, they relied on perceptual similarities. Moreover, Goddu et al. (2020) found that causal information plays an important role in analogical reasoning for children as young as 3 to 4 years of age. In their study, children again relied on perceptual similarities unless they had received causal information that helped explain analogies.

Furthermore, inductive reasoning encompasses analogical reasoning (Goswami, 2014). Chen et al. (1997) found that 13-month-old infants reason analogically, as they were able to generalise the solution of a problem to other problems. Infants learned to remove a barrier, bring a string into reach with the help of a cloth, and pull the string to reach a toy. Afterwards, infants were confronted with problems involving different barriers, cloths, strings, and toys. Infants recognised the structural similarity between the problems and could therefore apply their means-to-end reasoning to the other scenarios through analogical reasoning. The authors also discovered that even at 10 months of age, infants could use analogical reasoning if at least one feature of the situation remained the same, e.g., if the barrier did not change. The researchers conclude that the younger infants did not have the same amount of prior knowledge as the 13-month-olds.

Therefore, the younger infants considered surface features over relational similarity. Namy and Gentner (2002) found that prior knowledge plays an important role for preschool children as well. If children know little about a concept or an entity, they will mainly consider surface features. Children between the ages of 4 and 7 already have prior knowledge or theories about stability (Bonawitz et al., 2012; Krist, 2010). Therefore, they might not only consider surface features such as colour or size, but also more complex characteristics such as the mass.

Abduction. Abduction implies the inference of rules and regularities. In contrast to induction, abduction introduces new principles and refers to the process of creating explanations (Johnson-Laird & Khemlani, 2017). Abduction encompasses the resolution of causal inconsistencies through explanations, which depend on understanding and prior knowledge (Johnson-Laird & Khemlani, 2017; Lombrozo & Vasilyeva, 2017). Thus, if children observe an event that violates their causal understanding, they will search for alternative explanations. For example, if a child assumes that blocks which are supported at their geometrical centre will remain stable, but observes that they tumble, they will note an inconsistency and might look for alternative explanations.

Furthermore, abduction encompasses (1) Bayesian priors (Griffiths, 2017), (2) the coordination of theory and evidence, (3) hypothesising, and (4) variable control (Leuchter & Hardy, 2020). (1) Bayesian priors indicate an initial belief and its probability and might be used for abductive reasoning (Bonawitz et al., 2012; Gopnik, 2013). (2) To use abductive inference, children have to understand the value of evidence for theories. If a child has not acquired this understanding, they will not search for alternative explanations when they observe evidence that contradicts their prior beliefs. The reason for this is that children will not view the evidence as undermining their theory. For example, if children assume that supporting the centre is sufficient for blocks to remain stable, they might ignore evidence indicating that the mass is important. Preschool children are able to use observed evidence to predict the outcome of a new experiment, suggesting that they infer regularities from evidence (Schulz, Gopnik, & Glymour, 2007). However, even adults face problems to successfully coordinate theory with evidence (Sodian, 2018). The role of Bayesian inference and theory-evidence coordination for children’s knowledge acquisition and their intuitive theories is discussed in Chapter 1.1.7. (3) Abduction can take the form of hypothesising. A child might have the hypothesis that the mass of a block is important for it to remain stable. Children between the ages of 4 and 6 face problems with the formation of hypotheses (Piekny & Maehler, 2013). In this study, the children were asked to generate hypotheses about fantasy animals. Preschool children mostly failed to generate hypotheses. However, Ruffman et al. (1993) showed that if children are confronted with perfect covariation and are simply required to form a single hypothesis, they succeed in hypothesis formation. (4) The ability to control variables is a prerequisite to test for the validity of one's abductive inference. A child examining if the centre is sufficient for blocks to remain stable, might place blocks of different shapes and sizes on a supporting surface with their centre supported. The control-of-variables strategy refers to the ability to test hypotheses with unconfounded experiments (Klahr & Li, 2005). Children between the ages of 4 and 6 are able to control variables to conduct unconfounded experiments if they are taught to do so (van der Graaf et al., 2015). Moreover, Kushnir and Gopnik (2005) found that 4- to 6-year-old children realised if their own interventions were confounded by the interventions of an experimenter and did not draw conclusions from these confounded interventions. This result suggests that children understand, at least in part, that confounded experiments do not offer reliable conclusions.

The findings on children's ability to use deductive, inductive, and abductive inference suggest that preschool children can use all three types of inference if they are confronted with tasks that are developmentally appropriate.

The three reasoning processes, deduction, induction, and abduction relate to other cognitive processes such as generalisation, categorisation, formation of concepts, and causality that support the formation of theories about science principles, while general and specific reasoning processes lay the foundation for these cognitive processes (Gelman, 2013; Gopnik, 2013; Kuhn, 2013; Zimmerman, 2007). All of the mentioned processes are relevant for children's learning and studies on children's capabilities concerning these processes may provide further indications of how children's development can be fostered. Therefore, these cognitive processes are discussed in the following sections in the light of the three reasoning processes as well as the two perspectives.

1.1.3 Generalisation

Generalisation is a part of deductive reasoning, because deduction comprises the inference from a generalised rule, feature, or category to a specific case (Grégoire, 2001). If a child has concluded that the mass needs to be supported, they can deduce that this rule will apply to a specific block. Generalisation can be a form of inductive reasoning (Goswami, 2014). The inductive process of generalising underlying features can take place in three forms (Rehder, 2017b). (1) Object to category, e.g., all asymmetrical blocks remain stable if their mass is supported. (2) Category to category, e.g., all asymmetrical and symmetrical blocks remain stable if their mass is supported. (3) Object to object, e.g., block 1 and block 2 both remain stable, because their mass is supported. These three processes are more likely when the premises are similar, typical, diverse, monotone, or have causal relations (Hayes & Heit, 2013; Rehder, 2017b). Furthermore, generalisation in abductive reasoning takes the form of generalising rules and regularities to form categories or concepts. For example, a child might generalise the rule that the mass needs to be supported in order for blocks to remain stable.

Generalisation can be investigated in light of the general and specific perspectives. Studies have examined whether children apply general (perceptual features) or specific processes (prior knowledge). Gelman and Markman (1986) found that 4-year-old children understand which features are generalisable and which are not, and only generalise the relevant features. In their study, children were asked to generalise either by category membership, e.g., is a shark likely to breathe like a tropical fish (category) or like a dolphin (perceptual similarity). Children mostly generalised characteristics by category membership, not perceptual similarity. In another study, Gelman and Markman (1987) came to similar results with 3- to 4-year-old children. Concerning stability, young children may understand that the colour is not a generalisable feature concerning blocks' stabilities, because different blocks have different colours. Furthermore, they might understand that the blocks' centres, or their masses are generalisable features and use these to form categories, i.e., induction, or explanations, i.e., abduction. Additionally, Hayes and Thompson (2007) discovered that 5- to 8-year-old children rather generalise causal relations opposed to perceptual similarities when it comes to fantasy animals, indicating that children understand causality. Children had no prior knowledge or theories about the fantasy animals in their study and still they did not rely on perceptual similarities, but included their knowledge that causality is generalisable, while perceptual features are not necessarily. All three studies underline the specific view that knowledge guides learning. However, in all of the studies, some children also relied on perceptual similarities, as suggested by the general perspective. Seemingly, children are more likely to rely on their prior theories or knowledge, but might still consider surface similarities.

The above-mentioned studies have uncovered that children mainly use causal information to generalise properties, but some children rely on perceptual similarities as well. Generalisation is a prerequisite for categorisation, because a child has to generalise certain characteristics of entities in order to form categories.

1.1.4 Categorisation

According to Rakison and Lawson (2013), categorisation can be defined as the grouping of entities that are alike or at least similar and the cognitive process of finding a superordinate characteristic all group members share. Categorisation also includes deductive reasoning, since children generalise properties to new entities (Rehder, 2017a). For example, a child might generalise the rule that the mass needs to be supported to blocks of different shapes and sizes. Moreover, necessary processes for categorisation are differentiation and induction to scrutinise whether a certain feature or characteristic can be generalised across a number of entities to form a category. Categories are central for the use of past experiences to understand new situations and are essential for classification, inference, and communication (Markman & Rein, 2013). Categorisation can encompass induction, as children form categories or categorise new entities into existing categories. Children might categorise symmetrical and asymmetrical block as blocks. Categorisation can involve abductive processes as well if underlying rules, regularities, or causalities are introduced that determine category membership or that category members share. A child might determine that all blocks share the characteristic that they are stable if their mass is supported.

According to Markman and Rein (2013), there are five different kinds of categories. (1) *Feature-based categories* have characteristic features in common, e.g., blocks are made out of wood. (2) *Goal-derived categories* share a goal, e.g., a child might have the goal to build a house including doors, windows, a garage, different rooms, and furniture. (3) *Relational categories* are organised around a relationship between items, e.g., a counterweight stabilising a construction. (4) *Role-governed categories* are defined by the role in a relational structure, e.g., the importance of the support of the mass for objects to remain stable. (5) *Thematic categories* refer to items that co-occur, e.g., in block play co-occurring items or entities might be blocks of different shapes and sizes, the ground that is built on, etc.

Children’s categorisation is discussed in the light of general and specific processes. Studies in support of general processes found that children between 5- and 6-years of age categorise fantasy animals mainly based on perceptual similarities such as labels instead of their prior knowledge (Sloutsky & Fisher, 2004). However, Hayes et al. (2008) could not replicate this result. Moreover, Rakison and Hahn (2004) found that 5-year-old children use both perceptual as well as nonobvious properties to determine category membership, because they have learned that nonobvious entities are often related to category membership, while perceptual features are not. Indeed, their results suggest that children use statistical information such as distribution to categorise, underlining the view of the child as a data analysist.

The specific view postulates that children have intuitive theories about categories in general and in specific, which can be adjusted. Children have theories about what features define a category and understand that these features are not necessarily perceptual but might be nonobvious (Gelman & Markman, 1986; Rakison & Hahn, 2004; Rakison & Lawson, 2013). Studies in support of specific processes found that 4-year-old children base their inductions mainly on category membership instead of perceptual similarity (Gelman & Markman, 1986, 1987), indicating that they have a theory about the importance of underlying similarities over perceptual similarities. Gelman and Coley (1990) have reported similar results with 2-year-olds. Simons and Keil (1995) found that 3-year-olds distinguish between living and non-living things and appeal to biological features when sorting animals into categories. Moreover, Gentner and Namy (1999) found that 4-year-old children, who learned a novel label for multiple members of a category, e.g., all members were referred to as blickets, were more likely to choose the category match (e.g., apple and banana) than the perceptual match (e.g., apple and balloon). However, if children only learned a new label for a single instance, e.g., only a single object was referred to as a blicket, they are equally likely to select category and perceptual matches. This study highlights the role of comparisons for categorisation. These findings indicate that children rely on theories for categorisation.

Opfer and Bulloch (2007) found evidence for both general and specific reasoning processes. Children in their study used perceptual similarities, i.e., a general process, to categorise a new biological species if they had received no causal information. However, if children had received causal information about the target origin, they used that prior knowledge for categorisation, i.e., a specific process. This finding highlights the role of causality for categorisation and implies that children use both general and specific processes depending on external circumstances such as provided information.

From an early age on, categorisation is influenced by causality. Causality may affect categorisation in three ways through (1) *coherence*, (2) *explaining away evidence*, and (3) *essentialism* (Rehder, 2017a).

(1) *Coherence* indicates that entities are more likely to be categorised into new or existing categories if an underlying causal relation between the new entity and other entities or the category as a whole are detected (Gelman, 2013; Rehder, 2017b). For example, if the mass is viewed as the deciding factor for an object’s stability, an object whose mass is supported is more likely to be categorised as stable. Gopnik and Sobel (2000) discovered that 2- to 4-year-old children use both causalities as well as perceptual features for categorisation. They conducted a study in the blicket detector paradigm to exclude effects of possible prior knowledge. Children were asked to categorise objects that had causal relations or perceptual similarities in common as either blickets or not blickets. Even the 2-year-olds quickly learned that causality was a more reliable predictor of category membership than perceptual similarity. By the age of 3 to 5, children are more likely to form categories on the basis of causality than perceptual features if they are

prompted to think about underlying causal relations (Walker et al., 2014). Hayes and Rehder (2012) found that 5- to 6-year-olds use coherence for categorisation. In their study, children were more likely to categorise instances into a certain category if the underlying causal relation was present.

(2) *Explaining away evidence* implies that entities are less likely to be categorised if other plausible causes exist (Kuhn, 2014; Simons, 2000). For example, if an object's middle is a sufficient explanation for its stability, the mass might not be considered.

(3) *Essentialism* refers to causal models embedded within categories that are deemed essential for the category as a whole. For example, the mass always needs to be supported for an object to remain stable. Young children can also reason causally from evidence that they observe to unobservable features that are essential for determining category membership (Gopnik & Sobel, 2000; Sobel & Buchanan, 2009; Sobel & Kirkham, 2007). In a series of studies with the blicket detector, Sobel et al. (2004) and Sobel and Kirkham (2006) found that 2- and 4-year-old children can infer which of two blickets activates the detector. Children observed that blicket A and B together activated the detector, afterwards children either observed that block A alone activates or does not activate the detector. If A activated the machine, children inferred that A is a blicket. If A did not activate the machine, children inferred that B is a blicket. Therefore, they determined the essential causal feature of the category blicket. However, Rakison and Hahn (2004) found no evidence for essentialism in 4-year-old children, as they did not weigh nonobvious features more important for categorisation than perceptual features. Yet, children in their study did not have any reason to assume an underlying causal relationship between category membership and the nonobvious feature, e.g., mammals or insects might sleep in trees. Beyond that, in certain instances the nonobvious properties even contradicted children's prior knowledge about the category, e.g., mammals lay eggs. The missing evidence for essentialism might thus stem from the missing causal relation. Concerning stability, children might have already acquired prior theories such as that supporting an object's mass is sufficient for the object to remain stable. Therefore, they might view support of the mass as essential for stability.

Categorisation is intertwined with concept formation and concepts may include mental representations of categories. In the context of stability, children might have a stability concept and therefore the development of concepts might contribute to the understanding of children's reasoning about stability.

1.1.5 Concepts

Concepts are mental representations of the properties, features, and structures of categories and their members, and are thus considered the basis of human cognition. Concepts store knowledge and support generalisation and the formation of new categories. Moreover, concepts are parsimonious, as they decrease memory capacity (Rakison & Lawson, 2013). Examples for concepts are (1) information about single entities (Medin & Schaffer, 1978), e.g., characteristics of specific symmetrical or asymmetrical block shapes; (2) an amalgamation of relevant features, i.e., prototypes (Medin & Schaffer, 1978), e.g., a prototypical block might be made of wood, has a symmetrical rectangular shape, and remains stable if its centre is supported; (3) a list of necessary and defining characteristics (Bruner, 1964; Bruner et al., 1966; Rakison & Lawson, 2013), e.g., the theory that the mass needs to be supported for asymmetrical objects to remain stable. Similar to categorisation, concepts can be used for deductive reasoning to generalise features to other, maybe new entities (see Rehder, 2017a). Concepts can be formed on the basis of induction and categorisation, as children use bottom-up processes to form concepts, i.e., a form of accommodation (Piaget, 1950). Moreover, concepts allow for induction, because new evidence or entities can be included into already existing concepts (Gelman, 2013; Rakison & Lawson, 2013), the process Piaget (1950) called assimilation. Abductive processes may be used for the formation of concepts, as children hypothesise about relations and test these hypotheses (Gelman, 2013).

Love (2017) summarises five possible ways, in which humans form concepts, (1) *rule-based*, (2) *protype-based*, (3) *exemplar-based*, (4) *hybrid models*, and (5) *multiple systems models*. (1) The *rule-based* approach suggests that concepts are defined through logical rules. Therefore, a person must test whether these rules are true, which implies an abductive process. However, if concepts were solely rule-based, useful information would be discarded to keep the rule as parsimonious as possible. Furthermore, concepts are more complex than just simple rules. For example, a rule-based approach to stability might suggest that blocks always tumble when their centre is unsupported. However, blocks may tumble for a number of reasons. Their centre or their mass might not have been adequately supported or a child might have knocked them over, etc. (2) Concepts may be *prototype-based*. For example, a prototypical block might be made of wood, is rectangular, and symmetrical. This suggests that humans recognise all properties and form a prototype that is a summary of the concept, but do not have any information about frequencies. Evidence for a typicality based concept formation has been found for young children at 6 years of age (Rhodes et al., 2008). In this study, children and adults were presented with photographs of animals varying in concept typicality and diversity. Participants received two sets of photographs, one with two diverse examples, e.g., golden retriever and hairless Chinese crested, and a non-diverse example, e.g., golden retriever and Labrador for the category dog. The participants were asked from which pair of photographs, the diverse or the non-diverse, they could learn most about the concept. Children were likely to rely on typicality and ignore the role of diversity, e.g., they chose the non-diverse pair, while adults were likely to investigate typicality in diverse samples. Concerning stability, children might rely on typical blocks, e.g., symmetrical and rectangular, when asked to rate stabilities of different blocks and incorrectly judge that supporting the centre is sufficient for all blocks. They may ignore the role of diversity and thus blocks with different shapes such as asymmetrical blocks. (3) The *exemplar-based* approach implies that not only a prototype, but every single example of a concept is stored. For example, for the stability concept, every stable and tumbling block would be stored. This viewpoint recognises the importance of frequency, variability, and correlations, and allows for the integration of new examples into the concept by looking at the similarity between a new example and all existing ones. This approach's downside is the high required memory load making an exemplar-based concept too complex to be useful. (4) *Hybrid models* infer that humans do not solely store prototypes or all examples, but selectively choose important features that define the concept. For example, a child might determine that the support of the mass is the defining feature for stability. (5) *Multiple systems models* indicate that humans do not have a singular concept but multiple concepts that may exist at the same time. For example, children may have acquired the theory that the mass is a deciding factor for stability, while still maintaining the theory that the geometrical centre might be important. Over time, they may find that the mass theory is more successful in predicting stability and use this theory more and more frequently.

Children's concepts are discussed in the light of general and specific processes as well. Gelman (2013) claims that the two processes are not separate, but complement each other in the formation of concepts. She suggests that concepts (1) are formed early in life, but develop throughout childhood. Accordingly, in a study, Legare and Gelman (2008) found that new concepts develop, while old concepts often exist at the same time. They discovered that even into adulthood scientifically correct and supernatural concepts can co-exist (see multiple systems models). For example, children might acquire knowledge that the mass has to be supported for an object to remain stable, but may at the same time still regard the centre as important. (2) Concepts underlie general principles, but are embedded in specific knowledge. A concept's structure is mostly informed by general principles such as perceived visual similarity (Bhatt & Quinn, 2011; Quinn et al., 2008), typicality (Rhodes et al., 2008), and analogies (Namy & Gentner, 2002). However, a concept's content is influenced by specific principles, which highlight children's theories about specific contents such as psychology and biology (Schulz, Bonawitz, & Griffiths, 2007). For example, a child might learn about stability by observing co-variation that is interpreted in the light of their prior theory. (3) Children actively construct concepts, but are influenced by experiences and input. Children may at times refuse to adopt correct concepts by ignoring the evidence (Kuhn, 2014; Simons, 2000). However, children's learning is also highly influenced by their surroundings. Children's stability knowledge might be influenced by their time spent with building blocks, but they may ignore the evidence they observe during building. (4) Concepts are flexible and thus open to new knowledge, but also constrained to certain principles, e.g., blocks may come in different shapes, sizes, and colours, but they are also constrained to certain materials such as wood or plastic. (5) Concepts are informed by statistical frequencies as well as theories. This highlights the different views of the child as a data analyst or a theorist. Young children and even infants are very sensitive to statistical patterns and draw conclusions from them (Sloutsky & Fisher, 2004). However, children also form concepts about non-observable states and principles such as mental states and incorporate causality into their concepts (Carey, 2009; Gopnik & Meltzoff, 1997; Rehder, 2017a). It is possible that children start to form concepts from observable content and then use theory-based abductive processes such as hypothesis testing to enrich them. Therefore, children might combine general learning mechanisms such as conditioning and association and specific mechanisms such as the use of conceptual and causal information. Gelman (2013) explains that surface features often correlate with causal relations, and statistical regularities are indicative of theories. Therefore, children might search for complex features such as the mass when determining stability in line with observed evidence and their prior theories.

Many of the studies on generalisation, categorisation and concept formation have demonstrated that causality is important for each of the three processes (e.g., Hayes & Thompson, 2007; Schulz, Bonawitz, & Griffiths, 2007; Sobel & Kirkham, 2006). Children are more likely to generalise, categorise and form concepts correctly if they are aware of underlying causal relations. Therefore, causality seems to be a crucial factor in children's cognitive development.

1.1.6 Causality

According to Gopnik (2013) causality is at the heart of children's cognitive development and most concepts and intuitive theories are comprised of causal generalisations. She further hypothesises that causal relations are at the core of concepts and theories and few concepts, such as numerical and spatial concepts, are not of causal nature. Rehder (2017a) explains that not all features of a concept or a theory are a part of the causal model. Features belonging to the causal model, however, determine a concept, a category or a theory, whereas other features provide indirect evidence. The support of the mass is a determining feature of the stability concept. Deduction is often concerned with causal relations (if A, then B; if A changes, then B changes). A child might assume that if a block is supported at its mass (A), then it will remain stable (B). If a block is pushed along the supporting surface so its mass is unsupported (A changes), then it will tumble (B changes). Moreover, there are a number of studies concerned with inductive inference in causality (Bright & Feeney, 2014; Gopnik & Sobel, 2000; Griffiths et al., 2011; Schulz, Bonawitz, & Griffiths, 2007). The probability of a consequence B leads children to search for a possible cause A. If a block tumbles, children might search for the cause A. After a possible cause A was determined, the same process can be abductive if prior probabilities for the hypothesised cause A are included or a person conducts experiments to ascertain if A really causes B. The search for causal relations is influenced by prior knowledge or theories, as they influence which probable causes a person even considers (Gopnik, 2013; Griffiths, 2017). For example, a child might determine that the block's colour is not important, but assume that their centre or their mass might be.

According to Waldmann (2017), theories on causal reasoning can be embedded into three larger frameworks. (1) The *process framework* comprises theories referring to continuous causal and mainly physical processes such as tumbling blocks. (2) The *disposition framework* involves the interaction of two causal agents, such as the interaction of a person pushing over a block tower and the blocks tumbling. (3) The *dependency framework* is concerned with relations such as if A, then B, e.g., if the mass is supported, the blocks will remain stable. This framework explains how people learn and infer from statistical patterns.

In developmental psychology, three possible origins of causal reasoning in infancy are discussed, (1) *innate representations of motor or force events*, (2) *agents*, and (3) *covariation* (Muentener & Bonawitz, 2017). (1) Causal reasoning in infants is presumed to develop through the *innate or early representation of motion events or transfer of physical force*, i.e., a specific theory. From about 6 months of age, infants are surprised if they observe a temporal or spatial gap between the movement of an object A and the movement of an object B. However, if A nudges B and B starts moving right away, infants will show no signs of surprise. This indicates that they understand the causal link between the movement of objects A and B (Cohen & Amsel, 1998). Moreover, Kotovsky and Baillargeon (2000) found similar results with motion events. (2) Infants might reason about *agents* and their actions, i.e., goal-directedness, which is a specific view of children's innate or very early knowledge, because knowledge about goal-directedness is viewed as specific and not as derived from a general learning mechanism such as covariation. For example, Woodward et al. (1993) found that 7-month-old infants expected an object to move another object, but did not expect this for two persons. This indicates that infants understand that people have agency and can move of their own accord. (3) Causal reasoning might develop through information about *covariation*, which is a general process. This view posits that infants reasoning might be general from the very beginning. Sobel and Kirkham (2006) discovered that 8-months-olds are already sensitive to statistical patterns. The infants were confronted with two objects A and B that caused the appearance of an object at position C, but not the appearance of an object at position D. In the backwards blocking condition, infants observed that B caused C, thus eliminating A as a cause. In a second condition, infants saw that B caused D. When infants saw A, they looked longer at position C. Thus, the infants probably determined that A causes C.

The development of children's causal reasoning is discussed in the light of general and specific processes as well. A general process influencing this development might be learning through the observation of covariation, i.e., associative learning. If children observe that objects supported at their mass will remain stable, they may infer that the mass determines stability. Accordingly, Kushnir and Gopnik (2005) found that 4- to 6-year-old children base their judgments of causal relations on covariation and infer the strength of this relationship by the frequency of co-occurrence. Moreover, children also learn from statistical patterns found in their own explorations (Bonawitz et al., 2019). In this study, 4- to 5- year-old children first learned that there were multiple categories, categorised the objects, and last inferred the causal relationships between these categories. Muentener and Bonawitz (2017) conclude that children apply statistical learning strategies that go beyond mere associative learning.

Furthermore, children might also employ specific information in the form of prior causal knowledge that is embedded in specific beliefs or theories. A child with a prior theory about the importance of mass for stability might pay attention to the mass instead of the colour, shape, or material, and therefore facilitate the complex input. For example, Muentener and Carey (2010) found that 8-months-old infants do not solely rely on covariation patterns, but infer that humans are causal agents. This implies that infants do not only rely on general learning processes, but use their specific psychological knowledge that humans can act as agents as well. Furthermore, 18-moths-old children used their prior knowledge that humans are agents, while inanimate objects are not, as they inferred an adult's intended actions (Meltzoff, 1995). Additionally, while prior knowledge might facilitate children's causal learning in a specific domain, children do not necessarily need to understand underlying complex processes to draw causal conclusions (Muentener & Bonawitz, 2017).

Both approaches might be reconciled, because they complement each other. Schulz and Gopnik (2004) found that young children from about 4 years of age are more likely to accept domain-consistent causes. However, after observing a large amount of statistical evidence they transferred their causal knowledge to other domains. Schulz, Gopnik, and Glymour (2007) came to similar results concerning biological and psychological domains. Children integrate evidence, but 4-year-olds still choose domain-consistent causes that were consistent with their prior theories more often than the correct cross-domain cause. Additionally, 3-year-olds did not integrate statistical evidence and were unable to choose cross-domain causes. Sobel and Munro (2009) expanded these findings in a series of studies. They discovered that 3-year-old children are able to transfer causes across domains, although they are more likely to do this if they understand the reason for the causality. Concerning, children's knowledge and theories about stability, their reasoning only needs to take place in a single domain, i.e., stability. Since studies found that children are able to transfer causes across domains, but face problems with it, supporting children's search for domain-consistent causal relations such as the importance of the mass for stability might facilitate children's knowledge or theory acquisition about stability.

A theory that is concerned with children's intuitive theories and the statistical probabilities that these intuitive theories are correct, and therefore combining general and specific processes, is called *theory theory* (Gopnik, 2013). Moreover, theory theory values the role of causality. Additionally, it encompasses deductive conclusions from an intuitive theory concerning novel entities, inductive reasoning from entities or events to a new intuitive theory, and abductive inference by including Bayesian priors. Therefore, theory theory combines different reasoning processes with general and specific processes.

1.1.7 Theory theory, Bayesian inference, and theory-evidence coordination

Human beings – like all biological organisms – try to achieve homeostasis (Schwartenbeck et al., 2013). In the context of children's knowledge acquisition, Piaget (1950) referred to this concept as cognitive equilibrium. Following this assumption, Friston (2012) states that humans try to minimize the chance of surprises that might potentially have negative consequences by constructing and adapting knowledge, a process that Piaget (1950) referred to as assimilation and accommodation. The basic idea of intuitive (science) knowledge is that people possess intuitive theories about physical regularities as well as other concepts such as psychological or biological processes (Carey, 2009). Intuitive theories explain things that are observable, e.g., a tumbling block, with abstract principles that are often unobservable, e.g., the principle of force (Gerstenberg & Tenenbaum, 2017). The theories and thus the explanations may change especially during childhood and adolescence and become more sophisticated (Carey, 2009). The study of children's cognitive development tries to explain how children's intuitive theories develop into the complex and abstract representations that allow adults to explain and predict phenomena they come across, as well as make plans in order to influence their surroundings (Gopnik & Wellman, 2012).

Theory theory derives from a constructivist view and offers a framework that facilitates the explanation of this development. Theory theory postulates that children construct intuitive theories about the surrounding world, which share at least five characteristics with complex and abstract scientific theories (see Carey, 2009; Gerstenberg & Tenenbaum, 2017; Gopnik, 2013; Gopnik & Wellman, 2012). (1) Both encompass causal representations of the surrounding world. These representations can include complex constructs that are not directly observable and allow scientists and children to explain and shape their surroundings. For example, theories about stability might embed the support of the centre or the mass as the deciding feature. (2) They may be hierarchically organized and are coherently structured. This means that theories explaining relations in a (science) domain are divided into more specific theories concerned with a particular relation in that domain. (3) They provide possible explanations for regularities. Therefore, theories influence how a scientist or a child explains evidence. For example, a child with a mass theory is likely to view the support of the mass as the deciding factor. (4) They allow predictions of regularities. Based on their theories, scientists and children can infer what might have happened in the past, what might be happening in the future, and what might happen if they tried something new or manipulated their surroundings. (5) They can be adjusted in the face of counterevidence. Therefore, if a child or a scientist observes evidence that conflicts with their theoretical assumption and cannot be assimilated into their theory, they may react in one of two ways. They may either ignore or explain away the evidence that conflicts with their intuitive theory (Kuhn, 2014; Simons, 2000), or they may change their explanations for certain subordinate relations or even their general assumptions about regularities (Bonawitz et al., 2012; Gopnik & Wellman, 2012). Children may come across such counterevidence by chance in their everyday lives either through their own interventions, e.g. in their play, or through observing the interventions of others. Such change or adjustment often takes place gradually (Gopnik & Wellman, 2012). It becomes increasingly more likely if a person observes repeated counterevidence that forms a pattern indicating a systematic cause rather than a coincidence (Koslowski & Masnick, 2014). For example, if a child observes repeated counterevidence for their theory that support of a block's centre if sufficient for the block to remain stable, they might adjust their theory that for some blocks the centre is sufficient and for others not.

While scientists test their theories specifically and deliberately and design experiments for this purpose, children, for the most part, do not test their theories with carefully designed experiments, but adjust their theories when they are confronted with evidence in their everyday lives (Gopnik, 2013). Schulz and Bonawitz (2007) found that, even though children's play is not systematic, it facilitates children's causal learning and children try to resolve confounded evidence in their play. Moreover, according to Gopnik and Meltzoff (1997), children pass through the same developmental processes and therefore have similar representations about the same objects at roughly similar times in their lives, while scientists have different representations according to the theories in their field of research. Research has provided support for theory theory, inter alia, in the domain of balance (Bonawitz et al., 2012) and biology (Schulz, Bonawitz, & Griffiths, 2007).

Concerning deduction, induction, and abduction, theory theory encompasses all three processes. Thus, deductive conclusions about new entities or principles can be drawn from intuitive theories, and intuitive theories can include conditional inferences (Gopnik, 2013) such as *If the mass is supported, the blocks remain stable*. Moreover, intuitive theories may be formed through induction. If a child observes that asymmetrical blocks remain stable when their centre is unsupported, but their mass is supported, they might infer that the mass is the deciding factor and integrate this new knowledge into their theory or form an entirely new theory. Last, theory theory also allows for abductive inference through the inclusion of Bayesian priors.

Researchers have applied probabilistic models, specifically *Bayesian inference*, to the theory theory framework to value the role of probability and prior knowledge on learning processes (Gopnik, 2013; Gopnik & Wellman, 2012; Schulz, Gopnik, & Glymour, 2007). Bayesian inference can be probabilistic in three ways (Gopnik, 2013). (1) The relations within the theory itself may be indeterminate, e.g., a block functioning as a counter-weight may but does not necessarily need to determine a construction's stability. Other variables such as distance might be responsible for determining the stability. (2) The relationship between theory and evidence may be probabilistic, e.g., a block functioning as a counter-weight can stabilise or destabilise a construction. (3) One theory might be more probable than others, e.g., a theory about a weight stabilising a construction might be more probable than a theory about the blocks' colours.

Furthermore, Bayesian inference indicates how a learner changes their theory after being confronted with a set of evidence and how children might combine theory and evidence (Schulz, Gopnik, & Glymour, 2007). Following the probabilistic approach, the construction of a theory derives from its probability before and after consideration of the produced evidence as well as the likelihood of the produced evidence if the theory is correct (Bonawitz et al., 2012). This can be exemplified by Bayes' rule presented in the following formula:

$$P\ (H/E) = (P\ (E/H)\ (P\ (H))\ \%\ P\ (E)$$

In this formula, P (H/E) represents the posterior, i.e., children's belief in a theory after evaluating the observed evidence, e.g., children's belief in centre theory after observing evidence for the importance of the mass. P (E/H) refers to the likelihood indicating the probability of the evidence given the prior theory, e.g., children with a centre theory might find the evidence for mass unlikely. P (H) equals the prior, i.e., the initial belief in a theory or the probability of a theory before observing evidence, respectively, e.g., indicated by the consistency with which children apply centre theory before observing evidence. P (E) represents the probability for the evidence itself, e.g., the probability of observing evidence for the importance of the mass. Thus, Bayesian inference allows for the inclusion of children's prior theories and the combination of a theory with new evidence concerning both the probabilities for the theory before and after observing the evidence (Gopnik, 2013). Therefore, if a theory has a high a priori probability, a child might need to observe a lot of evidence in order to accept a new theory. Otherwise they might keep their intuitive theory and dismiss or explain away the new evidence (Gopnik, 2013; Griffiths et al., 2011; Hawkins et al., 1984). Thus, children who have very strong and consistent presumptions are less likely to adapt their theory than children who have weaker and inconsistent presumptions. Studies have yielded results in favour of Bayesian inference (Bonawitz et al., 2012; Gopnik et al., 2001; Griffiths et al., 2011; Kushnir & Gopnik, 2007; Sobel et al., 2004).

Summarising, theory theory and its Bayesian application integrate the general and specific perspectives (Gopnik, 2013). Children's intuitive theories are always specific, e.g., theories about stability, and at the same time rely on statistical probabilities, e.g., in the form of Bayesian inference. This integration allows the explanation of young children's reasoning including its development and lays the foundations for its promotion.

While Bayesian inference suggests that children draw conclusions from evidence and adapt their intuitive theories accordingly, studies concerned with children's ability to *coordinate theory and evidence* call into question whether young children can draw conclusions from patterns of covariance and whether they can relate these conclusions to a theory (Koerber et al., 2005; Piekny & Maehler, 2013; Ruffman et al., 1993). In these studies, the children did not face problems relating theory with evidence if they observed perfect covariation, but when evidence was presented in the form of imperfect covariation they had trouble with theory-evidence coordination. The ability to evaluate imperfect covariation develops during the preschool years (Piekny et al., 2014). The studies on theory theory presented children with perfect covariation, which explains why the children in their studies changed their theories after observing the evidence (Bonawitz et al., 2012; Schulz, Bonawitz, & Griffiths, 2007). Furthermore, Gopnik et al.

(2001) found that children interpreted the presented evidence correctly. Sobel et al. (2004) expand on this result with their study on children's ability to make inferences from indirect evidence of perfect covariation in the form of data they had not directly observed.

Studies on theory-evidence coordination consider children's prior knowledge as well. Results yielded that prior knowledge plays an important role in children's interpretation of evidence, because children often interpret evidence in the light of their prior beliefs (Croker & Buchanan, 2011). However, Koerber et al. (2005) found that 3- to 4-year-old children are able to interpret covariation that contradicts their prior beliefs. Children might even change their theories after observing such contradicting evidence when presented in a way that they can interpret, e.g., perfect covariation. Moreover, Köksal-Tuncer and Sodian (2018) found that preschool children between the ages of 4 and 6 are able to generate hypothesis and generate disconfirming evidence when confronted with false claims. These results correspond with the assumptions of theory theory and Bayesian inference that children evaluate the probability of the theory before and after observing the evidence as well as the probability of the evidence itself and under the assumption that their prior theory was true. Following the analysis of these probabilities, children might either change their theory or dismiss the evidence (Gopnik & Wellman, 2012; Griffiths et al., 2011; Kuhn, 2014).

Concluding, children face certain problems with reasoning due to their developmental constraints that might affect their science learning. However, these constraints can be addressed by creating a learning environment that considers at least five factors. (1) Children should be familiar with the science content and already have developed prior knowledge or theories, as this facilitates their reasoning (Muentener & Bonawitz, 2017). (2) Children mainly face problems with interpreting imperfect covariation, but can interpret perfect covariation (Koerber et al., 2005). Children's science learning can be facilitated if they are confronted with perfect instead of imperfect covariation. For example, evidence should always confirm the importance of the mass. (3) Children's prior theories play an important role in the way that children relate theory and evidence and adjust their prior intuitive theories to a more scientifically correct one (Gopnik, 2013). Moreover, children often have similar intuitive theories about science regularities (Gopnik & Meltzoff, 1997). Therefore, interventions should consider these intuitive prior theories and confront children with counterevidence that contradicts these prior theories. Concerning theories about stability, evidence should disconfirm children's intuitive, but scientifically wrong, prior theories. (4) Children whose prior theories seem very probable to them, e.g., as they are very consistent, are less likely to adjust their theories and more likely to explain away the evidence (Gopnik, 2013; Griffiths et al., 2011; Hawkins et al., 1984). Thus, children should observe a sufficient amount of evidence. (5) Children should receive enough time to observe repeated evidence, e.g., for the mass, that challenges their intuitive theories, e.g., their theory that the centre is sufficient, as theory adjustment often happens gradually (Gopnik & Wellman, 2012).

However, children do not necessarily learn from the observation of counterevidence, since many possible theories might explain the evidence (Gopnik, 2013). By considering children's developmental constraints, e.g., by playing with the children and enriching the play with scaffolding materials and an adult's verbal support, children might be more likely to adjust their intuitive theories to a more scientifically correct one (see Copple & Bredenkamp, 2009; Zosh et al., 2018). A suitable science domain to investigate children's intuitive theories or knowledge and how to promote them is stability.

1.1.8 Children's knowledge and theories about stability

Children have intuitive theories about physics, a part of science, as early as infancy that they adjust towards more and more sophisticated theories during childhood and adolescence. Intuitive knowledge or theories about physics can often be studied with non-verbal tasks, rendering them especially suitable for young children (Wilkening & Cacchione, 2014). This also involves knowledge about stabilities (Baillargeon et al., 2014). Therefore, children's knowledge and their theories about stabilities can be investigated to learn more about children's cognitive development and their knowledge acquisition and theory adjustment in particular.

Studies on infants mostly used the violations of expectation method to investigate their knowledge of stabilities. In a typical violations of expectation task, infants will observe an expected and an unexpected event. It is hypothesised that infants look longer at an unexpected event than at an expected event, because they possess physical knowledge of an expected event, but are surprised by an unexpected event (Baillargeon et al., 2014). In a series of studies Baillargeon and her colleagues investigated whether infants understand that objects must be supported in order to remain stable. Needham and Baillargeon (1993) found that 4.5-month-old infants looked longer at a box that was pushed beyond a supporting surface and floated in mid-air compared to three other conditions, in which the box was either supported by the surface, held by a hand, or tumbled to the floor. This indicates that infants understand that an object must be supported in some way for it to remain stable. In another study, Baillargeon et al. (1992) found that infants at 6.5-month of age understood that partial support might be insufficient for a box to remain stable. The infants looked longer at an event in which only 15% of a box was supported than at two other events in which either the whole box or 70% of the box were supported. The researchers found that infants younger than 6.5-month of age looked at all three conditions for approximately the same amount of time, indicating that they thought that any amount of support was sufficient. Furthermore, Baillargeon and Hanko-Summers (1990) investigated whether 7.5- to 9.5-month-old infants knew that objects need adequate support in order to remain stable. The infants in their study were surprised when the symmetrical object remained stable even though it was not adequately supported by a small surface. However, when presented with asymmetrical instead of symmetrical objects, the infants looked equally long at the possible and the impossible event, suggesting that their knowledge of stability was limited when they were presented with asymmetrical objects. These studies show that infants have an intuitive knowledge of support, but that their knowledge is limited.

Siegler (1976) investigated children’s understanding of balance with the balance scale task. He found that young children at the ages of 5 and 6 only consider the weight on each side of the fulcrum when rating its balance. From the age of 9 onwards, children start to include the distance into their reasoning. But even some adolescents at the age of 16 face problems with the integration of the cross-product between weight and distance. In another study, Siegler and Chen (1998) found that 4- and 5-year-olds mostly consider the weight or had no consistent theory of balance. Very few 5-year-olds included the distance into their explanation.

Krist and his colleagues investigated 2- to 8-year-old children’s knowledge about stabilities with different methods. Krist et al. (2005) asked 4- to 8-year-old children to balance symmetrical and asymmetrical objects on a beam scale and found that performance increased with age. Additionally, children of all ages faced problems balancing asymmetrical objects, and faced less problems balancing symmetrical objects. Krist (2010) investigated whether 3- to 6- year-olds would rate the stabilities of photographs and animations of asymmetrical and symmetrical blocks that were either sufficiently or insufficiently supported correctly. The symmetrical blocks were placed on a supporting surface with 75% of the block being either supported or not supported, the asymmetrical blocks were placed in a way that 50% were supported by the surface. For both animations and photographs, children’s performance increased with age and again children faced problems with the asymmetrical, but not with the symmetrical blocks. In a similar study, Krist (2013) came to the same results with 3- and 4-year-olds. Last, Krist et al. (2018) applied an eye-tracking method with children between the ages of 2 and 6, and again found that performance increased with age and children had an easier time with symmetrical than asymmetrical objects. These results suggest that young children fail to consider the object’s mass, but do consider its geometrical centre. From a developmental perspective, most objects that children encounter in their everyday lives are more or less symmetrical. Therefore, they may learn to estimate their stability correctly at an earlier age (Krist, 2010). However, since the asymmetrical objects used in Krist’s studies were supported by 50%, they could only be rated correctly with mass knowledge. Children who only considered the centre could only guess.

In a series of studies, Pine and her colleagues came to similar results. In all studies, they asked children between the ages of 4 to 9 to balance symmetrical and asymmetrical blocks on a fulcrum. Most children succeeded with the symmetrical blocks, but not with the asymmetrical ones (Pine et al., 2002; Pine & Messer, 2000; Pine & Messer, 2003). They concluded that most children have a centre theory, and fail to consider the mass (and the distance). Moreover, in all their studies, they uncovered that children could learn how to balance asymmetrical blocks and learn about the importance of the mass through different interventions: The children (1) either observed the experimenter and explained why they could balance the blocks or they only observed and did not provide explanations (Pine et al., 2004; Pine & Messer, 2000); (2) played with the blocks freely (Pine & Messer, 2003); (3) had a group discussion (Pine & Messer, 1998); (4) either observed the experimenter or balanced the blocks together with the experimenter (Pine et al., 1999); or (5) observed the experimenter when balancing blocks or received direct feedback after they had failed to balance the blocks themselves (Pine et al., 2002). These results indicate that fostering children’s knowledge about stability is possible.

Bonawitz et al. (2012) categorised 4- to 7-year-old children according to their balancing of asymmetrical blocks. Children, who balanced by trial and error, were categorised as having a non-differentiated theory (*No theorists*), and all 4- and 5-year-olds fell into this category. Children, who balanced the blocks at their centre, were coded as being *Centre theorists*, and children, who considered the mass, were categorised as *Mass theorists*. Approximately 50% of the 6- to 7-year-olds were categorised as Centre and 50% as Mass theorists. Children then either observed evidence supporting centre theory or mass theory, which either conflicted or confirmed their prior theory. The children were asked to explain the evidence. Most No theorists explained the evidence with an undifferentiated theory, while the Centre as well as Mass theorists in the theory consistent condition remained with their theory. Centre theorists in the mass condition, however, were likely to adopt a mass theory. Mass theorists in the centre condition were likely to explain away the evidence that contradicted their belief. Afterwards the children were free to play with the blocks and were then again asked to balance an asymmetrical block. Many No theorists moved to Centre theory. These results expand the results from Pine’s studies and indicate that young children can acquire a more sophisticated theory when they are confronted with belief-conflicting evidence and are free to explore the evidence during play.

The results on children’s knowledge of stability imply that young children possess different knowledge/theories of stability, and that their mass knowledge/theory can be supported through various interventions. Therefore, children's stability knowledge/theory lends itself to studying the effects of prior knowledge as suggested by the literature on theory theory and theory-evidence coordination as well as the investigation of other age-appropriate intervention forms such as play. Nevertheless, the presented studies did not include potential effects of other interindividual covariates. Beyond cognitive components such as reasoning and knowledge acquisition, motivational components such as academic self-concept play an important role (Belland et al., 2013; OECD, 2014; Pintrich, 2003). According to theories concerned with academic self-concept, academic success and the corresponding self-concept are intertwined and fostering either leads to an increase in the other (Eccles, 2009; Marsh et al., 2012; Marsh & Craven, 2006). This relation has been identified even in young children (Guay et al., 2003). Therefore, a positive science self-concept may support children's engagement with science regularities and phenomena such as stabilities, and thus result in a sense of achievement in science learning (Marsh et al., 2012). Thus, science self-concept is discussed in the following chapter.

1.2 Academic self-concept in science

Shavelson et al. (1976) define self-concept as a person's perception of themselves as shaped by their experiences with their environment. They claim that a person's actions are influenced by their self-concept and their self-concept is expressed through behaviour (Ryan & Deci, 2000), e.g., by a person's choices, persistence and energy that is spent on a subject (Patrick & Mantzicopoulos, 2015). Self-concept is the evaluative component of one's ability in a domain (Jansen, 2017) and one of the most frequently investigated constructs in developmental and educational psychology, because it is a mediating factor for many desirable outcomes such as persistence, intrinsic motivation and accomplishments (Guay et al., 2004; Marsh et al., 2012; Wigfield & Eccles, 2000). Many researchers view self-concept as an important pillar of children's development (Marsh et al., 2002; Marsh et al., 2012; Patrick & Mantzicopoulos, 2015).

Shavelson et al. (1976) describe seven characteristics of self-concept, claiming that self-concept is (1) organised in the sense that experiences are integrated, and (2) multidimensional consisting of an academic domain and non-academic domains such as social and emotional self-concept. Moreover, it is considered to be (3) hierarchically organised with a general factor and subfacets, and (4) relatively stable. Self-concept (5) develops with age and becomes more and more differentiated during childhood and adolescence. In addition, self-concept is (6) evaluative, as a person compares themselves to their own prior achievements, other persons, or others' evaluations. Last, self-concept can be (7) differentiated as math self-concept is more closely related to math performance than, e.g., verbal or science performance. Marsh and Shavelson (1985) revised the original model by Shavelson et al. (1976), because research suggested that the individual academic facets were unrelated, contradicting the existence of a superordinate academic self-concept facet. Instead, they claim that self-concept can be divided into multiple academic facets such as mathematics, language, and science self-concept.

Following Marsh and Shavelson (1985), research on mathematics and language self-concepts suggests that they are the two ends of a spectrum of academic domains and are not related (Jansen et al., 2015; Marsh et al., 2014). This finding has been explained with the internal/external frame of reference model, which states that children tend to evaluate their ability in either mathematics or language through two different reference processes (Marsh, 1986). First, they compare their achievements in mathematics and language to the achievement of their peers. Accordingly, the achievement in both subjects is related. Thus, self-concepts in both subjects were hypothesised to be related as well. However, in the second reference process, the internal reference, a child compares their ability in mathematics with their ability in language. Self-concepts in math and language should be negatively related, because children tend to do better in one of the subjects (for an overview see, Marsh, Parker, & Craven, 2015). The combination of these two reference processes suggests that mathematics and language self-concepts are not related. Therefore, they are treated as two maximally different ends of a spectrum. Dimensional comparison theory extends the internal/external frame of reference model by incorporating other domains such as science that lie on the spectrum between math and language and are expected to correlate with math and language self-concept to some extent (Figure 1; Marsh, Parker, & Craven, 2015). Accordingly, Jansen et al. (2015) found relations between adolescents’ math and

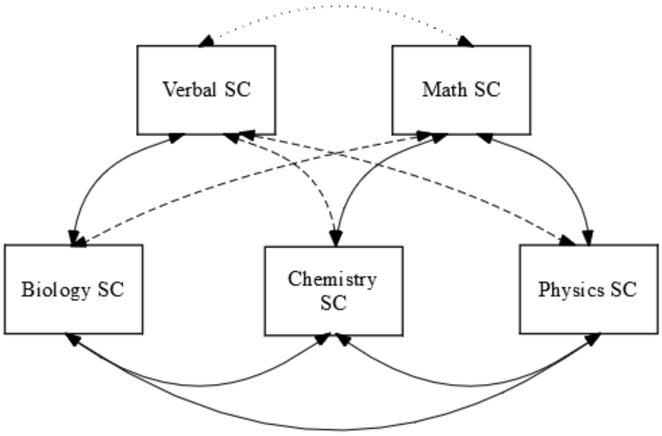


Figure 1. Hypothesised relationships between different self-concept domains. Solid lines indicate strong relations. Dashed lines indicate weak relations. Dotted lines indicate no relations.

language self-concepts and their science self-concepts in chemistry, physics, and biology. Therefore, the self-concepts in different science domains lie on the spectrum between math and language self-concepts.

Academic self-concept in a domain (physics, math, language, etc.) has long been constructed as containing specific competence beliefs and affective-motivational components that can be subsumed under one factor (Marsh, 2007). However, theories on the relationship between motivation and achievement, such as expectancy-value theory (Eccles et al., 1983), construct motivation as a value component, while competence beliefs are a part of the expectancy facet. According to this theory, motivational and competence components are related and influence knowledge acquisition (for overviews see, Eccles et al., 1983; Eccles, 2009; Kaplan et al., 2012; Schunk et al., 2008), because highly motivated children with high expectancies are more persistent and try to overcome obstacles in the learning process (Guo et al., 2015; Marsh et al., 2005). Studies have produced evidence that academic self-concept can be differentiated into a motivational and a competence component as early as kindergarten and elementary school (Arens et al., 2011; Arens et al., 2016; Arens & Hasselhorn, 2015). The motivational component can be conceptualised as a subfacet of the intrinsic value a person attributes to a task (*How much do I want to/do I enjoy building with blocks?*), while the competence component functions as a subfacet of a person’s expectancy (*How good am I at building with blocks?*; see Trautwein et al., 2013; Wigfield & Eccles, 2000). Both aspects are interdependent (Denissen et al., 2007), and influence a person’s achievement. This relationship has been investigated for preschool children (Arens et al., 2016), elementary school children and adolescence in mathematics (Guo et al., 2015; Lauermann et al., 2017; Wigfield & Eccles, 2000) and STEM (Andersen & Ward, 2014; Ball et al., 2017). Moreover, motivational and competence beliefs are associated with each other and their association in a specific domain increases during the elementary school years (Wigfield et al., 1997). Wigfield et al. (2009) suggest two reasons for this association. First, children may be more motivated in what they perceive themselves to be competent at. Second, motivational beliefs influence choice behaviour such as spending time playing with building blocks which may increase achievement and competence beliefs.

Concerning the development of self-concept, children as young as 4 years can differentiate between different (academic) aspects of their self, thus allowing the measurement of young children’s self-concept (Marsh et al., 2002). For children between the ages of 5 to 7, self-concept typically consist of specific competencies such as cognitive abilities, whereas interpersonal attributes such as being popular, smart or friendly towards others are missing for the most part and form in later childhood (Harter, 2015). Furthermore, preschool children tend to overestimate their abilities, which leads to an overly positive academic self-concept (Eccles, 2009; Harter, 2015; Helmke, 1999). Young children tend to engage in all-or-none thinking (Harter, 2015). This is supported by children’s development, because (1) at approximately 3 years of age, children start to include their own evaluations and those of attachment figures into their self-concept (Rochat, 2013). (2) At around 5 years of age they start to link representations to one another, for example, linking in the form of opposites, e.g., things or persons are either good or bad (Case, 1985; Griffin, 1992). Since most parents tend to encourage and praise their children, children acquire a sense of being good at a number of things and therefore they assume that they are all good and thus cannot be bad. This contributes to their overly positive self-concept (Eccles et al., 1993; Harter, 1999, 2015). During the primary school years, children’s academic self-concept tends to become more realistic, as they are confronted with feedback and start to compare themselves to their peers (Helmke, 1999).

There has been extensive research on children’s and adolescents’ self-concepts in language and mathematics (e.g., Guo et al., 2015; Helmke, 1999; Marsh et al., 1988; Skaalvik & Valås, 1999). But research on science self-concepts has been sparse. Even though research suggests that the perception of one’s own abilities and motivational prerequisites matters more than actual ability for educational outcomes (Marsh, Lüdtke, et al., 2015; Murayama et al., 2013) and that a high self-concept is especially important in subjects that children tend to find tiresome or hard such as science (Jansen, 2017). Moreover, it is hypothesised that educational outcomes such as knowledge acquisition matter for the development of academic self-concept as well. This is described in the reciprocal effects model that is concerned with the interaction of academic self-concept and corresponding academic success (Marsh & Craven, 2006). Results in support of the reciprocal effects between academic self-concept and academic achievement have been found for mathematics, language and general school achievement for primary school and high school students (e.g., Guay et al., 2003; Guay et al., 2010; Marsh et al., 2018; Marsh & Martin, 2011). Studies on this relationship in the science domain are few, but point to a positive relation between science achievement and science self-concept for primary school children (Denissen et al., 2007) and adolescents (Denissen et al., 2007; Jansen et al., 2014; Jansen, 2017; Möller et al., 2006). The relation between achievement or knowledge acquisition and science self-concept for preschool children has only been investigated for math self-concept (Arens et al., 2016). The authors found effects of early math achievement on math competence self-concept with a cross-lagged panel model. However, they found no effects for competence self-concept on achievement and no reciprocal effects for motivational self-concept and math achievement. The relation between preschool children’s academic self-concept and their corresponding achievement might differ from the relation in school children. Arens et al. (2016) reason that the difference might stem from the preschool education system in Germany, because education mostly consists of learning during free play instead of formal instruction. Therefore, children might not have related their academic self-concept to their actual achievement. Moreover, in a recent study, the reciprocal effects model has been called into question for young children, as different statistical approaches lead to different conclusions (Ehm et al., 2019). Evidence supporting this relationship was uncovered with a classic cross-lagged panel model and a full-forward cross-lagged panel model. A random intercept and a change score model did not support the reciprocal effects model. Further studies on the relationship in preschool children are needed.

Nevertheless, the reciprocal effects model and its assumptions allow for the derivative of successful interventions that allow positive experiences with and promote positive attitudes towards science learning and may thus lead to higher knowledge acquisition in the sciences. Interventions targeting academic outcomes, e.g., science skills also affect corresponding self-concept (for a meta-analysis see, O'Mara et al., 2006). However, interventions undermining competence beliefs might not have the desired effects on academic achievement or knowledge acquisition (Marsh et al., 2012). Moreover, young children are highly motivated to learn about science (Oppermann et al., 2017). This high motivation can serve as an introduction to science education by enabling children to work with materials they are familiar with or by familiarising children with new materials. Nayfeld et al. (2011) found that children’s engagement with the science areas of their preschool increased after they were introduced to the materials and had understood how to handle them. Furthermore, playful interventions with materials that children are familiar with, e.g., building blocks, could promote children’s science knowledge and theories and at the same time maintain their high motivation and competence beliefs (see Bonawitz et al., 2011; Zosh et al., 2018). Children, who have a high mastery motivation will be more persistent and will probably acquire more knowledge about the underlying science principles. Moreover, following the reciprocal effects model, children who acquire more knowledge will also have more positive and more stable competence beliefs, as they are shaped through experiences (Shavelson et al., 1976). Therefore, academic self-concept should be considered when developing interventions that promote specific achievement and learning gains.

Concerning science learning, Patrick et al. (2009) and Samarapungavan et al. (2011) discovered that an inquiry-based approach supported preschool children’s science knowledge and their science self-concept. Children in the inquiry group had acquired more science knowledge and had a higher science self-concept compared to a control group, which had received their kindergarten’s traditional science education. However, the researchers did not investigate the relationship between science knowledge acquisition and science self-concept. In another study, Marsh and Richards (1988) implemented an intervention aiming at enhancing adolescents' math and reading achievements and their corresponding self-concepts. They found that the intervention enhanced both achievement and corresponding self-concepts. However, in a study with primary school children, Craven et al. (1991) demonstrated that feedback on positive abilities and performance enhanced reading and math self-concepts, but not corresponding achievement. Despite these ambiguous results, the studies by Marsh and Richards (1988), Patrick et al. (2009), and Samarapungavan et al. (2011) imply that enhancing science knowledge and corresponding science self-concept is possible. Self-concept is promoted through positive experiences that highlight competence and might be gained during play (Trawick-Smith, 2012). Accordingly, play could support both knowledge acquisition and corresponding self-concept, because children gain positive experiences with the subject, while also learning about the underlying science concepts (Weisberg et al., 2013; Zosh et al., 2018).

Drawing from these explanations, three major research gaps in research on the relation between children's academic self-concepts and their academic achievement can be derived. Research is sparse on (1) children younger than 6 years; (2) the relation between science achievement and science self-concept; and (3) how to promote both achievement and academic self-concept in a science domain. This dissertation aims to contribute to these discussions.

Besides science self-concept, which might interact with science achievement, other prerequisites might contribute to children's knowledge and theory acquisition. Possible prerequisites might be intelligence, e.g., fluid and crystallised, and spatial intelligence such as mental rotation ability, which all have been found to contribute to knowledge and theory acquisition (for an overview see, Flynn & Blair, 2013; Newcombe et al., 2013). Furthermore, self-related components such as interest in block play might contribute as well, because interest leads to increased involvement (Crouch et al., 2018).

1.3 Facets of intelligence and self-related constructs as possible prerequisites

Possible prerequisites for children's science learning are intelligence and self-related constructs. Intelligence is considered one of the key prerequisites for learning, and is conceptualised as a *g*-factor that consists of two sub *g*-factors, fluid intelligence—indicated by figural perception and figural reasoning—and crystallised intelligence—indicated by language capacity (Cattell, 1963; Flynn & Blair, 2013). Furthermore, it contains different facets such as spatial ability, e.g., mental rotation ability, that are related to children's STEM success (Baddeley, 1986; Newcombe et al., 2013).

Interest is a self-related construct that supports meaningful engagement in an activity, such as block play, and through that engagement might foster knowledge acquisition (Crouch et al., 2018). Therefore, it is worthwhile to consider both intelligence and interest as possible prerequisites for knowledge acquisition in science.

1.3.1 Fluid intelligence

Figural perception and figural reasoning imply the ability to perceive figural content, mentally represent it, abstract it, and identify spatial regularities. Thus, they are related to spatial abilities (Newcombe et al., 2013), and both are indicators of fluid intelligence (Cattell, 1987; Weiß & Osterland, 2013). Fluid intelligence is defined as the ability to solve new or unfamiliar problems (Cattell, 1963), and influences learning and mental functioning (Flynn & Blair, 2013).

Concerning children’s knowledge and theories about stability, a higher ability to represent objects mentally and identify underlying spatial principles and similarities between different objects might contribute to children’s knowledge acquisition and theory adjustment. For example, children with higher abilities might detect an object’s mass as the deciding factor for stability easier than children with lower abilities.

1.3.2 Crystallised intelligence

Language capacity is an indicator for crystallised intelligence, which influences the development of mental functioning and knowledge acquisition (Brydges et al., 2012; Cattell, 1987; Flynn & Blair, 2013; Thorsen et al., 2014). Crystallised intelligence is an accumulation of knowledge already acquired (Cattell, 1963).

Concerning children’s knowledge about stabilities, crystallised intelligence might relate to children’s knowledge and theories about stabilities that they already acquired through their experiences. Furthermore, children might have more complex theories regarding stability and therefore might find it easier to integrate new knowledge into their theories (see Gopnik & Wellman, 2012; Thorsen et al., 2014).

1.3.3 Mental rotation ability

Mental rotation is specified as the ability to recognise that two pictures depict the same object or figure from different perspectives (Shepard & Metzler, 1971). Mental rotation ability belongs to the broad category of spatial abilities, more precisely, it indicates intra-object representations and the ability to change them (Newcombe et al., 2013). The ability to mentally rotate objects emerges in children as young as 4 years of age, even though their abilities are far from fully developed (Newcombe et al., 2013; Quaiser-Pohl, 2003; Quaiser-Pohl et al., 2010).

According to Jäncke et al. (2001), mental rotation has three prerequisites. First, a visually perceived object has to be mentally represented. Second, the spatial features of the mental representation have to be uncoupled from its other features. Third, the spatial features have to be processed. Concerning the relation with children's knowledge of stability, the same underlying mental processes might apply. Children need to represent the block construction mentally, abstract the spatial features and imply the construction's stability. Furthermore, mental rotation ability is an important prerequisite for interest and success in STEM (Webb et al., 2007). This relation might be especially relevant for science knowledge with spatial components such as knowledge about stabilities.

1.3.4 Interest

Hidi et al. (2017) define interest as an individual's cognitive and affective involvement in a content or an activity. A prerequisite for interest in a topic is that a person at least has heard or has already acquired some knowledge about it. Children who are interested in a certain content believe that they can have mastery experiences when engaging in it and are more likely to engage with the content and seek a deeper understanding of it than children with a lower interest (Renninger & Hidi, 2016). Therefore, interest supports meaningful engagement and therefore knowledge acquisition (Crouch et al., 2018). Children who are not interested in a topic might become more interested if they learn or uncover something that they find fascinating or rewarding. Afterwards they are more likely to reengage with the content (Hidi et al., 2017; Renninger & Hidi, 2016).

Interest can be triggered through the person themselves, attachment figures or environmental influences and develops through interaction with a content (Hidi et al., 2017; Renninger & Hidi, 2016). According to Renninger and Hidi (2016), interest develops in four phases. First, a person shows situational interest that results from short-term changes in cognition and/or affect affiliated with a topic or an activity. A person might need support from others, e.g., for children this might be parents, teachers or peers, or, in school and kindergarten, through the instructional design to maintain their interest. If the interest is sustained, the person enters the second phase, in which the interaction with the content continues over time. In this phase, children start to acquire knowledge about the content and associate it with positive emotions. In the third phase, the interaction with the content becomes even more regular. Children will reengage with the content independently of others and the instructional design. They have knowledge about the content, reflect about it and are likely to pose questions and try to answer them. In the fourth phase, a person has a well-developed interest and engages with the content regularly and intensely. Children are more likely to persevere through challenges and overcome frustration associated with the content.

Young children are very interested in science before they have started school (Oppermann et al., 2017). They like to listen to informational texts about topics from biology and physics, e.g., levers (Mantzicopoulos & Patrick, 2010). Furthermore, children engage in science activities in their free play time in preschool or kindergarten by reading science books, playing with magnets, blocks, etc. (Rubin et al., 1978; Saçkes et al., 2011). Furthermore, studies have uncovered a positive relation between science interest and science achievement for adolescents (Jansen et al., 2016) and for preschool girls (Leibham et al., 2013).

Concluding, following Hidi et al. (2017) and Renninger and Hidi (2016), a higher interest in block play might enhance children's engagement with the content and therefore their knowledge about science phenomena associated with block play, such as knowledge about stabilities. Moreover, higher knowledge might lead to higher interest in block play as well.

Chapter 1 was concerned with children's cognitive developmental, their science self-concept and prerequisites for theory and knowledge acquisition. They lay the foundation for children's learning and provide information concerning possible ways to support knowledge and theory acquisition by highlighting children's developmental constraints and prerequisites. Chapter 2 is concerned with possible ways of promoting children's development, while considering their constraints and prerequisites. One possibility of implementing support in an appropriate manner is playful learning (Copple & Bredenkamp, 2009; Golinkoff et al., 2006b; Patrick & Mantzicopoulos, 2015)

2 Support

How might science learning and children's trust in their own science abilities be supported from an early age on? Playful learning is considered developmentally appropriate for young children (Copple & Bredenkamp, 2009; Golinkoff et al., 2006b; Patrick & Mantzicopoulos, 2015). Through an adult's guidance, children's knowledge acquisition and their theory adjustment can be supported (van de Pol et al., 2010). Children's cognitive developmental constraints, e.g., their reasoning, may be considered during play and addressed through targeted guidance provided by an adult (Golinkoff et al., 2006b; Schulz & Bonawitz, 2007). Moreover, children's motivational and competence beliefs may be promoted as well (Belland et al., 2013; Pintrich, 2003).

2.1 Play

Play has been a research focus for researchers from many different disciplines such as psychology and biology for decades (e.g., Piaget, 1952; Vygotsky, 1978). Theorists believe that from an evolutionary biological perspective, play is not necessary for survival (Burghardt, 2011), but must have helped individuals acquire skills and understand their environment; and while there must have been other ways to acquire these skills and knowledge, play was probably the less costly and less difficult alternative (Bateson, 2011; Pellegrini & Pellegrini, 2012). Findings from developmental psychology show that play supports a variety of important constructs, such as children’s academic self-concept (see Weisberg et al., 2013), their knowledge acquisition and understanding (Borriello & Liben, 2018; Fisher et al., 2013; Verdine et al., 2019), causality (Schulz, Bonawitz, & Griffiths, 2007), as well as their social skills (Hinkley et al., 2018; Reynolds et al., 2011). Bateson (2011) considers play as so crucial for human development that he refers to it as *developmental scaffolding*. Due to these theoretical considerations and research findings, play is considered to be developmentally appropriate practice for young children (Copple & Bredenkamp, 2009; Fisher et al., 2011), and thus many mandate that children receive more time to play and allowing teachers and parents to use the pedagogical possibilities of it (Fisher et al., 2011; Golinkoff et al., 2006a; Hirsh-Pasek et al., 2009).

Children start to play in infancy. In that phase, their play is mostly sensorimotor in structure and function, as they act on objects without regarding their physical characteristics. For example, an infant might shake their teddy bear and visibly enjoy the activity. By the second year, children’s play becomes more representational and children start to combine objects in meaningful ways. Their play becomes increasingly complex. For example, a child might sing their teddy bear to sleep and later the child might act as if the teddy bear sings another toy to sleep. By preschool, children’s play has two dimensions, i.e., type of activity and social participation. Children might build a block tower on their own, or build a city out of building blocks together with other children (for an overview, see Pellegrini & Bjorklund, 2004; Pellegrini & Smith, 1998; Rubin & Smith, 2018).

Despite the interest researchers and theorists hold for the subject, they have struggled to define play (Rubin et al., 1983; Zosh et al., 2018). Most definitions have certain elements in common. The two core characteristics that are acknowledged in almost all definitions of play are (1) the voluntariness with which children engage in play and (2) the intrinsic motivation/joy children experience while playing (e.g., Burghardt, 2011; Rubin et al., 1983; Trawick-Smith, 2015; Zosh et al., 2018). These distinguish play from other types of behaviour, e.g., consummatory behaviour (Rubin et al., 1983). Furthermore, voluntariness and intrinsic motivation/joy are necessary, but not sufficient conditions for a behaviour to be playful, as other behaviours might share the characteristics, e.g., having your favourite meal (Burghardt, 2011).

Moreover, play is considered to be (3) process-oriented rather than goal-oriented (Daubert et al., 2018; Pellegrini, 2013; Rubin et al., 1983; Rubin & Smith, 2018). However, play can have goals, which are self-imposed by the player (Rubin & Smith, 2018). A playing child might have a goal in mind, e.g., building a high tower out of building blocks, but might lose interest in that goal while playing and instead decide to build a zoo or a farm or the process of building itself might become more important and more joyful than actually achieving the goal. This characteristic helps distinguish play from behaviour that has an externally imposed goal that needs to be reached, but is intrinsically motivated (i.e., joyful work), or behaviour without a goal (Rubin et al., 1983).

Furthermore, play is (4) child-directed and (5) contains elements of choice (Weisberg et al., 2016; Zosh et al., 2018). This indicates that children maintain control of the activity at all times and can decide what they want to do and how they would like to continue the activity.

Concluding, play might be defined as voluntary, intrinsically motivated/joyful, process-oriented as opposed to goal-oriented, child directed, and containing elements of choice (Burghardt, 2011; Daubert et al., 2018; Pellegrini, 2013; Rubin et al., 1983; Rubin & Smith, 2018; Trawick-Smith, 2015; Weisberg et al., 2013; Weisberg et al., 2016; Zosh et al., 2018).

Some researchers have defined play as a category, with a behaviour being either play or not play (Burghardt, 2011; Pellegrini, 2013; Pellegrini & Pellegrini, 2012). This view stems from an evolutionary biological/psychological perspective and defines play rather narrowly, so that many activities children engage in could not be considered play. For example, Pellegrini (2013) called into question whether block play should be considered play, because children might have a goal. However, following Rubin et al. (1983), block play can be considered play as children are focused on the process of building and their goals are self-imposed. Furthermore, if we assume that activities that have a goal are not play, all activities designed to help children learn cannot be play by definition. Even though, children might perceive the activity as play and all other characteristics are fulfilled (Zosh et al., 2018).

Therefore, many developmental psychologists view play as a continuum (e.g., Borriello & Liben, 2018; Fisher et al., 2013; Rubin et al., 1983; Weisberg et al., 2013; Weisberg et al., 2016; Zosh et al., 2018). The continuum view of play constructs play as a spectrum with differences in who initiates the play, who directs the play and if there is a learning goal (Table 1; Zosh et al., 2018). According to this view, free play fulfils all characteristics of play. Guided play and games have a learning goal and are initiated by an adult, but the child has complete control of the activity and directs it, engages in it voluntary and enjoys it. The adult merely offers support during play (Zosh et al., 2018). Guided play and games can be distinguished, as play is not rule-governed (Daubert et al., 2018). Zosh et al. (2018) define two other activities on the continuum, namely, co-opted play, in which a child initiates a play, but an adult takes control of the activity, and playful instruction, which is merely direct instruction with a playful overtone. They do not consider these two forms play, because the children lose their agency. Furthermore, guided play is related to guided discovery (see Weisberg et al., 2016; Zhang, 2016), as the two principles share certain characteristics. Both rely on guidance without offering solutions to teach children about regularities (Reuter & Leuchter, 2020). However, guided discovery differs from guided play in one important aspect, as children are expected to take part in the discovery activity. Therefore, intrinsic motivation/joy is not a necessary characteristic of guided discovery, but is one of the two main characteristics of play found in almost all definitions (e.g., Burghardt, 2011; Rubin et al., 1983; Trawick-Smith, 2015; Zosh et al., 2018).

Table 1

Dimensions of guided play following Zosh et al. (2018)

	Free play	Guided play	Games	Guided discovery	Co-opted play	Playful instruction	Direct instruction
Voluntary	Yes	Yes	Yes	No	Yes	No	No
Joyful	Yes	Yes	Yes	No	Yes	Yes	No
No external goals	Yes	No	No	No	No	No	No
Child directed	Yes	Yes	No	Yes	No	No	No
Elements of choice	Yes	Yes	Yes	Yes	No	No	No

The view of guided play being play is in line with the view of Rubin’s research group and Piaget’s and Vygotsky’s work (Daubert et al., 2018; Piaget, 1952; Rubin et al., 1983; Rubin & Smith, 2018; Vygotsky, 1978). None of these researchers claim that play needs to be free of goals, but rather that the process of playing is more important to the player than a (learning) goal. In guided play, the children still direct the activity, they engage in it voluntary and can enjoy it, they can choose what they would like to play and how and if they would like to continue playing (Fisher et al., 2011; Weisberg et al., 2016; Zosh et al., 2018). The children might even have self-imposed goals that are different from the learning goals the adult has in mind. Furthermore, reaching the learning goal is not mandatory, meaning that children are encouraged to pay attention to certain characteristics of the material during the play, but they are free to explore the materials themselves and play with them as they wish.

Guided play can take at least two forms. The adult can either initiate a play by providing materials, e.g., suggestions for building with building blocks, or support the children verbally while they play (Weisberg et al., 2016). Concerning the effectiveness of guided play by means of learning, Zosh et al. (2018) hypothesise that during guided play six measures may support children's learning: encouraging children to *explore*; fostering their *engagement* in the play; helping children to *make sense* of things by relating them or asking questions; encouraging children to share their knowledge with their *peers*; structure the play so that it is *iterative*; and maintain children's *joy* for the activity.

Concluding, guided play is viewed as a suitable way for children to learn about a variety of subjects, including language (Borriello & Liben, 2018; Ferrara et al., 2011), mathematics (Verdine et al., 2019), and science (Reuter & Leuchter, 2020). Guided play can maintain children's intrinsic motivation (Bonawitz et al., 2011), while still allowing an adult to scaffold the activity (see van de Pol et al., 2010). Following the ideas of Vygotsky (1978), an adult may help children master challenging tasks, which they might not have been able to achieve by themselves, and acquire new insights through scaffolding, which might also enhance children's science self-concept.

2.2 Scaffolding

Scaffolding is a concept derived from Vygotsky's social-cultural theory (see e.g., Vygotsky, 1978) that was further developed by Wood et al. (1976) to explain how an experienced person can support a novice in their learning. Regarding guided play, scaffolding can support children's learning (Fisher et al., 2013; Weisberg et al., 2016) as well as their corresponding self-concept. Scaffolding during guided play can take place in the form of (1) material scaffolds that structure the learning content, and (2) verbal scaffolds that can be adapted to a learner's progress and target their motivational and competence self-concepts (Belland et al., 2013; Guthrie et al., 2004; Kleickmann et al., 2016; Martin et al., 2019; Miller & Wang, 2019; van de Pol et al., 2010).

Material scaffolds are most effective when they are carefully designed to consider children's prior knowledge or intuitive theories about the learning content, and create a link between prior knowledge or theories and the new content (Leuchter et al., 2014). They can be used to initiate and structure a guided play (Weisberg et al., 2016), as well as highlight counterevidence to an intuitive prior theory that many children have about a science regularity (see Gopnik & Wellman, 2012). Regarding children's theories about stability, scaffolding materials in the form of photographs could show a counterweight stabilising or destabilising an object, thus offering evidence for a mass theory, while at the same time negating children's centre theory. Children could rebuild these photographs, and observe evidence contradicting their prior theory. Furthermore, material scaffolds can draw attention to important and new aspects (DeLoache, 2014), e.g., by highlighting that the support of the centre is insufficient for asymmetrical blocks, children might find it easier to realise that their centre theory is incorrect.

Verbal scaffolds are adaptable to a learner's progress (Kleickmann et al., 2016) and are especially important for heterogenous learning groups, e.g., preschool children (Sylva et al., 2007). In guided play, an adult can play along with the children and offer verbal support during the play to encourage higher order thinking (Chin, 2007; Haden, 2010; Kleickmann et al., 2016; Martin et al., 2019; Weisberg et al., 2016) and motivational and competence self-concepts (Belland et al., 2013). Examples for verbal scaffolding techniques that support children's cognitive activities are activating prior knowledge, asking for the child's reasoning, providing explanations, encouraging comparisons, and modelling (for an overview see van de Pol et al., 2010). For motivational and competence self-concepts, scaffolding can encompass promoting the perceptions of challenge, competence and success (Belland et al., 2013)

Children's *prior knowledge* can be activated through questions and asking children for their presumptions. By referring to knowledge that a learner has already acquired the process of assimilating new aspects into existing schemata is facilitated (Gurlitt & Renkl, 2010; Mayer, 1997; Piaget, 1950; Weinert & Helmke, 1998). For example, an adult might ask a child about their knowledge about building with blocks and stabilities. Moreover, an adult can *ask for children's reasoning* through questions that encourage children to explain their underlying thoughts and assumptions. This supports children in structuring their thinking processes (Hsin & Wu, 2011). During a guided block play, an adult could ask children for their reasoning behind stabilities, and urge them to explain why a structure remains stable or tumbles. Additionally, an adult can *provide explanations* to help integrate observations of evidence into children's interpretation of a regularity (Murphy & Messer, 2000; Renkl, 2002). An adult explaining why a counterweight can stabilise a block might help children organise their knowledge and structure thinking processes (see Richey & Nokes-Malach, 2013). *Encouraging comparisons* supports children’s understanding of similarities and differences between entities and helps generalising underlying regularities or concepts (Hsin & Wu, 2011; Richey & Nokes-Malach, 2013). An adult could encourage children to compare the stabilities of asymmetrical and symmetrical block constructions to support children’s mass knowledge or mass theory acquisition. Furthermore, *modelling* refers to an adult performing certain goal-directed behaviours and thinking styles by thinking aloud, which invites children to imitate (Hmelo-Silver et al., 2007; Mayer, 2004). This scaffolding technique offers support especially for young children (Vygotsky, 1978). For example, an adult might show children how to stabilise block constructions.

For motivational and competence self-concepts, an adult can *promote the learner’s perception of challenge* by assuring the learner that they can solve a task they may struggle with. This might enhance their expectancy of success (Britner & Pajares, 2006). Moreover, *by referring to the learner’s competence* an adult can underline the competences that the learner has already acquired and the competences that can be acquired through working on a task. This can reassure learners and highlight new strategies (Kaplan & Maehr, 2007). Last, *referring to a learner’s success* and encouraging them to explain their strategy for solving a task can encourage them to take a closer look at a content and invoke a sense of pride (Pintrich, 2003; Schunk et al., 2008).

Few studies are concerned with the comparison of the effect of guided play with material and verbal scaffolds and guided play with material scaffolds on children’s knowledge acquisition. In a study by Leuchter and Naber (2019), material and verbal scaffolds increased 6- to 7-year-old children’s knowledge about force. This form of guided play was superior to only material scaffolds, only verbal scaffolds, and free exploration. Accordingly, Hadzigeorgiou (2002) came to the same results concerning children’s knowledge about mechanical stability.

Material and verbal scaffolds can be implemented into children’s play to support their knowledge acquisition and theory adjustment regarding a scientific regularity, while still maintaining the core characteristics of play as being intrinsically motivated and voluntary (Zosh et al., 2018). Therefore, play that is supported by scaffolds is developmentally appropriate even for preschool children (Golinkoff et al., 2006a). Concerning children’s theory adjustment, scaffolds can relate to children’s prior knowledge and theories, present counterevidence, and support children in coordinating their theory with new evidence (Gopnik & Wellman, 2012). As scaffolds facilitate knowledge acquisition and theory adjustment, children might also feel a sense of achievement. According to the REM, this might contribute to children’s science self-concept

(Marsh & Craven, 2006). Moreover, scaffolds can target children's science self-concept as well (Belland et al., 2013).

Summing up, the investigation of children's knowledge and theory acquisition, and their science self-concept may serve, inter alia, the purpose of laying the foundation for science education. Moreover, by examining possibilities to promote children's development, while considering their developmental constraints and prerequisites, new insights into how best to support their development may be acquired.

3 The present dissertation

One purpose of science education is the promotion of children's science reasoning and knowledge acquisition with regard to their motivational and competence self-concept and their developmental constraints (Belland et al., 2013; Copple & Bredenkamp, 2009; OECD, 2014). Young children's cognitive development lays the foundation for their science learning and understanding, and theories such as theory theory are concerned with children's learning (Gopnik & Wellman, 2012). Suggestions on how to promote children's science knowledge and theories derived from theory theory involve taking children's prior theories into consideration and confronting children with counterevidence that cannot be assimilated into these prior theories (Bonawitz et al., 2012). However, research on early science education is sparse and children's motivational and competence self-concepts are not considered in many learning settings or only play minor roles. This raises multiple questions: How can science competence in specific domains be measured? How do children reason about causalities in the form of stabilities? What intuitive theories do they possess? How can science knowledge and theories be fostered while still maintaining motivational and competence self-concepts? What are influences of interindividual prerequisites such as intelligence? Answers might be achieved through playful interventions that combine children's prior experiences with new information.

To examine young children's development of science knowledge and theories as well as their science self-concept, the following research questions were formulated:

1. Is knowledge about stability a unidimensional construct?
2. How do different forms of play affect children's knowledge and theories about stabilities?
3. What is the effect of prior knowledge or theories on knowledge acquisition or theory adaptation in the science domain of stability?
4. How do different forms of play affect children's motivational and competence self-concepts?

5. Are there reciprocal effects between young children's motivational and competence self-concepts and their stability knowledge acquisition?
6. Is knowledge about stability related to interindividual prerequisites such as intelligence, mental rotation ability, and interest in block play?

To investigate these research questions, three studies with preschool children aged 5 and 6 with a playful intervention, a pre-, post-, and follow-up test were conducted. The pretest was administered approximately two weeks before the one-hour playful intervention and the immediate posttest. The follow-up was conducted approximately ten weeks after the posttest. (1) Children's knowledge about stabilities and (2) children's interest in block play were assessed as group tests; (3) their motivational and competence self-concepts, (4) their mental rotation ability, and (5) their reasoning about stabilities were assessed in single interviews at all three measurement points. Moreover, (6) children's language capacity as an indicator for crystallised intelligence was assessed in a single interview, and (7) figural perception and figural reasoning as indicators of their fluid intelligence were measured as a group test at the pretest only. A transfer test assessing children's mass knowledge was administered as a group test at the follow-up. Testing time at the pretest was approximately 30 minutes for the group tests and 60 minutes for the single interviews; at the post- and follow-up tests, testing time took approximately 20 minutes for the group tests and 30 minutes for the single interviews.

The playful intervention was implemented with three play groups that differed in the support they received, therefore, two guided play groups and a free play group were implemented. To achieve ecological validity, the children played in small groups of children led by one of six female experimenters for approximately 60 minutes. In all three groups, the experimenter praised children's efforts and encouraged them. The children were parallelised into three groups according to their language capacity, resulting in triplets with the same language capacity. For example, a child with a language capacity of $T = 50$ was paired with two other children with a language capacity of $T = 50$ and then each of the children was assigned to one of the play groups.

The first guided play group, the Material group, received material scaffolds in the form of building blocks and photographs of different block constructions, which varied in the number of blocks and their complexity. The blocks varied in shape (cuboids, triangles, and other shapes), size, and colour (brown, black, yellow, red, and green). The material scaffolds were presented to structure the play. Moreover, the scaffolds linked new content, e.g., asymmetrical block structures' stabilities, to children's presumed experiences with block play. Furthermore, asymmetrical as well as symmetrical block structures were used aiming at drawing children's attention to differences in the stability of symmetrical compared to asymmetrical constructions (Hsin & Wu, 2011). The material scaffolds were developed prior to the study and tested in play sessions with children to ensure that children could rebuild the structures shown on the photographs and had fun playing with the materials. In total, the children played five different activities, which are presented in more detail in Article 2, and its supplementary materials[1], as well as in Article 3.

The second guided play group, the Verbal group, received the same material scaffolds and additional verbal scaffolds to promote their mass knowledge, reasoning about stabilities as well as their motivational and competence self-concepts. Mass knowledge and reasoning were fostered through modelling, activating prior knowledge, encouraging comparisons, providing explanations, and asking questions or asking for reasoning; motivational and competence self-concepts were fostered by promoting perceptions of challenge, competence and success (Belland et al., 2013; van de Pol et al., 2010). In both guided play groups, the children were free to play with the blocks they chose and could choose if they wanted to play with a friend or on their own. Moreover, the children could choose to build something entirely different than the constructions presented on the photographs to maintain the playfulness of the situation.

The free play group, Free play group, received unstructured building blocks that came in a large wooden box. The experimenters told the children that they were free to build whatever they liked and could build by themselves or with a friend. The experimenter did not intervene during the play.

The dissertation is based on three articles that have been submitted to or published in peer-reviewed journals. Table 2 provides an overview over the three articles.

The first article *Measuring preschool children's knowledge of the principle of static equilibrium in the context of building blocks: Validation of a test instrument*[2] focuses on the validation of the Centre-of-Mass Test measuring children's knowledge about stability, using item response theory. The article was divided into two studies. In Study 1 the construct structure was tested, and in Study 2, the construct validity of stability knowledge was investigated. The Centre-of-Mass Test's conformity with a 1PL-testlet model with the subtests estimation of stable and unstable constructions was confirmed. Therefore, the knowledge of stability can be construed as a unidimensional construct. Moreover, stability knowledge is related to fluid and crystallised intelligence.

---

[1] All supplementary materials including the tests for stability knowledge and science self-concept can be found in the appendices.

[2] Manuscript submitted to British Journal of Educational Psychology, impact factor: 2.506. Status: Published.

The second article *The impact of a construction play on 5- to 6-year-old children's reasoning about stability*[3] is concerned with children's intuitive theories about stability and the effects of guided and free play, children's prior theories as well as their intelligence on these intuitive theories. Theory theory suggests that children form intuitive theories about their environment that might be adjusted in the face of counterevidence (Gopnik & Wellman, 2012). We investigated which theories young children have about stabilities and if a playful intervention can support children in acquiring mass theory, a more scientifically correct theory than centre theory or theories unconcerned with stabilities (Bonawitz et al., 2012). At pretest, most children explained stabilities by referring to the constructions' centres or reasons that were not concerned with stability. The playful intervention supported children' mass theory acquisition with children in the Verbal group acquiring mass theory most often. The consistency of children's prior theories had an influence on their acquisition of mass theory as well. Children with inconsistent prior theories in the Verbal group acquired a mass theory most often; children with consistent prior theories in the Free play group acquired a mass theory least often of all groups. Fluid as well as crystallised intelligence had a positive effect on children's acquisition of a mass theory. Implications for science education in kindergarten are discussed.

The third article *Construction play promotes change in 5- to 6-year old children's science knowledge about stabilities and science self-concept*[4] focuses on the effect of the playful intervention on children's mass knowledge and motivational and competence self-concepts. Moreover, the reciprocal effects between knowledge acquisition and motivational and competence self-concepts were investigated and differences between the play groups were of interest. The guided play groups gained more mass knowledge than the free play group. Furthermore, motivational and competence self-concepts remained stable in the guided play groups, whereas they declined in the free play group. However, no evidence for reciprocal effects between mass knowledge and motivational or competence self-concept was found. It is concluded that guided play with or without verbal support may be an effective way to foster children's stability knowledge and their science self-concept. Reasons for the missing reciprocal effects are discussed.

This dissertation closes with a general discussion combining the results of the three articles and discussing them in the light of their theoretical background and practical implications.

---

[3] Manuscript submitted to Frontiers in Psychology, impact factor: 2.067. Status: Published.

[4] Manuscript submitted to Child Development, impact factor: 4.891. Status: Submitted.

Table 2.

Overview over the three presented articles.

	Article 1 **Measuring preschoolers' knowledge of static equilibrium**	**Article 2** **Play's impact on children's reasoning**	**Article 3** **Play fosters children's knowledge and self-concept**
Topic	Validation of a test instrument to measure children's stability knowledge	The impact of guided and free play on children's intuitive theories about stability	The effect of guided and free play on children's stability knowledge and science self-concept
Aims	(1) Validating a test instrument to measure children's stability knowledge using item response modelling (2) Investigating construct validity	(1) Investigating children's intuitive theories about stability (2) Examining if guided play with material (+ verbal scaffolds) and free play affect the consistent application of theories (3) Investigating the effect of children's prior theories on their consistent use (4) Investigating the relation of theory development and intelligence	(1) Examining if guided play with material (+ verbal) scaffolds and free play affect stability knowledge and science self-concept (2) Investigating the reciprocal effects between knowledge acquisition and science self-concept
Experimental variables	- A measure for stability knowledge - Children's science self-concept - Fluid and crystallised intelligence - Mental rotation ability - Children's interest in block play	- Three different forms of play: Guided play with material (+ verbal) scaffolds, and free play - Fluid and crystallised intelligence - Children's prior intuitive theories	- Three different forms of play: Guided play with material + verbal scaffolds, material scaffolds, and free play
Outcome variables	- Children's knowledge about stability	- Children's theories about stabilities	- Children's stability knowledge acquisition - Children's science self-concept
Sample	Study 1: 217 preschool children Study 2: 166 preschool children	183 preschool children	183 preschool children

Article 1:

Measuring preschool children's knowledge of the principle of static equilibrium in the context of building blocks: Validation of a test instrument[5]

Anke Maria Weber and Miriam Leuchter
Institute for Children and Youth Education, University of Koblenz-Landau, Germany

Background. Preschoolers' knowledge of the principle of static equilibrium is an important research focus for understanding children's science content knowledge. Hitherto studies have mainly used behavioural observation with small samples. Thus, extending these studies with a validated test instrument is desirable.

Aims. The aim was to validate an instrument (the Centre-of-Mass Test), which is concerned with preschoolers' knowledge of the principle of static equilibrium, using item response theory. In Study 1, the construct structure was tested, and in Study 2, its relation with stabilities of symmetrical blocks, figural reasoning, figural perception, mental rotation, level of interest, self-concept, motivation, and language capacity was investigated.

Samples. A total of 217 5- and 6-year-old children participated in Study 1 and 166 5- and 6-year-old children in Study 2.

Methods. All tests were administered as paper–pencil picture tests in groups and single interviews.

Results. In Study 1, the Centre-of-Mass Test's conformity with a 1PL-testlet model with an overall knowledge of static equilibrium and with two subtests, estimation of stable and unstable constructions, was confirmed. Using a 95% binomial distribution, children were categorized into three knowledge categories: geometrical-centre, centre-of-mass, and undifferentiated knowledge. In Study 2, knowledge of the principle of static equilibrium showed positive correlations with figural perception and reasoning, language capacity, and estimation of the stabilities of symmetrical objects.

Conclusions. The Centre-of-Mass Test measures knowledge of the principle of static equilibrium as a unidimensional construct and mirrors preschoolers' estimations found in previous studies. The acquisition of a more sophisticated static equilibrium knowledge is related to spatial knowledge and language capacity.

---

[5] Weber and Leuchter (2020)

Keywords: development of knowledge of static equilibrium; preschool children; science learning; validation study

1 Introduction

The view on science and technology as part of early education has changed profoundly in recent decades. First, science and technology are considered to be an important field in our technology-based society, and thus, both are core aspects of preschool and primary school curricula (National Research Council, 2012). Second, studies on children's early cognitive development have indicated that young children's learning potential in the field of science has long been underestimated, showing that young children have intuitive but often naïve knowledge about their surrounding environment (Baillargeon, 1994; Gopnik, 2012; Metz, 1995). For example, infants as young as 4 months of age are able to build intuitive knowledge about the continuity and solidity of physical objects (Baillargeon, 1987; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Given these curricular demands, there is growing consensus that science education should start as early as preschool age, which in our country refers to 5- and 6-year-olds in their last year before starting primary school, to facilitate children's science learning later on (Eshach & Fried, 2005; Gelman & Brenneman, 2004; Samarapungavan, Patrick, & Mantzicopoulos, 2011). However, the researchers cited here are not advocating for a formal education, but rather for children learning through play. Nevertheless, the current literature on preschool children's understanding of science is limited. (a) Important contributions have addressed preschoolers' science-related process competencies, such as theory-evidence-coordination or variable control strategy (for an overview see, Kuhn, 2014; Zimmerman, 2007). (b) Lately, researchers have been paying more attention to motivation in learning science (Oppermann, Brunner, Eccles, & Anders, 2017; Samarapungavan et al., 2011). (c) Many studies have examined the developmental aspects of conceptual knowledge of science throughout childhood (for an overview see, Wilkening & Cacchione, 2014). However, a limited number of studies have examined young children's conceptual knowledge with a perspective on education. Most of these studies focus on floating and sinking (Kallery, 2015; Leuchter, Saalbach, & Hardy, 2013; Potvin & Cyr, 2017), while others assess children's concepts of light and shadow (Hsu, Tsai, & Liang, 2011), living things (Opfer & Siegler, 2004), the growth and development of specific species (Samarapungavan et al., 2011), and aspects of the concept of force (Leuchter & Naber, 2019).

However, expanding our understanding of children's science content knowledge lays the groundwork for understanding children's learning and is an important prerequisite for developing preschool interventions and curricula. Therefore, standardised tests in certain scientific domains are needed. Such tests enable us to measure children's naïve concepts and, based on these insights, build a curriculum that transforms those naïve concepts into scientific concepts; thus, the tests are applied both before and after the intervention. While standardised tests are widespread in mathematics and language (e.g., PPVT 4: Dunn et al., 2015; mathematical subtest of the Woodcock Johnson Tests: Mather & Woodcock, 2001), we identify a lack of standardised instruments for testing science content knowledge in different domains. This lack may be due to the broad range of fields and domains of science learning, such as biology, physics, chemistry, etc.

The principle of static equilibrium as a physics concept is interesting for studying children's science content knowledge from two perspectives: (a) It is considered an important specific concept in science and engineering curricula (National Academy of Engineering and National Research Council, 2009). (b) The principle of static equilibrium is a part of children's everyday lives, e.g., when playing with blocks, which is an activity that children spontaneously engage in (Pine & Messer, 2003). To our knowledge, few studies have analysed children's knowledge of the principle of static equilibrium in its relation to block play (e.g., Bonawitz, van Schijndel, Friel, & Schulz, 2012) and have mainly used behavioural observation with small samples. Thus, upscaling first insights into children's knowledge of the static equilibrium is required and an instrument measuring this knowledge has yet to be developed.

1.1 Content considerations for developing an instrument for assessing knowledge of the principle of static equilibrium

The principle of static equilibrium is concerned with forces acting on objects that, as a result, are either at rest or in motion (Riley & Sturges, 1993). The centre of mass corresponds to the geometrical centre of a symmetrical, but not an asymmetrical, object (Figure 1). Focusing on the centre of mass leads to correct estimations of the stabilities of both symmetrical and asymmetrical objects, but it is mandatory for the correct estimation of asymmetrical objects' stabilities. Thus, for a symmetrical object, if more than half of it is supported, it will always remain in place because it is in a state of static equilibrium. However, this principle does not always hold for asymmetrical objects, as their centre of mass has to be supported for them to remain in place. If it is not supported, the object will fall down. Therefore, the correct estimation of the stability of an asymmetrical object indicates an understanding of the principle of static equilibrium, which covers only one physical aspect of balance, such as the estimation of stability, while studies concerned with balance mostly address the multiplicative elements of balance strategies in the context of beam scales (e.g., Siegler, 1976).

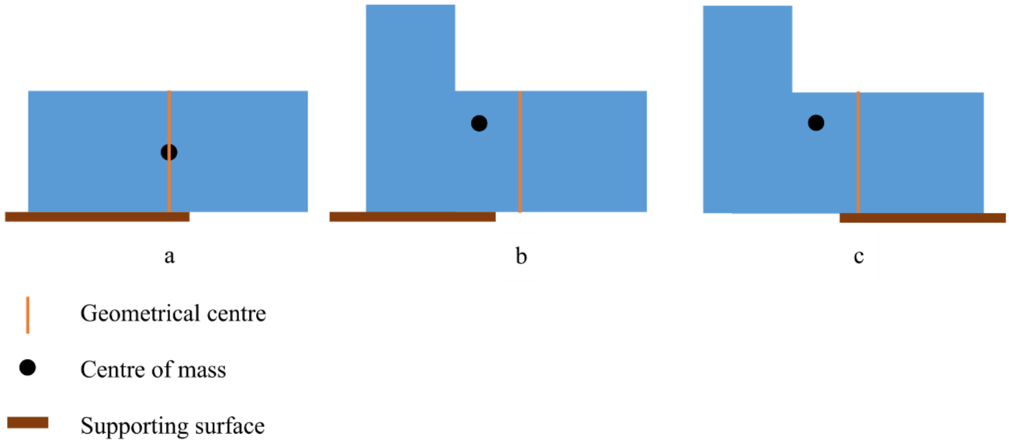


Figure 1. Static equilibrium of symmetrical (A on the left) and asymmetrical objects (B in the middle, C on the right).

Studies from developmental psychology have produced results that can be interpreted as indicators of an understanding of static equilibrium. Thus, considering these studies supports our goal to develop a test for assessing knowledge of static equilibrium in the context of block play. For example, Needham and Baillargeon (1993) showed within the habituation paradigm that 4.5-month-old infants are aware that an unsupported object falls. In another study, Baillargeon, Needham, and Devos (1992) investigated whether infants expect a box to be stable or not depending on contact with a supporting surface. These results suggested that from 6.5 months, infants have a concept of how much contact is needed for the box to be stable.

Siegler (1976) investigated children’s knowledge about balance by putting different weights at different distances on a fulcrum and asking children about the fulcrum’s balance. From 9 years of age, children start to consider the relationship of weight and distance instead of viewing them as separate, indicating a growing knowledge of the principle of static equilibrium. However, Pine, Lufkin, Kirk, and Messer (2007) found that only few preschoolers even consider the weight, much less the distance, of objects on a balance beam.

Krist, Horz, and Schönfeld (2005) asked children between ages 4 and 8 to actively balance symmetrical and asymmetrical objects on a beam scale. In another study, Krist (2010) applied a rating task with photos. In both studies, the researchers found preschool and primary school children’s performance on the estimation of symmetrical objects (e.g., cuboids) to be superior to their estimation of asymmetrical objects (e.g., tetrahedrons), and achievement increased with age (Krist, Atlas, Fischer, & Wiese, 2018). These results suggest that preschool children may disregard the centre of mass and concentrate on the geometrical centre. Moreover, as children’s performance was comparable in all three studies, the results indicate that children’s performance is not influenced by type of assessment.

Bonawitz et al. (2012) also found that children’s use of a centre-of-mass theory increased between ages 4 and 7. Moreover, these authors asked children to balance different symmetrical and asymmetrical objects on a beam scale and categorised the children into three different categories (Karmiloff-Smith and Inhelder, 1974): (1) approximately 1/3 of the children balanced the objects at their geometrical centre, (2) 1/3 balanced the objects according to their centre of mass and thus successfully balanced asymmetrical objects, and (3) 1/3 exhibited an undefined balancing pattern.

However, concerning our goal of measuring children’s knowledge of the principle of static equilibrium, we can identify four restrictions of these studies. First, they mainly aimed at measuring developmental psychological theories such as rule application (Siegler, 1976), representational redescription (Pine & Messer, 2003), and theory theory (Bonawitz et al., 2012); children’s knowledge of the principle of static equilibrium was only measured as a by-product. Therefore, instruments specifically designed to measure the principle of static equilibrium are still missing, and the validity of the measures used for the assessment of the knowledge of the principle of static equilibrium has not been addressed in the studies. Second, the few studies that were concerned with children’s understanding of balance used only single objects (Bonawitz et al., 2012; Krist, 2010). In children’s everyday lives, however, single objects play a minor role compared to block constructions. Third, some of the instruments used in prior studies test aspects of static equilibrium but do not consider all of its features: In Krist’s (2010) test, the stabilities of the symmetrical blocks could be easily rated by considering the amount of contact with the platform, whereas the asymmetrical blocks were supported by exactly half of the platform, making rating their stability with a geometrical-centre knowledge an issue of guessing. Thus, this test cannot provide valid results for measuring knowledge of the principle of static equilibrium. Fourth, most studies on children’s balance concepts were conducted in a one-to-one laboratory environment and coded using observation methods, which are very time-consuming. Laboratory studies have a high internal validity ensured by controlled conditions, which may come at the cost of ecological validity, as the presented materials have little to do with children’s everyday lives.

Thus, a time-efficient test instrument in an ecologically valid context, such as block play, that allows the valid and reliable measurement of children's centre-of-mass knowledge when rating objects' stabilities is needed. Keeping these requirements in mind, we developed a test instrument aiming at measuring children's knowledge of the principle of static equilibrium. Our test is based on block constructions, because children are more familiar with them from their everyday lives. Moreover, the test consists entirely of asymmetrical block constructions that can only be rated correctly by applying centre-of-mass knowledge. In our instrument, all block structures are placed with their centre of mass clearly either on the supporting surface, thus remaining stable, or with their centre of mass beyond the supporting surface, thus tumbling over. The instrument must guarantee that those children using a geometrical-centre approach rate the items incorrectly. Therefore, the geometrical centre of the asymmetrical block structures is clearly supported by a supporting surface if the block structure tumbles (Figure 1, C), but hovers in mid-air if the block structure remains stable (Figure 1, B), owing its stability entirely to the location of its centre of mass. Furthermore, the test duration is only a couple of minutes and since the test is picture based, it can be presented to groups of children.

Thus, in line with Bonawitz et al. (2012) and Karmiloff-Smith and Inhelder (1974), an adequate test should allow the identification of (1) children with a geometrical-centre knowledge, who consider the geometrical centre as a determining factor when rating an object's stability; (2) children with a centre-of-mass knowledge, who consider mass the deciding factor; and (3) children possessing knowledge of neither, who show an undifferentiated pattern and guessing.

1.2 Item structure, dimensionality, and reliability of a test on static equilibrium in the context of block play

Robitzsch (2016) argued that if every item of a test is assumed to measure the construct in question equally well, a 1PL model should be used, and if the testlets are deemed irrelevant to the construct in question, the model may be split into testlets. Knowledge of the principle of static equilibrium was assumed to be a unidimensional construct in former studies (e.g., Bonawitz et al., 2012; Krist, 2010; Siegler, 1976). However, this knowledge might consist of different sub-facets, such as stable and unstable constructions. Thus, the use of a 1PL-testlet model might be an adequate route to take.

1.3 Construct validity

In developing a test for knowledge of the principle of static equilibrium in the context of block play, we must consider its construct validity, which may be examined by assessing its relations with aspects of intelligence related to spatial skills—mental rotation ability, figural perception and figural reasoning—as well as language capacity and geometrical-centre knowledge.

Mental rotation. Mental rotation describes the ability to determine that two pictures represent the same object from two different perspectives (Shepard & Metzler, 1971). Jäncke, Kleinschmidt, Mirzazade, Shah, and Freund (2001) name three prerequisites for mental rotation: (a) the transformation of a visually perceived object into a mental representation, (b) the uncoupling of the spatial features from other sensory impressions, and (c) the processing of the spatial features. The same cognitive process might influence children's knowledge of static equilibrium. To estimate the stability of a structure, children first need to build a mental representation of that structure with its spatial features uncoupled from possibly distracting sensory impressions such as colour. Through processing the spatial features, a child can estimate the stability by taking into account spatial features, such as contact with supporting platform and number of objects on the supporting platform.

Figural perception and figural reasoning. Figural perception and figural reasoning are indicators of fluid intelligence, which is viewed as one of the most important prerequisites for knowledge acquisition (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). Moreover, figural reasoning and figural perception describe an individual's ability to perceive figural content and abstract it (Cattell, 1987; Weiß & Osterland, 2013). These skills may be indicators of children's ability to represent objects mentally and detect underlying (spatial) principles (Weiß & Osterland, 2013). Thus, these skills might be related to the knowledge of static equilibrium because, to estimate stabilities, the figural content of a structure, such as the amount of contact with the supporting platform and the structure's spatial position, must first be perceived. Thereafter, these figural elements must be abstracted, for example, by considering the weight of the individual objects.

Language capacity. Language capacity describes a person's vocabulary and is considered an indicator for crystallised intelligence (Cattell, 1987). Similar to fluid intelligence, crystallised intelligence influences knowledge acquisition (Thorsen, Gustafsson, & Cliffordson, 2014). Furthermore, language capacity is important since questions assessing children's knowledge of static equilibrium often include rather complex phrases such as what happens when…? (Krist, 2010). Children with a high language capacity may thus better understand the task.

Geometrical-centre knowledge. Children with geometrical-centre knowledge will incorrectly estimate the balance of asymmetrical objects but will correctly estimate the balance of symmetrical objects. Children with centre-of-mass knowledge, on the other hand, will correctly assess the balance of both symmetrical and asymmetrical items. Therefore, the assessment of children's ratings of the stabilities of symmetrical objects provides additional information about the tests' validity.

1.4 Self-related correlates

The investigation of the connection of the knowledge of the principle of static equilibrium and self-related constructs such as academic self-concept, motivation, and interest is also informative since other studies have uncovered connections between these self-related constructs and knowledge in different domains (Marsh & Martin, 2011).

Self-concept. Self-concept has been defined as a person's perception of him- or herself that is shaped by the individual's experiences (Shavelson, Hubner, & Stanton, 1976). The reciprocal relationship between self-concept and achievement in a domain has been demonstrated in many studies on mathematics, language, and science (e.g., Jansen, 2017; Marsh & Martin, 2011; Marsh, Xu, & Martin, 2012). An individual uses success and failure as a reference for their expectations of their abilities in a certain domain (Helmke & van Aken, 1995; Juang & Silbereisen, 2002). In the context of building blocks, self-concept relates to, i.e., a child's self-perception of how well he or she builds difficult structures and how much he or she knows about building. However, children with a high self-concept in block building may build with blocks more often and may set themselves tasks that are harder to achieve because a high self-concept leads to increased motivation and effort, which subsequently lead to higher achievement in the domain of interest (Helmke, 1999; Marsh et al., 2012). Therefore, we assume that children with higher self-concepts have acquired more experiences with block building and, through block building, have also acquired more knowledge about balancing relationships and the centre of mass than children with low self-concepts.

Motivation. According to Ryan and Deci (2000), motivation is a driver of behaviour, and even young children differ in their motivations in different domains (Guay et al., 2010). Motivation is also linked to achievement since children who are motivated to learn more about a subject show increased effort and therefore gain more knowledge (Guo, Marsh, Parker, Morin, & Yeung, 2015; Marsh, Trautwein, Ludtke, Koller, & Baumert, 2005). Thus, children who are motivated to learn more about block building might engage in block play more frequently. These children may be persistent when they encounter obstacles, pay increased attention to the tasks at hand, and have an internal drive to understand complex phenomena that they might come across (Oppermann et al., 2017). Through their experiences, these children might have gained knowledge about static equilibrium.

Interest. Interest is defined as voluntarily cognitive and affective engagement with an activity (Renninger & Hidi, 2016). Interest is connected to knowledge acquisition because, on the one hand, interest can develop through engaging in an activity that supports knowledge acquisition, and on the other hand, it can help to gain new knowledge in a domain (Crouch, Wisittanawat, Cai, & Renninger, 2018). The positive relationship between interest and achievement in different domains has been uncovered for older children (Jansen, Lüdtke, & Schroeders, 2016) and for preschool girls' interest in science and their later science performance at the age of 8 (Leibham, Alexander, & Johnson, 2013). Interest in block play might, like self-concept and motivation, be linked to the time spent with building blocks, which might in turn be related to knowledge of static equilibrium.

1.5 Research questions

This paper consists of two studies. In study 1, we analyse the item structure of the Centre-of-Mass Test (COM Test), which measures children's knowledge of static equilibrium, using item response theory. The research questions for this study are as follows: (1) Is knowledge of the principle of static equilibrium a unidimensional construct? (2) Does the factor structure show a good reliability, and thus do all items measure knowledge of the principle of static equilibrium equally well?

Study 2 addresses possible correlates and the construct validity of the COM Test with the research questions: (3) Is the displayed knowledge of static equilibrium positively associated with the symmetrical block structures, aspects of intelligence related to spatial skills, mental rotation, level of interest, self-concept, motivation, and language capacity? (4) How are the classifications, geometrical-centre knowledge, centre-of-mass knowledge and undifferentiated knowledge related to the other constructs? We investigate correlation patterns and group differences to address these questions.

2 Study 1

2.1 Methods

In study 1, we tested the structure of the COM Test, which was constructed and piloted by Leuchter and Plöger (2014).

2.1.1 Participants

In total, 231 children participated in the study. However, 14 children were removed from further analysis because they showed a bias in their response pattern. Thus, 217 children (104 females, age: *M* = 5.52, *SD* = 0.50) remained in the sample, which was obtained via volunteering kindergarten teachers, who helped to contact the children and their parents. All participants were informed about the goals of the study and participated voluntarily and with their parents' consent, which was obtained in written form. Two hundred six children were of European, 1 of Central American, 2 of African, and 9 of Asian descent.

2.1.2 Item construction and selection

Based on Krist (2010), who showed that the presentation form, photos vs. handling, is irrelevant for children's performance, each item consisted of one photo of an asymmetrical block construction (16 dichotomous items in total, eight stable, eight unstable, Figure 2). The aim of study 1 is to examine the item structure of the COM Test.

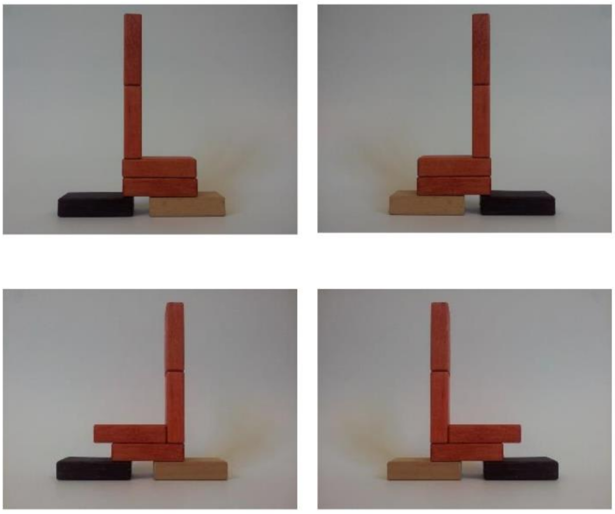


Figure 2. Example items of the COM Test: Unstable 1a and 1b in top row, Stable 3a and 3b in bottom row.

2.1.3 Procedure

Each child completed the COM Test, which was led by a female experimenter, in groups of six children with a paper-pencil procedure in a quiet room at their institution. The children were prevented from copying from each other, either by seating them back to back or by placing a small screen between them. The session started with the experimenter building a stable symmetrical block structure, which was presented on the first page. The experimenter asked the participants for a prediction before pulling the black block away. Would the red blocks fall or remain stable? After that, the experimenter asked the children to draw a circle around the photo to indicate stability or draw an "X" on the photo to indicate instability, based on their estimation. After each participant had decided, the experimenter pulled the black block away, reassuring the children that it did not matter if they had predicted incorrectly. Second, an unstable symmetrical block structure was built. Finally, the experimenter explained that the participants had all done very well, so she did not need to build the following structures shown on the photos, but she asked the participants to decide about each structure's stability quietly so as not to disturb the others.

2.1.4 Statistical analysis

The statistics program R (R Core Team, 2018), version 3.5.1, was used for data analysis using the R packages psych (Revelle, 2018) for the correlation structure, lavaan (Roseel, 2012) for bifactor modelling to assess dimensionality, eRm (Mair, Hatzinger, & Maier, 2018) for the investigation of differential item functioning, TAM (Robitzsch, Kiefer, & Wu, 2018) for item response modelling and WrightMap (Torres Irribarra & Freund, 2014) for the creation of a Wright Map.

2.2 Results

Four percent of the variance could be traced back to group membership, ICC = .04; thus, a second level could be neglected (Cunningham & Huguet, 2012).

Internal consistency was acceptable with $\alpha = .79$. The frequencies of correct answers for each item ranged from 16% to 56% (Table 1). Pearson's ϕ for categorical variables (Table 2) showed a high correlation both within the unstable and within the stable items. Stable and unstable items showed either no correlation or negative correlations, thus knowledge of static equilibrium might possibly be multifaceted.

Table 1

Frequencies of correct answers per item

			Answers	
Item	Picture	∑ Correct	∑ Incorrect	% Correct
Stable 1a		36	181	17
Stable 1b		39	178	18
Stable 2a		56	161	26
Stable 2b		72	145	33
Stable 3a		83	134	38
Stable 3b		66	151	30
Stable 4a		73	144	34
Stable 4b		52	165	24
Unstable 1a		122	95	56
Unstable 1b		109	108	50
Unstable 2a		74	143	34
Unstable 2b		90	127	41
Unstable 3a		106	111	49
Unstable 3b		82	135	38
Unstable 4a		122	95	56
Unstable 4b		96	121	44

The item response modelling used to investigate the first two research questions has three requirements that must be tested in advance: the dimensionality of the construct, the absence of differential item functioning, and local independence (Wells, Rios, & Faulkner-Bond, 2016).

2.2.1 Dimensionality of knowledge of the principle of static equilibrium

To address the dimensionality of the knowledge of static equilibrium, the correlation structure of the test items is investigated, and a corresponding bifactor model is specified. We assume knowledge of the principle of static equilibrium to be unidimensional but do not eliminate the possibility of different testlets. The correlation pattern (Table 2) suggests a stable and an unstable facet of knowledge. Moreover, the two-factor model is theoretically plausible, as it might require different mental processes to rate stable or unstable block constructions.

Yuan-Bentler correction offers a robust solution with sample sizes of approximately $n = 200$ (Li, 2016) and was used for model specification of the bifactor model with the two factors, stable and unstable, and a general statics factor. The Yuan-Bentler corrected bifactor model with uncorrelated factors following Chen, Hayes, Carver, Laurenceau, and Zhang (2012) showed a good fit, $\chi^2 = 120.39$, $df = 88$, $p = .012$, CFI = .95, RMSEA = .04, $p = .791$, SRMR = .06. All items loaded adequately high and significantly on their corresponding latent factor.

For exploratory reasons, we also considered a model with left/right factors, but statistical analysis revealed an unsatisfactory fit, $\chi^2 = 358.76$, $df = 150$, $p < .001$, CFI = .76, RMSEA = .08, $p < .001$, SRMR = .10.

Thus, we cannot assume that a model with left/right factors is relevant, but the results suggest a bifactor model for stable and unstable items. Before testing this model in depth by analysing the fit of the 1PL-testlet model, we will examine differential item functioning according to age and gender as well as local independence of the items (Wells et al., 2016).

2.2.2 Differential item functioning and local independence of the items

The second research question is addressed by specification of a 1PL-testlet model. A prerequisite for such a model is the absence of differential item functioning (DIF) and local independence. DIF inspects test fairness but does not make any statement about group differences on the ability level. An absence of DIF indicates measurement invariance, permitting the assumption that the factor structure is the same for different subgroups. An Anderson likelihood

Table 2

Correlations of the COM Test items

	Stable 1a	Stable 1b	Stable 2a	Stable 2b	Stable 3a	Stable 3b	Stable 4a	Stable 4b	Unstable 1a	Unstable 1b	Unstable 2a	Unstable 2b	Unstable 3a	Unstable 3b	Unstable 4a
Stable 1b	.15*														
Stable 2a	.22**	.30***													
Stable 2b	.16*	.36***	.26***												
Stable 3a	.16*	.30***	.16*	.27***											
Stable 3b	.19**	.29***	.16*	.30***	.26***										
Stable 4a	.10	.28***	−.02	.14*	.38***	.14*									
Stable 4b	.18**	.47***	.29***	.34***	.36***	.36***	.26***								
Unstable 1a	−.08	−.12	−.01	−.01	.03	−.04	.02	−.11							
Unstable 1b	.02	−.04	.10	.00	.08	.00	.14*	.00	.35***						
Unstable 2a	.07	.19**	.11	.15*	−.01	.20**	.08	.19**	.09	.25***					
Unstable 2b	−.02	.00	.00	.08	.11	.03	.19**	.03	.33***	.41***	.32***				
Unstable 3a	−.01	.00	.01	.07	.14*	−.08	.12	.08	.36***	.47***	.31***	.37***			
Unstable 3b	−.12	.01	.08	.10	−.20**	−.08	−.01	.07	.27***	.36***	.32***	.35***	.36***		
Unstable 4a	−.23***	−.02	−.01	−.01	−.05	−.10	.12	.04	.38***	.31***	.22***	.40***	.34***	.36***	
Unstable 4b	−.03	.03	.02	.00	.05	.02	.16*	−.07	.34***	.47***	.35***	.52***	.44***	.37***	.43***

Notes. The correlation coefficient is Pearson's ϕ for categorical variables. *$p < .05$. **$p < .01$. ***$p < .001$.

ratio test was calculated for age and gender. No DIF was found for the age groups or the gender variable.

Local independence was tested using the adjusted form of Yen’s Q3 (Yen, 1984), as specified in the TAM package (Robitzsch, Kiefer et al., 2018), mean aQ3 = .00; item residual correlations ranged from aQ3 = −.30 to aQ3 = .24. However, only six residual correlations reached values higher than the commonly used .20 cut-off, thus supporting the assumption of local independence.

2.2.3 Fit of the 1PL-testlet model

Next, we fitted a 1PL-testlet model with factor correlations set to zero, which showed a good fit for the data, with a stable dimension and an unstable dimension and an overall g factor representing knowledge of static equilibrium, RMSEA = .07, SRMR = .07 (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996). The overall reliability was high with $\omega_t = .89$.

Item difficulties higher than zero indicate an item that is more likely to be solved by a person with a high ability score, difficulties lower than zero indicate an item that is more likely to also be solved by a person with a low ability score. Item difficulties were intermediate to very difficult (Table 3). However, the unstable items were easier for the children to solve, as their difficulty ranged from −0.36 to 1.06, while the stable items all had difficulties higher than zero and ranged from 0.67 to 2.17.

Mean-square residual summary statistics measure the fit discrepancy of items to the model. The outfit is an unweighted measure and equals the average of the standardised residual variance across items and persons. The infit equals the individual variance weighted residuals. Most studies employ values between 0.70 and 1.30 to indicate that an item is productive for the measurement of the latent variable (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008). All items showed a good infit as well as outfit (Table 3).

Figure 3 provides a Wright Map, or person-item map. The plot visualises the person distribution on the latent construct on the left (stable, unstable, and overall knowledge of the static equilibrium from left to right) and the location of the item difficulties on the right side. The Wright Map visualises the higher difficulty of the stable items compared to the unstable ones. However, the items specifically cover the average and high ability spectrum, while children with a low knowledge of the principle of static equilibrium have a small probability of solving any of the items correctly. The differentiation in average and high ability ranges indicates that the instrument can measure learning gains after interventions, thus ensuring that only

Table 3

Item difficulties and fit statistics as estimated by the 1PL-testlet model

Item	Difficulty (1PL-testlet-model)	*SE*	Infit	Outfit
Stable 1a	2.17	.20	1.12	1.16
Stable 1b	2.05	.20	0.87	0.72
Stable 2a	1.45	.18	1.10	1.20
Stable 2b	0.97	.17	1.00	1.03
Stable 3a	0.67	.16	0.97	0.95
Stable 3b	1.14	.17	1.02	1.06
Stable 4a	0.94	.17	1.12	1.15
Stable 4b	1.58	.18	0.87	0.80
Unstable 1a	−0.36	.17	1.12	1.30
Unstable 1b	0.02	.17	0.97	1.02
Unstable 2a	1.06	.18	1.18	1.29
Unstable 2b	0.57	.17	0.93	0.84
Unstable 3a	0.10	.17	0.95	0.94
Unstable 3b	0.81	.17	1.02	1.02
Unstable 4a	−0.36	.17	1.02	1.00
Unstable 4b	−0.33	.17	0.86	0.81

children who have acquired a centre-of-mass knowledge solve most items correctly. Furthermore, the person distribution on both the overall statics factor and the unstable facet is normally distributed, indicating that few children solve the items correctly (centre-of-mass knowledge) or incorrectly (geometrical-centre knowledge), while most are guessing (undifferentiated knowledge). For the stable facet, the person distribution covers only the middle region, indicating that most children have undifferentiated knowledge, and even fewer solve all the items correctly.

Applying a 95% binomial distribution with σ = 1.96 to categorise the children into three categories resulted in classifying children with a total of four or fewer correct answers as applying geometrical-centre knowledge and children with a total of 12 or more correct answers as applying centre-of-mass knowledge. The other children, with a total of between 5 and 11 correct answers were classified as applying undifferentiated knowledge. In total, only twelve children consistently applied centre-of-mass knowledge, 70 applied geometrical-centre knowledge, and 135 children were classified as applying undifferentiated knowledge.

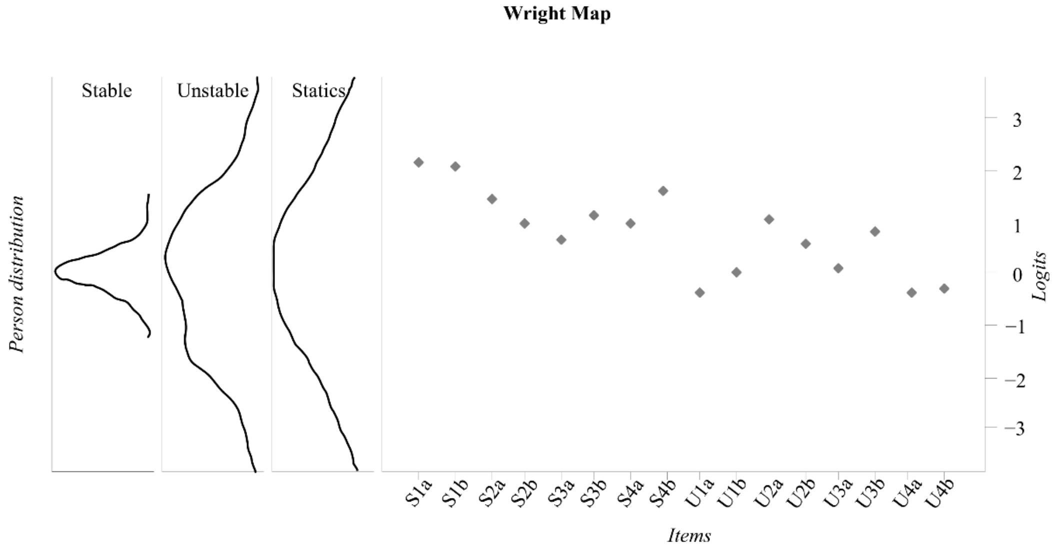


Figure 3. Wright Map on the 16 items of the COM Test and the *N* = 217 children. *Notes*. The left panel of the figure provides the latent abilities estimation of stable and unstable items and the overall knowledge of the static equilibrium from left to right. S1a to S4b = items Stable 1a to Stable 4b. U1a to U4b = items Unstable 1a to Unstable 4b.

2.3 Discussion

The 1PL-testlet model fit the data well, indicating that the test consists of an overall ability and two subtests. Furthermore, the results suggest that unstable items are easier to solve, denoting that children may acquire a concept of tumbling earlier than a concept of a counterweight stabilising an object. Developmentally, this result seems to be plausible because it is more salient to observe items tumbling than objects remaining in place because of the noise a tumbling object makes (Johnson, 2013). Additionally, a child's emotional response to a self-made block building tumbling over (e.g., distress, frustration) is probably more memorable than the reaction to a building remaining stable (pride, joy; Jiang, Waters, Liu, Li, & Yang, 2017).

Bonawitz et al. (2012) suggested three theoretical approaches that children apply to rate the stability of objects: geometrical-centre theory, centre-of-mass theory and no theory. We applied this categorisation to the COM Test, and the 1PL-testlet model showed that most items were too difficult for many of the children; those children could thus be categorised as having geometrical-centre knowledge. Only a small number of children solved most items correctly and thus could be categorised as having centre-of-mass knowledge. Therefore, the test differentiates between the children's theoretical approaches, as can also be seen in the Wright Map. The items needed to be difficult to ensure that only children applying centre-of-mass knowledge consistently could solve a high number of items correctly. As studies have found that most preschoolers have geometrical-centre knowledge or undifferentiated knowledge (Krist, 2010, 2013), the established difficulty of the COM Test is consistent with the results of previous studies, and the test is adequate considering its application for tracing learning gains after interventions. The COM Test has an advantage over behavioural observation because it is less time-consuming and can be conducted in groups. Thus, the results of the item response model have underlined the structure and usefulness of the COM Test for the measurement of knowledge of static equilibrium.

3 Study 2

3.1 Methods

3.1.1 Participants

Of the participants who took part in the first study, 166 preschool children (82 females; age: *M* = 5.52, *SD* = .49) agreed to take part in our second study. Consent was obtained from the children and parents as described in study 1. The sample structure corresponded to study 1.

3.1.2 Measures and procedure

In addition to the COM Test (Study 1), respondents completed the following test instruments assessing possible correlates:

- Geometrical-centre knowledge was assessed by completing the 12 items of Krist's Test (Krist, 2010), $\alpha = .74$, enabling the comparison of children's rating of symmetrical objects to the rating of the asymmetrical structures in the COM Test. The test procedure was similar to that of the COM Test. Each item consisted of a picture of a differently coloured block (e.g., a cuboid) with its geometrical centre either being supported or not supported by a yellow block. The children decided whether the block on each picture was adequately supported.

- A labyrinth subtest and a matrix subtest from the Culture Fair Test (Weiß & Osterland, 2013) were used for the assessment of figural perception and figural reasoning, two aspects of fluid intelligence related with spatial skills in line with the test collection from the SILC ("Spatial Intelligence and Learning Center (SILC)," 2018). Instructions were given according to the CFT 1-R manual (CFT 1-R handbook for more information on the testing parameters).
- Mental rotation ability was assessed with the Picture Rotation Test (PRT; Quaiser-Pohl, 2003), α = .81, range: 0 = all items solved incorrectly to 16 = all items solved correctly. The test consisted of 16 items, each containing four pictures (a reference picture, its correctly rotated version, and two incorrect mirrored and rotated versions). Children were asked to choose the correct picture. This test was conducted in a standardised single-interview procedure according to its manual.
- Interest in block play (Flottmann & Leuchter, unpublished manuscript), α = .76, was assessed with nine picture-based items presented in a small booklet. Two items were dropped from further analyses, as reliability analysis showed that they correlated negatively with the total scale. Each item contained a pair of pictures, one displaying block play, the other a different game, e.g., playing with dolls, or painting. Children chose whether they preferred block play or the other activity by circling the corresponding picture, and received a score ranging from 0 = always preferred the other activity to 7 = always preferred block play.
- Self-concept in building block structures (6 items), α = .79, and motivation for learning about balancing blocks (5 items), α = .75, were assessed with an interview adapted from the Young Children’s Science Motivation (Y-CSM) scale (Oppermann et al., 2017). This test was conducted in a standardised single-interview procedure according to its manual. Self-concept was assessed through asking e.g., “Do you find learning about building with blocks easy or not so easy? Can you show me how easy?”. Motivation was assessed through asking e.g., “Would you like to learn more about building with blocks? Show me how much”. The children received a triangle on a separate sheet of paper, which they could use to point at the corresponding answer, ranging from 0 = not at all to 3 = very much.
- Language capacity was assessed with the German version of the Peabody Picture Vocabulary Test (PPVT 4; Dunn et al., 2015), for which *t*-values are available. A *t*-value of 50 indicates an average language capacity, and smaller and higher values represent lower and higher than average language capacity, respectively. This test was conducted

in a standardised single-interview procedure according to its manual. The instrument consists of 19 sets with 12 items each. Every item consists of four different pictures. The children are given a word, and point to the picture illustrating this word until eight items in one set are answered incorrectly (PPVT 4 handbook for more information on the testing parameters).

The group tests (COM Test, Krist's Test, Interest, CFT 1-R) took about 30 minutes, and the single interview (PRT, Y-CSM, PPVT 4) took about 60 minutes per child. During the testing, breaks were provided whenever the child or the experimenter considered it necessary.

3.1.3 Statistical analysis

The statistics program R (R Core Team, 2018), version 3.5.1, was used for data analysis. The Amelia II package for multiple imputation (Honaker, King, & Blackwell, 2011) helped deal with missing values. Normal distribution is a requirement for multiple imputation and was met by all continuous variables with skew < |2| and kurtosis < |7| (West, Finch, & Churran, 1995). The results of the m = 20 imputed datasets were pooled with the semTools (semTools Contributors, 2016), MKmisc (Kohl, 2018), and miceadds (Robitzsch, Grund, & Henke, 2018) packages. Group comparisons were conducted with Kruskal-Wallis rank sum tests and Dunn's multiple comparison test using the package dunn.test (Dinno, 2017).

3.2 Results

Descriptive results of the test instruments are presented in Table 4. Study 1 showed that knowledge of the principle of static equilibrium can be constructed as a unidimensional construct. The mean of the COM Test was rather low, whereas children performed very well on Krist's Test. For the CFT-1R, comparisons to the expected means were impossible because we assessed only two subtests. The mean values of the PPVT 4 for language capacity and of the PRT for mental rotation comply with the expected means taken from Dunn et al. (2015) and Quaiser-Pohl, Rohe, and Amberger (2010).

To assess the fit of the adapted version of the Y-CSM for the assessment of motivation and self-concept, we ran a Yuan-Bentler corrected CFA with two factors—self-concept and motivation—and fixed factor loadings, which showed a satisfactory fit, $\chi^2 = 72.67$, $df = 52$, $p = .031$, CFI = .94, RMSEA = .06, $p = .062$.

T-tests uncovered gender differences for the COM Test, as the girls reached higher means than the boys, $t(155.93) = 2.19$, $p = .030$, $d = 0.34$, and for Krist's Test, $t(149.70) = -2.03$, $p = .044$, $d = 0.32$, with the boys achieving higher means. Furthermore, we found gender differences in interest, $t(146.25) = -8.78$, $p < .001$, $d = 1.37$, and in academic self-concept, $t(158.26) = -4.32$, $p < .001$, $d = 0.67$, both times in favour of the boys. To investigate possible interaction effects of gender with the aspects of intelligence, mental rotation, interest, self-concept, motivation or language capacity on centre-of-mass knowledge, moderated regression analyses were conducted. These analyses however did not yield any significant results.

Table 4

Pooled descriptive statistics of the tests and their facets

	M (*max*)	*SD*	M_{girls}	SD_{girls}	M_{boys}	SD_{boys}
COM Test overall	5.93 (16)	3.25	6.49	2.98	5.39	3.42
Krist's	10.24 (12)	2.13	9.90	2.33	10.58	1.86
ISpS	11.87 (30)	4.76	11.48	4.92	12.25	4.92
MR	11.46 (16)	3.61	11.36	3.52	11.56	3.52
Interest	3.65 (9)	2.05	2.48	1.40	4.80	1.94
SC	2.18 (3)	0.71	1.95	0.70	2.40	0.64
Mot	2.11 (3)	0.78	2.04	0.81	2.19	0.75
LC	52.49 (73)	8.96	51.61	9.29	53.35	9.29

Notes. ISpS = Intelligence related to spatial skills. MR = Mental Rotation. SC = Self-concept. Mot = Motivation. LC = Language capacity. *max* = maximal possible value of a test. M_{girls} = Mean of female participants on each scale. SD_{girls} = Standard deviation of female participants on each scale. M_{boys} = Mean of male participants on each scale. SD_{boys} = Standard deviation of male participants on each scale.

Again, children were categorised into the three categories, geometrical-centre, centre-of-mass, and undifferentiated knowledge, applying the 95% binomial distribution. Over all data sets, on average, 52 children (18 girls and 34 boys) were categorised as applying geometrical-centre knowledge with four or less correct answers, coded with 1; six children (three girls and three boys) fulfilled the criterion for centre-of-mass knowledge with twelve or more correct answers, coded with 2; and 108 were classified as applying undifferentiated knowledge (61 girls and 47 boys) with five to eleven correct answers, coded with 0. This categorisation was used for the correlations and the Kruskal-Wallis rank sum tests in the following sections.

Correlations are presented in Table 5. The COM Test's theoretical categories correlated positively with Krist's Test, aspects of intelligence, and language capacity. Thus, children who had undifferentiated knowledge performed badly on all of the other three tests. The positive correlation with language capacity might stem from its relation with intelligence.

To further investigate group differences, Kruskal-Wallis rank sum tests were applied for every data set and uncovered differences between the three knowledge classes for the other three tests, Krist’s Test: $H(2) = 10.51$ to $H(2) = 14.85$, $p < .005$; aspects of intelligence: $H(2) = 5.17$ to $H(2) = 9.08$, $p < .05$ for 17 data sets and $p < .07$ for three data sets; and language capacity: $H(2) = 10.19$ to $H(2) = 13.50$, $p < .007$. Further analyses with Bonferroni corrected Dunn’s multiple comparison tests revealed that the children applying geometrical-centre knowledge performed better than the children with undifferentiated knowledge on Krist’s test: $z(2) = 3.22$ to $z(2) = 3.85$, $p < .01$; aspects of intelligence: $z(2) = 2.20$ to $z(2) = 2.93$, $p < .05$; and language capacity: $z(2) = 3.13$ to $z(2) = 3.61$, $p < .01$.

Table 5

Correlations of the test instruments

	COM	Krist’s	ISpS	MR	Interest	SC	Mot
Krist’s	.25**						
ISpS	.19*	.22**					
MR	.01	.18*	.47***				
Interest	.12	.13	.04	.14			
SC	−.04	.06	−.06	−.12	.24**		
Mot	−.07	.09	.08	.00	.08	.51***	
LC	.16*	.15	.22**	.16*	.07	−.03	−.05

Notes. ISpS = Intelligence related to spatial skills. MR = Mental rotation. SC = Self-concept. Mot = Motivation. LC = Language capacity. **p* < .05. ***p* < .01. ****p* < .001.

3.3 Discussion

The second study was conducted to inspect the construct validity of the COM Test by investigating its relation with other constructs and possible group differences implied by the categorisation applied by Bonawitz et al. (2012). We expected positive relations between centre-of-mass knowledge and all other constructs and found them for Krist’s Test, and the aspects of intelligence, and language capacity. For Krist’s Test, children with geometrical-centre or centre-of-mass knowledge solved symmetrical items correctly. Furthermore, children with higher figural perception and figural reasoning skills and a higher language capacity were more likely to have either geometrical-centre knowledge or centre-of-mass knowledge. Jirout and

Newcombe (2015) found that spatial skills correlated with time spent playing with blocks. This result could explain the correlation found between aspects of intelligence and geometrical-centre or centre-of-mass knowledge. Unfortunately, we did not assess children's frequency of block play. This could be investigated in future studies.

However, there were no correlations with the other constructs. For mental rotation, we used a general test, whereas Casey et al. (2008) used a test with block constructions. Therefore, those authors probably found a correlation, whereas we did not. This difference can be investigated in future studies by applying both a specific and a general test for mental rotation. For interest and motivation, it can be assumed that children's prior experiences had little to do with knowledge of the principle of static equilibrium. For self-concept, children's self-concepts at a very young age are mostly incoherent, and they thus tend to overestimate themselves with little regard to their actual abilities (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Helmke, 1999).

Concerning the gender differences found for the COM Test, Krist's Test, interest, and self-concept, gender stereotypes come to mind (Desouza & Czerniak, 2009; Farrell, 1957; Lepola, Vauras, & Mäki, 2000; Saracho, 1994).

4 General discussion

The purpose of these studies was to take some initial steps towards validating a test instrument for measuring children's knowledge of the principle of static equilibrium. To achieve this goal, we conducted two studies, which, to our knowledge, were the first to systematically investigate children's knowledge of the principle of static equilibrium by examining its dimensionality, a standardised way of measurement, and its relation with other constructs. Study 1 revealed that the knowledge of the principle of static equilibrium can be constructed as a unidimensional construct with two subtests measuring children's knowledge of stability and instability. A 1PL-testlet model showed that all items measure knowledge of static equilibrium equally well. The unstable items are mediocre to difficult, while the stable items are difficult to very difficult, hinting at possible differences in acquisition of the concept of stability vs. instability. Thus, the COM Test can be used to investigate children's knowledge about balancing objects.

Our study contributes to understanding aspects of children's science content knowledge, specifically static equilibrium. This is a prerequisite for fostering children's understanding of science, which in itself is important (National Research Council, 2012). In line with prior research, our first study confirms that most children had undifferentiated knowledge or applied the geometrical-centre for balancing objects, with a standardised test underlining earlier findings from studies using classification methods by observation (Bonawitz et al., 2012; Krist, 2010; Siegler, 1976).

More precisely, in comparison with the results of Bonawitz and her colleagues, more children in the present studies had undifferentiated knowledge, while fewer children had geometrical-centre or centre-of-mass knowledge. One reason for this difference is that the children in our study were younger, 5 to 6 years old, while Bonawitz and her colleagues investigated 7-year-olds as well, who applied a centre-of-mass knowledge more often than the 6-year-olds did.

In study 2, we investigated knowledge of static equilibrium in relation to other constructs. Centre-of-mass knowledge was positively related to three aspects of intelligence, but we could not identify any relationship with mental rotation. Moreover, the correlation with the estimation of symmetrical objects' stabilities was positive as well.

Both crystallised and fluid intelligence were related with knowledge acquisition in the statics domain even for very young children. The positive correlation between crystallised intelligence, indicated by children's language capacity, and centre-of-mass knowledge illustrates that children with higher crystallised intelligence have more centre-of-mass knowledge (Thorsen, Gustafsson, & Cliffordson, 2014). However, the correlation also indicates that children with higher language capacity might have understood the instruction better. The positive relationship of figural perception and figural reasoning (both indicators for fluid intelligence) with centre-of-mass knowledge indicates that the skills to perceive and abstract figural content may be an underlying process for correctly estimating stability, such as position and amount of contact with the supporting surface (Weiß & Osterland, 2013). Nonetheless, the ability to mentally rotate an object does not seem to contribute to centre-of-mass knowledge, indicating that mentally representing the object from different perspectives might not be important. Furthermore, the positive relationship between geometrical-centre knowledge (evaluated with Krist's test) and centre-of-mass knowledge (assessed with the COM Test) underlines the validity of the COM test, as Krist (2010) claimed that geometrical-centre knowledge develops first.

Moreover, neither self-concept, or interest, nor motivation contributed to children's centre-of-mass knowledge. The lack of relation between self-concept, interest, or motivation and knowledge of the principle of static equilibrium may be due to the lack of prior experiences of children with the centre of mass. Because of these missing experiences, interventions targeting new experiences regarding the centre of mass are needed. In the domain of the principle of static equilibrium, such interventions could include the building of asymmetrical block constructions of varying difficulty, which could then be evaluated regarding their influence on children's centre-of-mass knowledge.

Taking a closer look at the test method, in contrast to Bonawitz et al. (2012) and Siegler (1976), we used a photo-based paper-pencil procedure instead of a resource-consuming behavioural observation. We chose this test format because Krist (2010) found that representation form is irrelevant to children's estimation, and it is thus irrelevant whether a child balances the blocks or views pictures of block structures.

We decided to develop the COM Test because in contrast to Krist's Test (Krist, 2010), the COM Test consists of asymmetrical block constructions, which can only be rated correctly with centre-of-mass knowledge because the weight has to be considered. In contrast, Krist's Test consists of symmetrical as well as asymmetrical items. The symmetrical items can be solved correctly by applying either geometrical-centre or centre-of-mass knowledge, as considering the geometrical centre is sufficient for correctly estimating the stability. However, the asymmetrical items are ambiguous because the block is placed on the supporting surface right at its middle. The items can be solved correctly by applying centre-of-mass knowledge, but children with undifferentiated knowledge as well as children with geometrical-centre knowledge can guess and will have a 50% chance of solving the items correctly. Therefore, Krist's Test does not adequately differentiate between the theoretical approaches. The COM Test fills this gap.

Another advantage of the COM Test is that it measures the principle of static equilibrium in a standardised way. Furthermore, the COM Test is time-saving, as it can be filled out by a group of children in a matter of minutes, rendering complex single-interview or observation procedures in a laboratory unnecessary. Since children are familiar with block constructions from their everyday lives, the test also ensures ecological validity.

The use of block constructions worked well for the preschoolers, as they had easy access to block play (Ferrara, Hirsh-Pasek, Newcombe, Golinkoff, & Lam, 2011). Furthermore, the COM Test can be used to measure children's learning gains after interventions targeting geometrical-centre or centre-of-mass knowledge since the test measures both, and thus developmental patterns and learning gains can be measured. Furthermore, the findings about children's different theoretical approaches can help to develop these new interventions and curricula, which can be used to support content knowledge in the statics domain.

Regarding the limitations of the studies, to further validate the COM Test, children's reasoning about building blocks should be assessed with interviews. Furthermore, to extend our findings concerning the relation between figural perception and figural reasoning and geometrical-centre and centre-of-mass knowledge, children's time spent playing with blocks should be assessed as a possible moderator of this relationship (Jirout & Newcombe, 2015). Following Casey et al. (2008) and Verdine, Irwin, Golinkoff, and Hirsh-Pasek (2014), relations between the COM Test and mental rotation measured with building blocks, as well as math abilities, should be investigated. More age groups could be tested to validate the construct structure in different age groups and find out more about developmental patterns. Additionally, the investigation of older children could help replicate Bonawitz et al. (2012).

Learning more about children's scientific knowledge of the principle of static equilibrium and related learning processes remains an interesting and important research focus, with many processes still to be uncovered. The investigation of content knowledge and underlying theoretical approaches remains crucial to the design of state-of-the-art tests and, based on those tests, curricula for kindergartens and other schools.

5 Acknowledgments

We would like to thank the German Research Foundation (DFG) for sponsoring the BauSpiel project. Furthermore, we would like to thank all children who participated in our study.

6 References

Baillargeon, R. (1987). Object permanence in 3½- and 4½-month-old infants. *Developmental Psychology, 23*(5), 655–664. https://doi.org/10.1037/0012-1649.23.5.655

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science*, *3*(5), 133–140. https://doi.org/10.1111/1467-8721.ep10770614

Baillargeon, R., Needham, A., & Devos, J. (1992). The development of young infants' intuitions about support. *Early Development and Parenting, 1*(2), 69–78. https://doi.org/10.1002/edp.2430010203

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology, 64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction, 26*(3), 269–309. https://doi.org/10.1080/07370000802177177

Cattell, R. B. (1987). Intelligence: It's structure, growth and action. Amsterdam: Elsevier.

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

Crouch, C. H., Wisittanawat, P., Cai, M., & Renninger, K. A. (2018). Life science students' attitudes, interest, and performance in introductory physics for life sciences: An exploratory study. *Physical Review Physics Education Research, 14*(1). https://doi.org/10.1103/PhysRevPhysEducRes.14.010111

Cunningham, S. D., & Huguet, N. (2012). Weighting and complex sampling design adjustments in longitudinal studies. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis* (pp. 43–69). Taylor & Francis.

Desouza, J. M. S., & Czerniak, C. M. (2009). Social behaviors and gender differences among preschoolers: Implications for science activities. *Journal of Research in Childhood Education*, *16*(2), 175–188. https://doi.org/10.1080/02568540209594983

Dinno, A. (2017). *dunn.test: Dunn's test of multiple comparisons using rank sums*. R package version 1.3.5 [Computer software]. Retrieved from https://CRAN.R-project.org/package=dunn.test

Dunn, L. M., Dunn, D. M., Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *PPVT 4. Peabody Picture Vocabulary Test*. Frankfurt: Pearson Assessment.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64*(3), 830. https://doi.org/10.2307/1131221

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology*, *14*(3), 315–336. https://doi.org/10.1007/s10956-005-7198-9

Farrell, M. (1957). Gender differences in block play in early childhood education. *Journal of Educational Research, 51*, 279–284.

Flottmann, J., & Leuchter, M. (2017). *Das Verständnis junger Kinder für das Konzept der potenziellen Energie – zum Zusammenhang von inhaltsbezogenem Wissen und Versuchsplanung. [Young children's understanding of the concept of potential energy – of the relation between content knowledge and experiment design.]*. WUM, Münster, Germany.

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*(1), 150–158. https://doi.org/10.1016/j.ecresq.2004.01.009

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*(6102), 1623–1627. https://doi.org/10.1126/ science.1223416

Guay, F., Chanal, J., Ratelle, C. F., Marsh, H. W., Larose, S., & Boivin, M. (2010). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *British Journal of Educational Psychology*, *80*(4), 711–735. https://doi.org/10.1348/000709910X499084

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, *37*, 161–168. https://doi.org/10.1016/j.lindif.2015.01.008

Helmke, A. (1999). From optimism to realism? Development of children's academic self-concept from kindergarten to grade 6. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 198–221). Cambridge University Press.

Helmke, A., & van Aken, M. A. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, *87*(4), 624–637.

Hsu, C.-Y., Tsai, C.-C., & Liang, J.-C. (2011). Facilitating preschoolers' scientific knowledge construction via computer games regarding light and shadow: The effect of the prediction-observation-explanation (POE) strategy. *Journal of Science Education and Technology*, *20*(5), 482–493. https://doi.org/10.1007/s10956-011-9298-z

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jäncke, L., Kleinschmidt, A., Mirzazade, S., Shah, N. J., & Freund, H.-J. (2001). The role of the inferior parietal cortex in linking the tactile perception and manual construction of object shapes. *Cerebral Cortex*, *11*(2), 114–121. https://doi.org/10.1093/cercor/11.2.114

Jansen, M. (2017). Academic self-concept in the sciences: Domain-specific differentiation, gender differences, and dimensional comparison effects. In F. Guay, H. W. Marsh, D. M. McInerney, & R. G. Craven (Eds.), *International advances in self research: Vol. 6. Self. Driving positive psychology and well-being* (pp. 71–112). Information Age Publishing.

Jansen, M., Lüdtke, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology*, *46*, 116–127. https://doi.org/10.1016/j.cedpsych.2016.05.004

Jiang, Z., Waters, A. C., Liu, Y., Li, W., & Yang, L. (2017). Event-related theta oscillatory substrates for facilitation and interference effects of negative emotion on children's cognition. *International Journal of Psychophysiology, 116*, 26–31. https://doi.org/10.1016/j.ijpsycho.2017.02.012

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. Sample. *Psychological Science*, *26*(3), 302–310. https://doi.org/10.1177/0956797614563338

Johnson, S. P. (2013). Object perception. In P. D. Zelazo (Ed.), *Oxford library of psychology: Vol. 1. Body and mind* (pp. 371–379). Oxford Univ. Press.

Juang, L. P., & Silbereisen, R. K. (2002). The relationship between adolescent academic capability beliefs, parenting and school grades. *Journal of Adolescence*, *25*(1), 3–18. https://doi.org/10.1006/jado.2001.0445

Kallery, M. (2015). Science in early years education: Introducing floating and sinking as a property of matter. *International Journal of Early Years Education*, *23*(1), 31–53. https://doi.org/10.1080/09669760.2014.999646

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, *3*(3), 195–212. https://doi.org/10.1016/0010-0277(74)90008-0

Kohl, M. (2018). *MKmisc: Miscellaneous functions from M. Kohl*. R package version 1.0 [Computer software]. Retrieved from http://www.stamats.de

Krist, H. (2010). Development of intuitions about support beyond infancy. *Developmental Psychology, 46*(1), 266–278. https://doi.org/10.1037/a0018040

Krist, H. (2013). Development of intuitive statics. *Zeitschrift Für Entwicklungspsychologie und Pädagogische Psychologie, 45*(1), 27–33. https://doi.org/10.1026/0049-8637/a000078

Krist, H., Atlas, C., Fischer, H., & Wiese, C. (2018). Development of basic intuitions about physical support during early childhood: Evidence from a novel eye-tracking paradigm. *Quarterly Journal of Experimental Psychology, 2,* 174702181773719. https://doi.org/10.1177/1747021817737196

Krist, H., Horz, H., & Schönfeld, T. (2005). Children’s block balancing revisited: No evidence for representational redescription. *Swiss Journal of Psychology, 64*(3), 183–193. https://doi.org/10.1024/1421-0185.64.3.183

Kuhn, D. (2014). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 497–523). Wiley-Blackwell.

Leibham, M. B., Alexander, J. M., & Johnson, K. (2013). Science interests in preschool boys and girls: Relations to later self-concept and science achievement. *Science Education, 97,* 574–593.

Lepola, J., Vauras, M., & Mäki, H. (2000). Gender differences in the development of academic self-concept of attainment from the 2nd to the 6th grade: Relations with achievement and perceived motivational orientation. *Hellenic Journal of Psychology, 7*(3), 290–308.

Leuchter, M., & Naber, B. (2019). Studying children's knowledge base of one-sided levers as force amplifiers. *Journal of Research in Science Teaching, 56*(1), 91–112. https://doi.org/10.1002/tea.21470

Leuchter, M., & Plöger, I. (2014). *Einschätzungen der Stabilität zusammengesetzter Bauklotz-Anordnungen 6-jähriger Kinder [Estimation of constructed block arrangements‘ stabilies by six-year-old children]*. Paper presented at the GSF conference, Leipzig, Germany.

Leuchter, M., Saalbach, H., & Hardy, I. (2013). Designing science learning in the first years of schooling. An intervention study with sequenced learning material on the topic of ‘floating and sinking’. *International Journal of Science Education, 36*(10), 1751–1771. https://doi.org/10.1080/09500693.2013.878482

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Mair, P., Hatzinger, R., & Maier, M. J. (2018*). eRm: Extended Rasch Modeling*. 0.16-0. [Computer software]. Retrieved from http://erm.r-forge.r-project.org/

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook: Vol. 1. Theories, Constructs, and Critical Issues* (pp. 427–458). https://doi.org/10.1037/13273-015

Marsh, H. W., Trautwein, U., Ludtke, O., Koller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Mather N., & Woodcock R. (2001). *Examiner's manual, Woodcock–Johnson Achievement Test–III*. Riverside.

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, *65*(2), 93–127. https://doi.org/10.3102/0034654306

National Academy of Engineering and National Research Council (2009). *Engineering in K-12 education*. National Academies Press. https://doi.org/10.17226/12635

National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, *47*, 121–148.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*(4), 301–332. https://doi.org/10.1016/j.cogpsych.2004.01.002

Oppermann, E., Brunner, M., Eccles, J. S., & Anders, Y. (2017). Uncovering young children's motivational beliefs about learning science. *Journal of Research in Science Teaching*, *24*(2), 195. https://doi.org/10.1002/tea.21424

Pine, K. J., Lufkin, N., Kirk, E., & Messer, D. (2007). A microgenetic analysis of the relationship between speech and gesture in children: Evidence for semantic and temporal asynchrony. *Language and Cognitive Processes*, *22*(2), 234–246. https://doi.org/10.1080/01690960600630881

Pine, K. J., & Messer, D. (2003). The development of representations as children learn about balancing. *British Journal of Developmental Psychology*, *21*, 285–301.

Potvin, P., & Cyr, G. (2017). Toward a durable prevalence of scientific conceptions: Tracking the effects of two interfering misconceptions about buoyancy from preschoolers to science teachers. *Journal of Research in Science Teaching*, *54*(9), 1121–1142. https://doi.org/10.1002/tea.21396

Quaiser-Pohl, C. (2003). The mental-cutting test "Schnitte" and the Picture Rotation Test – Two new measures to assess spatialability. *International Journal of Testing*, *3*, 219–231.

Quaiser-Pohl, C., Rohe, A. M., & Amberger, T. (2010). The solution strategy as an indicator of the developmental stage of preschool children's mental-rotation ability. *Journal of Individual Differences*, *31*(2), 95–100. https://doi.org/10.1027/1614-0001/a000017

R Core Team (2018). R: *A language and environment for statistical computing*. [Computer software]. Vienna, Austria. Retrieved from URL https://www.R-project.org/.

Renninger, K. A., & Hidi, S. (2016). *The power of interest for motivation and engagement*. Routledge.

Revelle, W. (2018). *psych: Procedures for psychological, psychometric, and personality research* (Version R package version 1.8.3) [Computer software]. Northwestern University, Evanston (Illinois).

Riley, W. F., & Sturges, L. D. (1993). *Engineering mechanics: Statics* (2nd ed.). Wiley.

Robitzsch, A. (2016). *Essays zu methodischen Herausforderungen im Large-Scale Assessment [Essays concerning methodological challenges in large-scale-assessment]*. Humboldt-Universität zu Berlin, Berlin.

Robitzsch, A., Grund, S., & Henke, T. (2018). *miceadds: Some additional multiple imputation functions, especially for mice*. (Version R package version 2.13-63) [Computer software]. Retrieved from https://CRAN.R-project.org/package=miceadds

Robitzsch, A., Kiefer, T., & Wu, M. (2018). *TAM: Test analysis modules*. (Version R package version 2.9-35) [Computer software]. Retrieved from https://CRAN.R-project.org/package=TAM

Roseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Samarapungavan, A., Patrick, H., & Mantzicopoulos, P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cognition and Instruction*, *29*(4), 416–470. https://doi.org/10.1080/07370008.2011.608027

Saracho, O. N. (1994). The relationship of preschool children's cognitive style to their play preferences. *Early Child Development and Care*, *97*, 21–33.

SemTools Contributors (2016). *semTools: Useful tools for structural equation modeling. R package version 0.4-14* [Computer software]. Retrieved from https://CRAN.R-project.org/package=semTools.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Siegler, R. S. [R. S.] (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481–520.

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, *8*, 33. https://doi.org/10.1186/1471-2288-8-33

Spatial Intelligence and Learning Center (SILC) (2018). Retrieved from https://spatiallearning.org/

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605–632. https://doi.org/10.1037/0033-295X.99.4.605

Thorsen, C., Gustafsson, J.-E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology*, *84*(4), 556–570. https://doi.org/10.1111/bjep.12041

Torres Irribarra, D., & Freund, R. (2014). *Wright Map: IRT item-person map with ConQuest* [Computer software]. Retrieved from http://github.com/david-ti/wrightmap

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology*, *126*, 37–51. https://doi.org/10.1016/j.jecp.2014.02.012

Weiß, R. H., & Osterland, J. (2013). *CFT 1-R*. Hogrefe.

Wells, C. S., Rios, J., & Faulkner-Bond, M. (2016). Testing assumptions of item response theory models. In C. S. Wells & M. Faulkner-Bond (Eds.), *Educational measurement: From foundations to future* (pp. 162–181). Guilford Press.

West, S. G., Finch, J. F., & Churran, P. J. (1995). Structural equation models with nonnormal variables: problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56–75). Sage.

Wilkening, F., & Cacchione, T. (2014). Children’s intuitive physics. In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 473–496). Wiley-Blackwell.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological*, *8*(2), 125–145.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001

Article 2:

The impact of a construction play on 5- to 6-year-old children's reasoning about stability[6]

Anke Maria Weber, Timo Reuter, and Miriam Leuchter
Institute for Children and Youth Education, University of Koblenz-Landau, Germany

Theory. Young children have an understanding of basic science concepts such as stability, yet their theoretical assumptions are often not concerned with stability. The literature on theory theory and theory-evidence coordination suggests that children construct intuitive theories about their environment which can be adjusted in the face of counterevidence that cannot be assimilated into the prior theory. With increasing age, children acquire a Center theory when balancing objects and try to balance every object at their middle, succeeding with symmetrical objects. Later, they acquire the basic science concept of stability through learning that the weight distribution of an object is of importance. Thus, they acquire a Mass theory and succeed in balancing asymmetrical objects as well. Fluid and crystallized intelligence might contribute to children's acquisition of Mass theory. Moreover, their Mass theory might be supported by implementing a playful intervention including (a) material scaffolds and (b) verbal scaffolds.

Aims. We investigated which theories children have about stability and whether these theories can be adjusted to Mass theory by implementing a playful intervention.

Method. A total of 183 5- to 6-year-old children took part in the study with a pre-post-follow-up intervention design. Children's Mass theory was assessed with an interview in which children explained constructions' stabilities. The children received a playful intervention with two differing degrees of scaffolding (material scaffolds or material + verbal scaffolds) or no scaffolding.

Results. At first few children used a Mass theory to explain their reasoning. However, after being confronted with counterevidence for the asymmetrical constructions, children changed their explanation and applied a Mass theory. More children

---

[6] Weber et al. (2020). The article was submitted to an American journal and is written in American English. The supplementary materials are presented in Appendices A to G.

in the play group with the highest degree of scaffolding, i.e., material + verbal scaffolds, acquired a Mass theory compared to the other groups. Fluid as well as crystallized intelligence contributed to children's acquisition of a Mass theory.

Discussion. Counterevidence can support children in their acquisition of a Mass theory. A playful intervention with scaffolding supports children even more.

Keywords: guided play; theory theory; theory-evidence coordination; free play; science learning; intelligence

1 Contribution to the field

Early science learning lays a foundation for the understanding of complex science concepts, which is relevant throughout school and later in life. Young children construct intuitive theories about science phenomena that they can adjust when they are confronted with evidence not compliant with their theory. In the science domain of stability, young children often explain stabilities with the construction's middle or with a theory not concerned with stability and neglect the weight. The question of how to best support such theory adjustments remains unanswered. Drawing on prior work on playful learning, we investigated whether guided block play with varying degrees of an adult's support enhanced children's learning compared to free block play in the science domain of stability. Children who received the most support were most likely to adjust their intuitive theories and explain stabilities with a construction's weight distribution. Intelligence contributed to children's learning in guided play as well as in free play. We make the case that playful learning in the form of guided play supports young children's science learning.

2 Introduction: Science education

Early knowledge acquisition in the science domain and scientific reasoning lay the foundation for the understanding of complex science concepts, which are relevant throughout the school years and in later life (Eshach & Fried, 2005; Trundle & Saçkes, 2012). Accordingly, studies have demonstrated that science learning and scientific reasoning can be promoted in the early years of childhood (Akerson et al., 2011; Cremin et al., 2015; Gelman & Brenneman, 2004; Klahr et al., 2011).

Children construct intuitive theories to explain what is happening around them, and adjust these theories continuously (Gopnik & Wellman, 2012). These theories encompass science concepts, which might be altered by confronting children with counterevidence (Bonawitz et al., 2012). Promoting young children's science learning aims at helping them adjust their theories and should consider children's developmental constraints by considering children's everyday activities, e.g., their play (Copple & Bredenkamp, 2009). One possibility for such science-related play could be construction play in the form of block play, which is an important leisure activity for young children (Borriello & Liben, 2018; Pellegrini & Gustafson, 2005; Rubin et al., 1978; Verdine et al., 2019). An adult's guidance might be integrated into children's play in the form of scaffolding, which might support children's science learning (Klahr et al., 2011; Mantzicopoulos et al., 2013; van de Pol et al., 2010).

Therefore, children's science theories might be supported through science-related play that focuses on children's experiences and encompasses an adult's scaffolding.

2.1 Theory theory, Bayesian inference, and theory-evidence coordination

Fostering scientific reasoning is one goal of science education (Chin, 2007; Klahr et al., 2011). Studies on children's scientific reasoning rely on literature concerned with theory theory, Bayesian inference and theory-evidence coordination. Theory theory is concerned with the adjustment of children's intuitive theories when they are confronted with evidence. Bayesian inference focuses on the interaction of intuitive theories with evidence, while theory-evidence coordination investigates the conditions under which children can interpret evidence.

According to theory theory, children construct intuitive theories about their environment, which have similarities with scientific theories (Gopnik & Wellman, 2012). Intuitive and scientific theories share at least five characteristics. (1) They encompass causal representations of the surrounding world, (2) may be hierarchically organized, (3) provide possible explanations for regularities, (4) allow predictions of regularities, and (5) can be adjusted in the face of counterevidence. In this process, not only the explanations for certain subordinate relations but also the general assumptions about regularities can change (Bonawitz et al., 2012; Gopnik & Wellman, 2012). Such adjustment occurs, often gradually, if a child is confronted with counterevidence that cannot be assimilated into their prior theory, either through their own interventions, e.g., in their play, or through observing the interventions of others (Gopnik & Wellman, 2012). According to Gopnik and Meltzoff (1997), children pass through the same developmental processes and therefore have similar representations about the same objects at roughly similar times in their lives. Most children do not test their theories with experiments but adjust them when they are confronted with evidence (Gopnik, 2013). Research has provided support for theory theory, inter alia, in the domain of balance (Bonawitz et al., 2012) and biology (Schulz et al., 2007).

Researchers have applied Bayesian inference to the theory theory framework to value the role of probability and prior knowledge on learning processes (Gopnik, 2013; Gopnik & Wellman, 2012; Schulz et al., 2007). Bayesian inference indicates how a learner changes their theory after being confronted with a set of evidence and how children might combine theory and evidence (Schulz et al., 2007). Bayesian models consider prior knowledge and its effect on inductive reasoning as well as how much a person believes one theory to be true. Furthermore, prior beliefs and evidence might interact, e.g., a child might interpret data according to their prior beliefs and dismiss counterevidence (Gopnik, 2013; Griffiths et al., 2011; Hawkins et al., 1984). Children with consistent presumptions will likely change their theory less easily than children with inconsistent presumptions (Gopnik et al., 2001; Griffiths et al., 2011; Sobel et al., 2004).

Studies on theory-evidence coordination have found that young children often face problems relating a theory with evidence (Koerber et al., 2005; Piekny & Maehler, 2013; Ruffman et al., 1993), which seemingly contradicts the results of studies on theory theory (Bonawitz et al., 2012; Schulz et al., 2007). However, taking a closer look at the studies on theory-evidence coordination, these studies showed that young children primarily face problems when evidence was presented in the form of imperfect covariation. Children were more likely to successfully coordinate theory with evidence when the evidence was presented in the form of perfect covariation, which is how the studies on theory theory presented evidence. For example, Sobel et al. (2004) found that children were even able to make inferences from indirect evidence of perfect covariation in the form of data they had not directly observed.

In conclusion, at least three factors might contribute to young children’s science learning with regard to their developmental constraints and should be considered. (1) Children can interpret perfect covariation but face problems with imperfect covariation (Koerber et al., 2005). Therefore, evidence should be presented in the form of perfect covariation. (2) Children have prior theories about science phenomena and often have similar theories at a certain age (Gopnik & Meltzoff, 1997). Therefore, these prior theories should be considered so that the presented evidence relates to children’s intuitive theories. (3) Children with consistent prior theories might need to see counterevidence repeatedly to adjust their theory because this adjustment often happens gradually (Gopnik & Wellman, 2012). Therefore, children should receive enough time to deal with the phenomenon.

The question remains how best to confront young children with evidence relating to their science theories. Children’s developmental constraints can be addressed by allowing for activities that occur in their everyday lives, e.g., play (Copple & Bredenkamp, 2009). Moreover, play might be enriched by scaffolding materials as well as an adult's verbal support (Zosh et al., 2018).

2.2 Material and verbal scaffolds in guided play

Play-based learning is the mandated pedagogy in early years' curricula in many countries (Hirsh-Pasek et al., 2009; Pyle et al., 2017) and is regarded as developmentally appropriate practice (Copple & Bredenkamp, 2009). Researchers have widely agreed that play can be characterized as a voluntary, intrinsically motivated, child-directed, and process- rather than goal-oriented behavior that contains elements of choice (Daubert et al., 2018; Pellegrini, 2013; Rubin et al., 1983; Trawick-Smith, 2015; Weisberg et al., 2013).

Zosh et al. (2018) define play as a spectrum that allows for different *types of play*, ranging from *free play* as voluntary, intrinsically motivated, and process-oriented behavior directed by the child to more goal-oriented and adult-directed forms of *playful instruction*. In between these two poles, *guided play* represents a blend of *free play* and *playful instruction* (Zosh et al., 2018). Guided play can be described as a playful activity that is directed by the child, i.e., the child is autonomous to decide what to do, for how long and at what pace. The adult's role in guided play is to prepare a play environment and support the children's activities to facilitate learning (Weisberg et al., 2013; Weisberg et al., 2016; Zosh et al., 2018).

Guided play shares strong commonalities with the guided inquiry principle, which has been identified as one of the most effective approaches to learning and teaching (Alfieri et al., 2011; Lazonder & Kamp, 2012; Mayer, 2004). Researchers have frequently framed guidance in inquiry-based science teaching within the scaffolding construct (Hmelo-Silver et al., 2007). In the scaffolding literature, both material scaffolds and an adult's verbal scaffolds are considered effective in guiding children's learning (Martin et al., 2019; van de Pol et al., 2010). Accordingly, guided play can take at least two different forms (Weisberg et al., 2016), guided play with *material scaffolds only* and guided play with *additional verbal scaffolds*.

In guided play with *material scaffolds*, the adult provides the children with purposefully designed and structured materials aiming at a specific learning objective (Weisberg et al., 2016). Research indicates that children's explorations with purposefully structured and limited materials can foster science learning (Cook et al., 2011; van Schijndel et al., 2015). In particular, learning materials are effective when they link the learning objective to children's prior knowledge (Leuchter et al., 2014) and focus children's attention on those aspects that are essential for understanding (DeLoache, 2014). For example, to foster children's stability concepts, the adult might provide the children with an assembly of building blocks and a variety of photographs. In guided play with material scaffolds only, the adult initiates the play activities by inviting the children to rebuild the block constructions depicted on the photographs and to explore whether the constructions remain stable or tumble. By building these constructions, the children are likely to face evidence (the construction remains stable or tumbles) that might be incompatible with their intuitive theory (Gopnik & Wellman, 2012). However, beyond initiating children's explorations, the adult does not intervene in the process.

Research suggests that children show more explorative behaviors when an adult takes a passive role (Bonawitz et al., 2011). In contrast, studies indicate that children's unguided explorations might not be sufficient to encounter the learning objective (Butts et al., 1994; Chen & Klahr, 1999; Klahr & Nigam, 2004; Sarama & Clements, 2009). In the stability example, children might rebuild the construction inappropriately and thus might witness incorrect evidence or imperfect covariation. Moreover, children might interpret evidence inappropriately to confirm their intuitive but incorrect theory.

In guided play with *additional verbal scaffolds*, the adult not only provides materials but additionally plays along with the children, supports the children's play verbally and encourages higher order thinking (Chin, 2007; Haden, 2010; Kleickmann et al., 2016; Martin et al., 2019; Weisberg et al., 2016). The adult can use a set of verbal scaffolding techniques to aid children's cognitive activities (for an overview see, van de Pol et al., 2010) and support the cognitive processes involved in theory-evidence coordination, thus helping children adjust their intuitive theory. *Activating prior knowledge* by asking questions and prompting the children to express their presumptions, e.g., whether a block construction will remain stable or tumble can facilitate the integration of new aspects into existing schemata (Gurlitt & Renkl, 2010; Mayer, 1997; Weinert & Helmke, 1998). Additionally, *asking for the child's reasoning*, e.g., by prompting the child to justify their presumptions about the block construction's stability allows the child to structure their prior knowledge and thinking processes (Hsin & Wu, 2011). *Providing explanations* may help the child coordinate their observations with an evidence-based interpretation of a phenomenon (Murphy & Messer, 2000; Renkl, 2002). *Encouraging comparisons* supports the child in identifying relational similarities or differences between entities by highlighting certain features (Hsin & Wu, 2011; Richey & Nokes-Malach, 2013). Furthermore, *modeling*, i.e., performing certain behaviors and thinking styles, offers a possibility for imitation (Hmelo-Silver et al., 2007; Mayer, 2004).

Research indicates that guided play with *additional verbal scaffolds* promotes children's science learning more effectively than free play (Fisher et al., 2011; Hadzigeorgiou, 2002; Pine et al., 1999; Reuter & Leuchter, 2020). However, there are only a few studies that have deliberately compared the effectiveness of *material scaffolds* with *additional verbal scaffolds* for children's science learning. Leuchter and Naber (2019) found that a combination of structured materials and verbal scaffolds supported 6- to 7-year-old children's learning in the physics domain of force more than only materials, only verbal support or free exploration. Similarly, the results of Hadzigeorgiou (2002) show that 4- to 6-year-olds perform more meaningful activities at an inclined plane to explore the concept of mechanical stability when they received structured materials and verbal scaffolding compared to children who received only materials or played freely.

Studying children's scientific reasoning in a playful context can aim at unraveling the interplay of material and verbal scaffolds. Concerning children's reasoning about science phenomena guided play can serve as a developmentally appropriate context to shed a light on (1) children's theory adjustment, (2) the way their prior theories interact with the evidence provided through the scaffolding materials, and (3) the conditions that may support children to coordinate theory with evidence.

2.3 The statics domain and children's beliefs about balance

Statics can be defined as the state of equilibrium of an object, which in turn is concerned with forces acting on objects that are either at rest or in motion (Riley & Sturges, 1993). Statics is therefore concerned with stability. If the middle of a symmetrical object is supported by a supporting surface, the object will remain stable. Therefore, the consideration of an object's geometrical center is sufficient when rating symmetrical objects. For an asymmetrical object, however, the mass must be considered because the geometrical center and the center of gravity do not correspond. If the center of gravity of an object is supported, the object will remain in place; however, if it is not supported, the object will tumble. According to Bonawitz et al. (2012), Krist (2010), and Siegler (1976), with increasing age children develop an understanding of the weight distribution so that they can estimate the stability of an asymmetrical object/construction.

Studies with infants have mostly employed the violation of expectation paradigm and have shown that infants have basic knowledge about stability (Baillargeon & Hanko-Summers, 1990). Studies with older children, however, have shown that even preschoolers face problems explaining why certain objects either remain stable or tumble. Krist and Krüger (2012) explain this discrepancy with different approaches of the violation of expectation paradigm and verbal explanations as a possible reason for this ability gap. They state that being surprised (violation of expectation) does not take as much cognitive reasoning as verbally explaining one's underlying theory.

Young children indeed hold misconceptions about balance. Siegler (1976) and Siegler and Chen (1998) placed different weights at different distances on a fulcrum and asked children to rate the fulcrum's balance. The researchers found that children from nine years of age started to consider both weight and distance, while younger children tend to view weight and distance separately. Other studies by Krist and colleagues have shown that between the ages of three to eight, children's abilities of balancing symmetrical and asymmetrical blocks and estimating symmetrical as well as asymmetrical objects' stabilities increase continuously independent of the type of assessment (rating photographs, Krist, 2010; eye tracking, Krist et al., 2018; balancing blocks, Krist et al., 2005). Even though children's estimation of asymmetrical blocks' stabilities increased, all three studies found children's performance on the estimation of symmetrical objects (e.g., cuboids) to be superior to their estimation of asymmetrical objects (e.g., L-shaped objects). As noted earlier, the center of gravity does not correspond to the geometrical center of an asymmetrical object. For symmetrical objects, however, considering their geometrical center is sufficient. Thus, children's difficulty in estimating the stability of asymmetrical objects indicates that they face problems considering the weight distribution.

Some studies have taken a closer look at children's theories about balance. Pine et al. (2007) asked 6- to 8-year-old children about their reasoning when balancing beams on a fulcrum and categorized their verbal utterances as well as their gestures into four categories: middle, weight, distance, and other. They found that most answers fell into the other or weight categories, and few children considered the distance. Moreover, Weber and Leuchter (2020) found that more than half of the 5- and 6-year-olds in their sample used an undifferentiated pattern when rating photographs of asymmetrical objects, approximately 1/3 applied Center knowledge, and less than 10% of children applied Mass knowledge.

The above studies have examined children's knowledge about stability from a developmental psychological perspective. However, it is also of interest if children's Mass knowledge can be supported in regard to their developmental constraints. Playful interventions with building blocks have supported the acquisition of different mathematical and spatial skills in other studies (e.g., Borriello & Liben, 2018; Ferrara et al., 2011; Fisher et al., 2013; Thomson et al., 2018; Verdine et al., 2019).

Regarding children's rating of stabilities, Pine and Messer (2003) found that 5- to 6-year-olds were able to balance more symmetrical as well as more asymmetrical blocks after playing with the blocks compared to a pretest. In another study, children between four and seven years of age first balanced symmetrical and asymmetrical objects on a beam scale (Bonawitz et al., 2012), and their balancing behavior was categorized into three categories (*No, Center, Mass* *theory*). Furthermore, the results indicated that younger children tend to use an undifferentiated pattern (No theory) and do not consider the center or the mass. Second, after balancing objects on a beam, children either played with a mass-consistent or a center-consistent toy on their own and freely. Afterwards, they again balanced an asymmetrical block. Children who had a Center theory before playing observed evidence that did not confirm their theory if they were in the mass condition. Many of these children adopted a Mass theory. Children who had a Mass theory before playing also observed evidence that did not confirm their theory if they were in the center condition. Most of these children did not alter their balancing behavior and instead explained away the evidence and remained Mass theorists. This outcome indicated that even a short presentation of counterevidence can support children’s learning, but that their prior theories need to be considered.

The different effects of free play and guided play with material and material + verbal scaffolds on children’s science learning in the domain of balance with regard to their prior theories have not yet been investigated. Furthermore, it remains unclear whether these adjustments remain stable over a longer period of time or if the children relapse into their prior intuitive theories.

2.4 Possible relationship with intelligence

Intelligence is one of the most important prerequisites for learning. The ability to solve or complete puzzles or patterns is considered an indicator of fluid intelligence (Cattell, 1987; Flynn & Blair, 2013). Two components of fluid intelligence are figural perception and figural reasoning as indicators of an individual’s ability to perceive and mentally represent objects and abstract certain characteristics (Cattell, 1987; Weiß & Osterland, 2013). In the context of stability, figural perception and figural reasoning might contribute to children’s Mass theory. To rate stabilities correctly, children must perceive, mentally represent and abstract the spatial features of the objects or constructions (Weber & Leuchter, 2020).

Language capacity is considered an indicator of crystallized intelligence and is one of the key indicators of mental ability in young children (Cattell, 1987; Flynn & Blair, 2013). Language capacity contributes to knowledge acquisition (Thorsen et al., 2014). Moreover, children with a higher language capacity might find it easier to articulate their reasoning and might profit more from verbal scaffolds.

2.5 The present research

The present study is concerned with the effects of three different types of construction play on children's science learning in the statics domain, specifically constructions' stabilities. We implemented two types of guided play (verbal + material scaffolds, material scaffolds) and free play. Following the literature on theory theory, Bayesian inference, and theory-evidence coordination, young children's science learning may be fostered by confronting children with evidence in the form of perfect covariation (Gopnik & Wellman, 2012; Koerber et al., 2005). Furthermore, children's prior theories, which they have acquired through their everyday activities, e.g., their play, should be considered so that the presented evidence relates to these theories, which can then help children interpret the evidence (Gopnik & Meltzoff, 1997; Gopnik & Wellman, 2012). For example, at the age of 5 to 6, children might explain and predict the stability of an object with a Center theory or have other theories (Bonawitz et al., 2012; Weber & Leuchter, 2020). Material scaffolds can be prepared in such a way that they show perfect covariation for Mass theory and contradict Center theory. Through verbal scaffolding, an adult can help the children connect the evidence presented through the material scaffolds with their prior (intuitive) theories. Thus, scaffolds may support children's theory adjustment from Other[7] or Center theory to Mass theory. Since theory adjustment often happens gradually (Gopnik & Wellman, 2012), children should receive enough time to explore stabilities. We designed playful interventions that consider these constraints and investigated the effects of the different kinds of play on children's theory adjustments in the statics domain. Moreover, we explored whether these adjustments remained stable over an extended period of time.

Finally, interindividual prerequisites might be partly responsible for children's theory adjustment and interact with the type of playful intervention that the children received. Research on theory theory, Bayesian inference and theory-evidence coordination suggests that children with a consistent prior theory might not adjust their theories as easily as children with an inconsistent prior theory (Gopnik & Wellman, 2012; Koerber et al., 2005). Thus, we are interested in the contribution of children's prior theories on their adjustments after being confronted with perfect evidence for Mass theory. Additionally, intelligence affects learning (Cattell, 1987; Flynn & Blair, 2013; Thorsen et al., 2014) and may thus contribute to theory adjustment as well. With respect to fluid intelligence, we hypothesize that figural perception and figural reasoning

---

[7] By Other, we do not imply that children have no theory at all, but rather treat it as a rest category for children's answers that were neither concerned with the center nor the mass of constructions (Bonawitz et al., 2012, p. 221).

facilitate theory adjustment. With respect to crystallized intelligence, we hypothesize that children with higher language capacity might profit more from verbal scaffolds than children with lower language capacity.

Therefore, we specify the following research questions:

(1) Do children explain their reasoning about stability with Mass theory, Center theory or Other?

(2) Can guided play with material scaffolds and with or without verbal scaffolds enhance children's consistent use of Mass theory compared to free play (a) directly after a playful intervention and (b) over an extended period of time? Does intelligence relate to the consistent use of Mass theory?

(3) Does the consistency of children's prior theories relate to children's consistent use of Mass theory in the different play conditions (a) directly after a playful intervention and (b) over an extended period of time? Does intelligence relate to the consistent use of Mass theory when prior theories are considered?

(4) Do children with a consistent Mass theory after the playful intervention perform differently on a transfer test than children who did not use Mass theory consistently? Does intelligence relate to performance on the transfer test?

3 Methods

3.1 Participants

In total, 183 children (88 girls, 95 boys), between the ages of five and six ($M = 5.55$, $SD = 0.51$), took part in the study. The participants visited 23 kindergartens in Germany (2 to 13 children per kindergarten), which all agreed to take part in the study and helped connect with the children and their parents. The kindergartens were located either in villages (700 to 3,000 inhabitants; $N = 83$ children), small cities (less than 20,000 inhabitants; $N = 10$ children) or medium sized cities (approximately 50,000 inhabitants; $N = 91$ children of whom 51 lived in the city center and 40 in the periphery). A total of 171 children were European, 9 were Asian, 2 were African, and 1 was Central American. All children and their parents were informed about the goals of the study, and all children took part voluntarily and with their parents' consent.

Some children dropped out of the study completely because, e.g., they moved or were ill on the dates agreed with the kindergartens, and some other children had missing values on some of the items. We used pairwise deletion because we decided to include the highest amount of data whenever possible. Therefore, the number of participants varies between different analyses.

3.2 Procedure

The study consisted of a pre-post-follow-up design with two guided play groups and a free play group. The pretest (T1) took place approximately two weeks before the play session and the immediate posttest (T2). The follow-up (T3) took place approximately ten weeks after the posttest. The duration of the play session was approximately one hour.

For the intervention, the children were parallelized into the three intervention groups according to their language capacity, which was assessed at pretest. Thus, matched samples were produced and each child in the Verbal group had a “language capacity twin” in the Material and in the Free play group. Both the *Material group* (59 children, 32 girls, 27 boys) and the *Verbal group* (64 children, 27 girls, 37 boys) received scaffolding materials in the form of building blocks. The Verbal group received additional verbal scaffolds. The *Free play group* (60 children, 29 girls, 31 boys) played with building blocks freely. The reason for the differences in group size is that each intervention was to be conducted in a group of approximately four to six children to achieve ecological validity. Therefore, five children in the Verbal group were not assigned to language capacity triplets in the other two groups. In total, there were 51 interventions with group sizes varying between 2 to 6 children per group.

The play was led by one of six female experimenters. To prevent experimenter effects, group*experimenter was varied, so that all experimenters had led all intervention groups, i.e., Verbal, Material and Free play group. In all intervention groups, the children were free to choose what they wanted to build or if they wanted to build with a friend or rather on their own. Furthermore, breaks were always possible, and the children were free to stop playing entirely (Rubin et al., 1983; Weisberg et al., 2016). For manipulation check, the play sessions were video or audio recorded with the permission of parents and children. Based on the recordings, we rated children’s playfulness according to Bundy et al. (2001) as well as children’s on-task behavior. High inferential ratings showed that all children in all recordings showed indications of playfulness, e.g., children sang, laughed, joked around with another or the experimenter, chose to build challenging constructions, and built together. Moreover, children’s on-task behavior did not differ between the groups.

Children in the Material group and the Verbal group received photographs of different block constructions, which differed in the number of blocks used and in their complexity. With each photograph came a box containing the building blocks needed for building the construction. The blocks differed in their shapes (cuboids, triangles, etc.), sizes, and colors (brown, black, yellow, red, and green). The materials were developed prior to this study and tested in play sessions to ensure that children were able to rebuild the structures shown on the photographs and had fun playing with the materials. The material scaffolds were implemented to structure the play and served as suggestions for the children. However, the children in the guided play groups were free to build constructions other than those we presented to them. In order for the activity to be enjoyable and playful for the children, we allowed the children to decide whenever they wanted to move on to the next activity. However, the experimenters could suggest another activity if they felt that the children started to lose interest.

The material scaffolds encompassed five activities and were presented in a standardized order (Supplementary Material 1 for all photographs):

(1) Black block (11 photographs): You can build the building shown on the photograph. Build the building and guess if the blocks remain stable or tumble.

(2) Add-a-block (8 photographs): These blocks on the photos were bewitched so they would remain stable. Can you rebuild the building, so that it is stable? (If a child did not succeed, the experimenter provided a green block): Look, here is a green block. Try to stabilize the building with it.

(3) Sliding (9 photographs): This is the sliding play. You rebuild the building on the photograph. Then, you slide the upper block along the lower block until it falls (experimenter models it). That's noisy, isn't it?

(4) Rebuild (11 photographs): You can just rebuild the building on these photographs and see how well you are doing. Some buildings are very easy to rebuild; others are more difficult. However, every single one will remain stable if built correctly.

(5) Stable/Tumble (8 photographs): The buildings on the photographs will sometimes remain stable, but at other times, the blocks are bewitched. Look at the photograph and predict "stable" or "tumble", and then try them out to see whether you were correct.

Additionally, the Verbal group received verbal scaffolds in the form of the activation of prior knowledge, asking for reasoning, the provision of explanations, the encouragement of comparisons, and modeling (Table 1; Hogan & Pressley, 1997; van de Pol et al., 2010; for the script see, Supplementary Material 2). The experimenters used this limited set of scaffolds presented in the script but applied them flexibly when playing with the children. All experimenters had received a training on how to apply scaffolding during play prior to leading the interventions.

The Free play group received a large box with the same building blocks as the guided play groups, but the blocks were unstructured. The experimenters did not suggest any buildings that the children could build but only told the children to play with the blocks freely.

During the play time, the experimenter praised the children’s efforts in all three groups and motivated them to try again if they encountered problems with building. Sometimes children would ask the experimenter for help with building, which she would provide in the Verbal group. However, in the Material group or the Free play group, she would friendly decline and suggest that children ask another child for help (Supplementary Material 3 for excerpts from the playful intervention).

Table 1

Scaffolding techniques used in the Verbal group (Hogan & Pressley, 1997; van de Pol et al., 2010).

Technique	Example
Activating prior knowledge	Have you ever seen something like this?
Asking for reasons	Can you explain this in more detail, so I can really understand what you think?
Providing explanations	Well done! If the heavy side of a block hovers in midair, the block will tumble.
Encouraging comparisons	Your building looks different than [another child’s building], doesn’t it? What is different? Is something similar?
Modeling	Look! (Experimenter also looks very closely/experimenter shows how to build a certain building.)

3.3 Measures

Children’s theories. Children’s theories about stability were assessed with a standardized single interview consisting of photographs of four symmetrical and four asymmetrical block constructions, which were always supported by a black block (Supplementary Material 4). The children were asked to estimate whether the block construction presented in the first photograph would remain stable or not if the black block was removed. Afterwards, the children received a total of five blocks, namely, four cuboid blocks consisting of two brown blocks, one black block and one yellow block (9 cm*3 cm*1.5 cm), as well as one smaller black cuboid block (3 cm*3 cm*1.5 cm). All the blocks were made of nonlacquered wood and colored by the researchers using acrylic colors to avoid slipperiness. The brown blocks had a narrow line drawn onto them with a pencil to facilitate finding the blocks’ middle for the children. The children were asked to rebuild the construction presented in the photograph. Then, the interviewer repeated the children’s former answer regarding the construction’s stability (stable or unstable) and asked them to explain the prediction. They answered, and the interviewer invited them to remove the black block and ascertain whether they had rated the stability correctly. Then, they proceeded to the second block structure and so on. The interviews were videotaped or recorded if both the parents and the child had consented to it; if not, the interviewer made notes on the child's theory. The same interview was administered at each point of measurement, i.e., the same test items were presented at T1, T2 and T3. The testing time was approximately 10 minutes.

Only the asymmetrical block constructions were used in the data analyses. The test started with two symmetrical items to familiarize the children with the test logic. The fourth and sixth items showed symmetrical constructions and were applied to ensure that children had positive mastery experiences during the testing because studies have found that symmetrical items' stabilities are easier for children to estimate than asymmetrical items' stabilities (Krist, 2010). The three asymmetrical items showed perfect covariance for Mass theory because the weight distribution always determined the stability. However, the second asymmetrical item as well as all the symmetrical items could also be rated correctly with Center theory, while the first and third asymmetrical items could not. Therefore, the evidence for Center theory was imperfect.

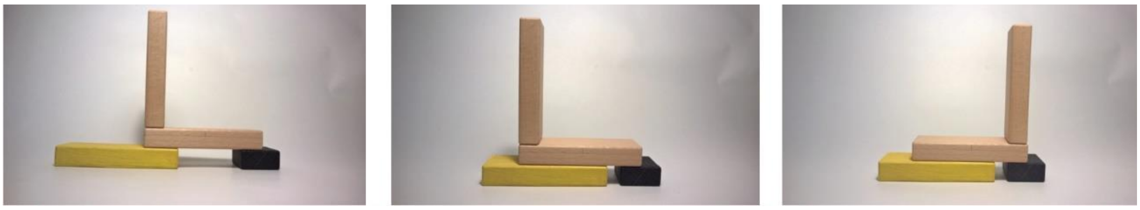


Figure 1. The asymmetrical constructions used to assess children's reasoning. Item 1 and 2 are stable constructions, item 3 is an unstable construction.

The children's answers were coded following the speech coding scheme from Pine et al. (2007), as shown in Table 2. If a child was unable to verbalize their answer, but, e.g., pointed at the vertical block, their answer was also rated as Mass theory. If a child indicated the middle with gestures, the answer was rated as Center theory. The children's explanations were coded by two independent raters, Cohen's κ > .90.

Regarding the items that were not used for further analysis, the distribution of the children's answers at T1 was as follows. For the first item (familiarization, symmetrical), 18 Mass, 118 Center, and 39 Other. For the second item (familiarization, symmetrical), 14 Mass, 124 Center, and 37 Other. For the fourth item (motivation, symmetrical), 19 Mass, 115 Center, and 42 Other. For the sixth item (motivation, symmetrical), 15 Mass, 96 Center, and 64 Other. The seventh item showed an asymmetrical construction but was removed from further analyses because it was inconclusive. The probability of the item remaining stable was 50% for statics reasons. Thus, there is no definite answer to this item. The third, fifth, and eighth items were asymmetrical items that were included in the data analyses and used for the assessment of children's stability theories (Figure 1).

Table 2

Coding scheme.

Coding	Speech	Example
Mass theory	The child refers to the weight being on one side of the brown blocks, mentions heaviness or talks about the importance of the vertical block.	"This side is heavier." "It's because of the block that's standing on the other."
Center theory	The child refers to the middle of the block or a bigger amount of the block resting on either the black or the yellow block.	"The brown block is resting more on the yellow block."
Other	Child speaks of something other than the two variables of interest (weight, middle), e.g., refers to the color.	"I don't know." "It tumbles, because it tumbles."

Transfer test. At the third point of measurement, a paper-pencil transfer test consisting of photographs of 16 asymmetrical block constructions was administered (Figure 2 and Supplementary Materials 5 for all transfer items), i.e., 8 stable constructions and 8 unstable constructions. The test was conducted in a group of up to six children who were seated back-to-back to prevent them from copying from each other. The test took approximately 10 minutes. The children were asked to rate the constructions' stabilities by either circling the photograph for a stable construction or crossing out the photograph for an unstable construction. Thus, children were not required to verbalize their knowledge. The constructions could only be rated correctly by considering the weight distribution because if the center of gravity was supported by the brown block but the geometrical center was not, then the constructions would always remain stable. However, if the center of gravity was not supported but the geometrical center was, then the construction would always tumble. For this instrument, the children did not need to explain their reasoning and only had to rate photographs. The children's content knowledge could be assessed to support the results of the reasoning test.

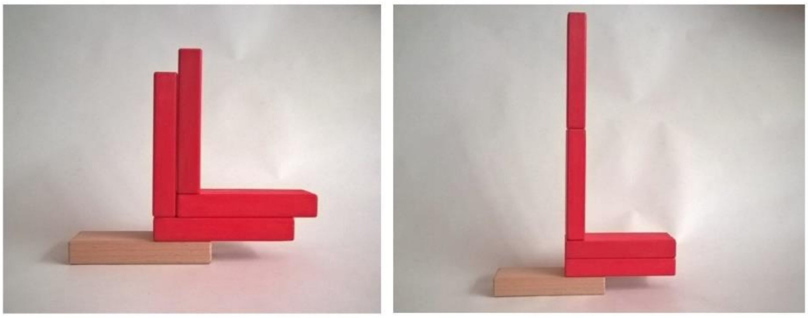


Figure 2. Example items of the transfer test, left is stable, right is unstable.

Aspects of intelligence. Visual perception and figural reasoning as two aspects of fluid intelligence were assessed with the labyrinths and matrices subtests of the Culture Fair Test (CFT 1-R; Weiß & Osterland, 2013) at T1. *t*-values were not available, as only two subtests were conducted (for more information on test parameters, please see CFT 1- R handbook).

Language capacity as an indicator for crystallized intelligence was assessed with the German version of the Peabody Picture Vocabulary Test (PPVT 4; Dunn et al., 2015) at T1. The PPVT is a picture-based standardized single interview for which *t*-values are available. The test consists of 19 sets with 12 items each consisting of four pictures per item. Five-year-old children start with set 4, and 6-year-old children start with set 5. For each item, the children receive a word and point at the corresponding picture. This procedure is continued until the child answers 8 out of 12 items in a set incorrectly (PPVT 4 handbook for more information). The PPVT 4 was also administered to ensure that all children understood and spoke the German language sufficiently.

The total testing time was approximately one hour at T1, 10 minutes at T2 and 20 minutes at T3. At T3, the follow-up test was administered before the transfer test. During the testing, breaks were possible whenever the child or the experimenter considered it necessary.

3.4 Data analyses

The statistics program R, version 3.6.2, (R Core Team, 2019) was used for data analyses. We used the survival (Therneau, 2015) package for the specification of survival analyses and the survminer (Alboukadel et al., 2019) package for forest plots.

In the first step, we investigated the number of children who had a Mass theory, a Center theory or Other on each point of measurement. The children received feedback about the correctness of their stability rating because after rating the stability, they removed the black block and could ascertain if they had answered correctly. Therefore, they had the opportunity to learn during testing. Thus, their answers were not independent and could not be summarized but instead were treated as individual events as the assumption of local stochastic independence was violated. Therefore, we used methods of risk-event analysis to analyze the group differences in the application of Mass theory (Singer & Willett, 2003).

4 Results

4.1 Children's use of Mass theory

To address the first research question concerned with children's use of causal relations, especially Mass theory, when explaining asymmetrical objects' stabilities, we investigated the percentage of children who applied mass for explaining each of the three asymmetrical constructions on the first point of measurement. The following results were obtained across all groups. For the first item, 11%[8] of children explained their reasoning with Mass theory, 41% with Center theory, and 48% provided another answer. For the second item, 20% of children explained their reasoning with Mass theory, 56% with Center theory, and 23% provided another answer. For the third item, 16% of children explained their reasoning with Mass theory, 48% with Center theory, and 35% provided another answer, Figure 3 for percentage shares of children's theories (Supplementary Material 6 for percentages of correct answers). To compare the probability of answering with a specific theory between items, we compared the proportions of children who had answered with a specific theory (either Mass, Center or Other) using z-tests of proportions. They revealed that children were not more likely to explain their reasoning with Mass for any of the three items, $z = 5.41$, $df = 2$, $p = .067$; however, the answer probabilities for Center theory, $z = 8.87$, $df = 2$, $p = .012$, and Other, $z = 23.90$, $df = 2$, $p < .001$, differed across the items.

The percentage of children applying each theory for each group at T1, T2 and T3 is presented in table 3. The use of Mass theory increased at T2 and T3, especially in the Verbal group.

---

[8] Rounded percentages.

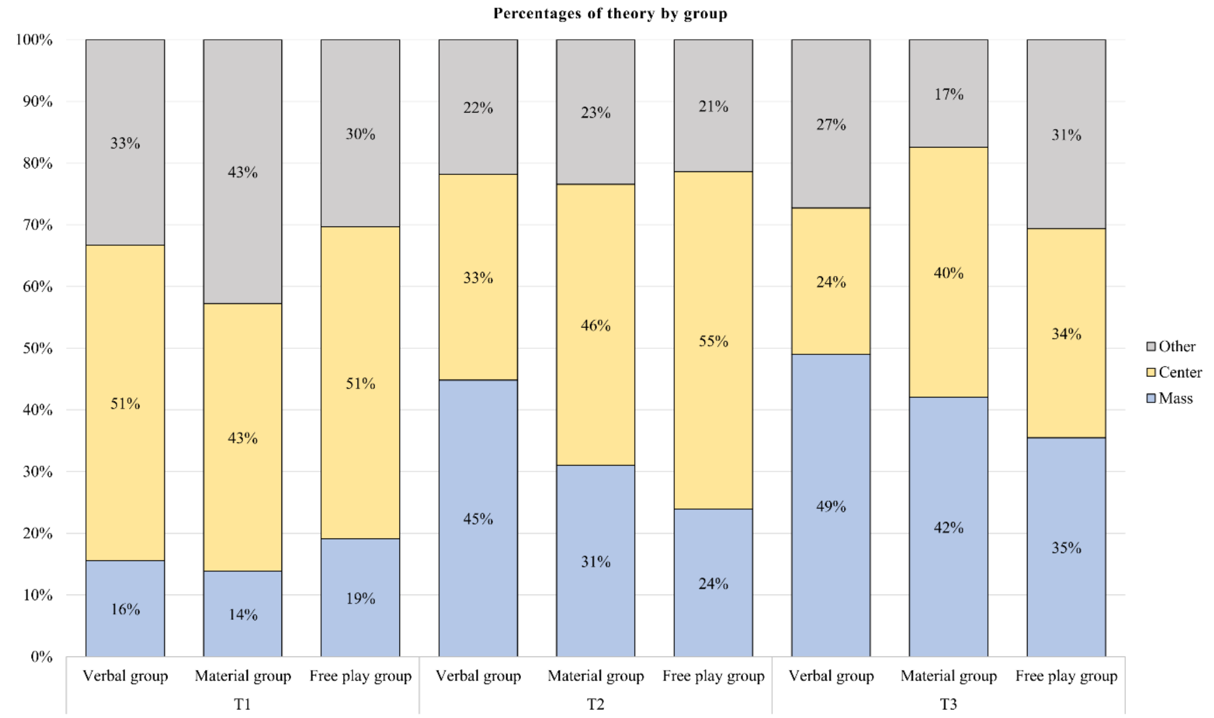


Figure 3. Percentages of theories applied by the children to explain all three reasoning items’ stabilities per measurement point and intervention group.

Table3

Number of children applying each theory in each group over all three points of measurement.

		Verbal group (N = 64)			Material group (N = 59)			Free play group (N = 60)		
	Item	Other	Center	Mass	Other	Center	Mass	Other	Center	Mass
T1	1	47%	40%	13%	51%	40%	9%	47%	42%	12%
	2	24%	60%	16%	30%	50%	20%	15%	59%	25%
	3	29%	53%	18%	49%	40%	11%	29%	51%	20%
T2	1	38%	25%	38%	32%	43%	25%	34%	51%	15%
	2	14%	39%	46%	13%	53%	34%	9%	60%	30%
	3	13%	36%	51%	25%	40%	35%	21%	53%	26%
T3	1	35%	23%	42%	40%	36%	24%	48%	29%	23%
	2	15%	27%	58%	10%	36%	55%	8%	46%	46%
	3	32%	21%	47%	2%	50%	48%	23%	54%	23%

Note. Percentages are given in rounded numbers.

4.2 Effects of guided play

To address the second research question concerning whether different kinds of play can enhance children's consistent use of Mass theory directly after the intervention as well as over a longer period of time, we used methods of survival analyses. Survival analysis is used to analyze the expected duration of time until an event takes place. In our case, the event is the children's consistent use of Mass theory after the playful intervention.

Consistency at T2. First, we used the binomial distribution to find a cut-off that guarantees that the probability of children finding the correct answer through guessing was below 10%. This enabled us to find how many correct answers might be given through guessing with $\sigma = 1.64$, i.e., $p < .10$, and a binomial probability of 1/3. Thus, we are able to categorize the children into children who explained their reasoning with Mass theory either consistently or inconsistently directly after the intervention at T2. Children with 3 out of 3 correct answers were rated as answering consistently, $p = .037$. Seven children who had explained their reasoning with Mass theory consistently at T1 (3 out of 3 Mass answers) were excluded from these analyses.

We defined each item as a point in time; therefore, time = 1 refers to the first item of T2, time = 2 to the second item of T2, and time = 3 to the third item of T2. The event of answering with Mass theory 3 out of 3 times could only take place at time = 3 or not at all. The detailed results of the Kaplan-Meier analysis are presented in Supplementary Material 7. The survival rate implies the percentage of children who remain either Center theorists or Other, and thus are not applying Mass theory consistently at T2. Therefore, the rate of children who applied Mass theory consistently is 100% minus survival rate, e.g., 100%−77% for the Verbal group. The results indicated that 23% of the children in the Verbal group explained their reasoning with Mass consistently at T2 compared to 9% of children in the Material group and 6% of children in the Free play group.

To investigate the results of the Kaplan-Meier analysis with a stricter procedure and to include the contributions of metric predictors, i.e., fluid and crystallized intelligence, a Cox regression (Table 4) was specified, likelihood ratio test = 12.87, $p = .012$. The Cox proportional hazard model assumes that the hazard curves for the groups should be proportional. This means that if child 1 is twice as likely to explain their reasoning with Mass theory than child 2 at an initial point in time, then at all later points in time, child 1 remains twice as likely to explain their reasoning with Mass theory consistently compared to child 2. In this particular Cox-regression, the event could only take place at time = 3. Therefore, the proportional hazard assumption was met.

Table 4

Cox-regressions for children's acquisition of Mass theory at T2

Development of consistencies between groups (Cox-regression)						
					95% CI	
	b	*SE*	*z*	HR	LL	UL
ΔFree play–Verbal	1.31*	0.65	2.01	3.70	1.03	13.26
ΔMaterial–Verbal	0.69	0.58	1.18	1.99	0.63	6.26
ΔFree play–Material	0.62	0.76	0.81	1.86	0.42	8.31
Fluid intelligence	0.04	0.06	0.69	1.04	0.93	1.16
CrI	0.08*	0.03	2.41	1.08	1.01	1.15

Interaction of crystallized intelligence with intervention group						
					95% CI	
	b	*SE*	*z*	HR	LL	UL
ΔFree play–Verbal*CrI	−0.05	0.11	−0.44	0.95	0.76	1.19
ΔMaterial–Verbal*CrI	0.15*	0.07	2.15	1.67	1.01	1.34
ΔFree play–Material*CrI	−0.20	0.12	−1.71	0.82	0.65	1.03
CrI Free play	0.16	0.10	1.58	1.18	0.96	1.44
CrI Material	−0.04	0.06	−0.68	.96	0.86	1.08
CrI Verbal	0.11**	0.04	2.65	1.12	1.03	1.22
Fluid intelligence	0.05	0.06	0.87	1.05	0.94	1.18

Notes. CrI = crystallized intelligence. HR = hazard ratio. CI = 95% confidence-interval. LL = lower level. UL = upper level. *p < .05. **p < .01.

The Cox regression showed group differences in the consistency of answering with mass between the Free play group and the Verbal group. However, there were no differences between the Material group and the Verbal group or the Free play group and the Material group. The Verbal group was almost four times as likely to explain their reasoning with Mass theory consistently compared to the Free play group, as indicated by the regression coefficient, and by a factor of HR = 3.70. Although neither the differences between the Verbal group and the Material group nor those between the Material group and the Free play group were statistically significant, descriptively, the Cox regression implied that the Verbal group had twice the chance of applying

Mass theory consistently than the Material group, HR = 1.99 and that the Material group had approximately twice the chance than the Free play group, HR = 1.86. Crystallized intelligence contributed to the consistent application of Mass theory, while fluid intelligence did not.

We tested whether crystallized intelligence interacted with the children's Mass theory in the three intervention groups. The analysis showed a difference between the Material group and the Verbal group dependent on crystallized intelligence, crystallized intelligence*ΔMaterial–Verbal, $b = 0.15$, $p = .032$. This indicates that children with high crystallized intelligence in the Verbal group profited more than children with high crystallized intelligence in the Material group (Table 5).

Consistency over T2 and T3. Next, we were interested in whether children's answers differed in their consistency over a longer period of time to check if the effect of the guided play was lasting. Therefore, we combined the three items of T2 and T3 into 6 points in time. Again, we used the binomial distribution with $\sigma = 1.64$, $p < .10$, and binomial probability = 1/3 to categorize the children into children who explained their reasoning with Mass theory either consistently or inconsistently. Children with ≥ 4 Mass explanations out of 6 when combining the items of the posttest and the follow-up were categorized as answering consistently, $p = .097$. The first point in time on which the event could take place was time = 4, i.e., the first item of T3, because the children had to answer four items with Mass theory to fulfill the event. The event could also take place at time = 5, i.e., the second item of T3, and at time = 6, i.e., the third item of T3. We specified a Kaplan-Meier analysis to investigate the percentage of children using Mass theory consistently (Figure 4 and Supplementary Material 7). Extending the descriptive results, the children in the Verbal group had the highest percentage of using Mass theory consistently at each point in time compared to the other two groups, Verbal group = 40%, Material group = 23%, Free play group = 15%.

Next, a Cox regression (Table 5) was specified, likelihood ratio test = 22.38, $p < .001$. First, we tested the proportional hazard assumption by correlating the scaled Schoenfeld residuals for group membership with time to ensure that the time and the residuals were independent. The hazard curves for the groups were proportional, as indicated by the global test, $\chi^2 = 5.56$, $p = .234$, as well as the group comparisons, all $p > .05$. The Cox regression showed group differences between the Verbal group and the Free play group, with the Verbal group having a higher chance of explaining their reasoning with Mass theory consistently by factor HR = 3.45. Again, there were no group differences between the Material group and the Free play group or between the Material group and the Verbal group. However, descriptively, the Verbal group had the highest chance of explaining their reasoning with Mass theory. Fluid and crystallized intelligence contributed to the consistent explanation with Mass theory. For the hazard ratios, see Figure 5.

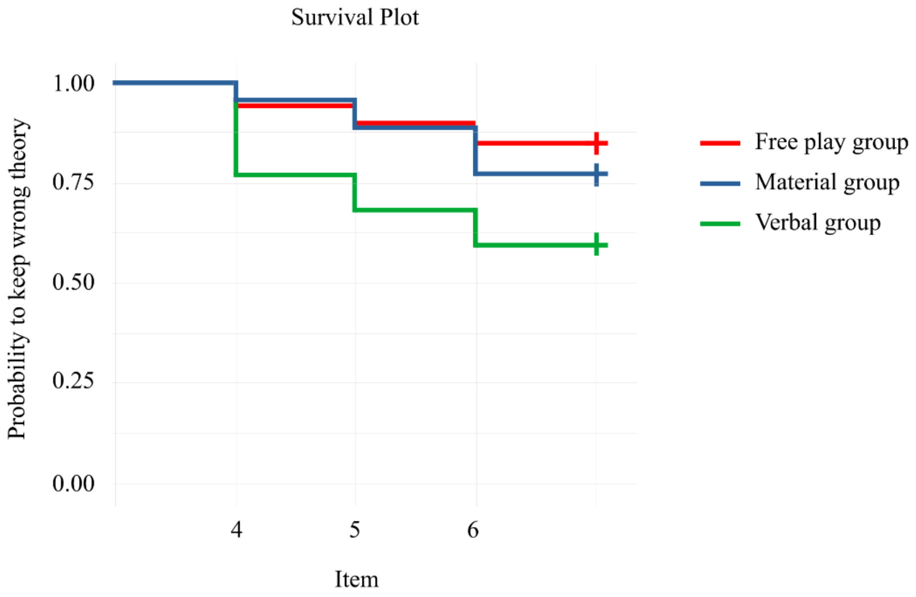


Figure 4. Survival curves of the three play groups.

Again, the interaction of crystallized intelligence and the intervention group was included in the Cox regression (Table 5). Crystallized intelligence contributed to the consistent use of Mass theory in the Verbal group and in the Free play group but not in the Material group. These differences were statistically significant. For low crystallized intelligence, the children in the Material group profited most from the intervention compared to the Verbal group and the Free play group.

Table 5

Cox-regressions for children's acquisition of Mass theory at T3 considering T2

Development of consistencies between groups (Cox-regression) over T2 and T3						
					95% CI	
	b	*SE*	*z*	HR	LL	UL
ΔFree play–Verbal	1.22**	0.45	2.74	3.40	1.42	8.16
ΔMaterial–Verbal	0.57	0.41	1.39	1.77	0.79	3.95
ΔFree play–Material	0.66	0.51	1.30	1.93	0.72	5.18
Fluid intelligence	0.10*	0.04	2.48	1.11	1.02	1.20
CrI	0.06*	0.02	2.56	1.06	1.01	1.11
Interaction of crystallized intelligence with intervention group						
					95% CI	
	b	*SE*	*z*	HR	LL	UL
ΔFree play–Verbal*CrI	−0.02	0.06	−0.37	0.98	0.87	1.10
ΔMaterial–Verbal*CrI	0.14*	0.05	2.57	1.15	1.03	1.28
ΔFree play–Material*CrI	−0.16*	0.07	−2.30	0.85	0.74	0.98
CrI Free play	0.11*	0.05	1.99	1.11	1.00	1.23
CrI Material	−0.06	0.05	−1.25	0.94	0.86	1.03
CrI Verbal	0.08**	0.03	2.61	1.09	1.02	1.16
Fluid intelligence	0.11**	0.04	2.61	1.12	1.03	1.21

Notes. CrI = crystallized intelligence. HR = hazard ratio. CI = 95% confidence-interval. LL = lower level. UL = upper level. **p* < .05. ***p* < .01.

4.3 Relationship of children's theory at T1 and children's consistent use of Mass theory

To address the third research question concerned with the relationship of the children's prior theories and their consistent use of Mass theory after the playful intervention, we categorized the children into those answering consistently or inconsistently at T1. For this method, we used the same criterion used for the prior analyses, i.e., children explaining their reasoning with either Center theory or Other 3 out of 3 times, $\sigma = 1.64$, $p < .10$, binomial probability of 1/3, were categorized as answering consistently at T1, and the other children were categorized as answering inconsistently. Hence, for the following analyses, the sample was divided into six groups, i.e., a consistent and inconsistent group for each of the three intervention groups. For categorizing children consistently answering with Other, we considered those children who had provided a theory neither concerned with the center nor the mass of constructions for all three items of the pretest.

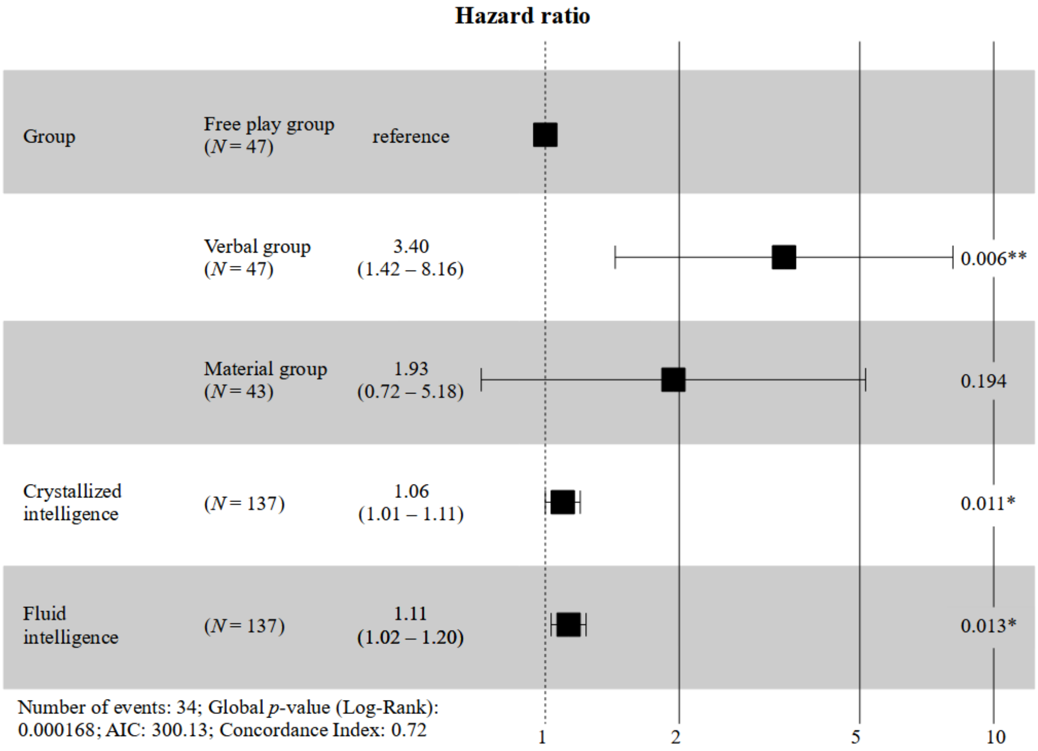


Figure 5. Forest plot with hazard ratios for the intervention groups including T2 and T3.

For the survival analysis, we applied the criterion of children explaining their reasoning with ≥ 4 Mass out of 6 to investigate whether children's prior theories relate to their acquisition of Mass theory. Kaplan-Meier analysis indicated that the children in the Verbal group who had answered inconsistently at T1 had the highest chance of explaining their reasoning with Mass consistently after the guided play, with 62% (Figure 4 and Supplementary Material 7 for the comprehensive results).

Next, a Cox regression (Table 6) was specified, likelihood ratio test = 26.15, $p < .001$. The hazard curves for the groups were proportional, as indicated by the global test, $\chi^2 = 7.13$, $p = .416$, as well as the group comparisons, all $p > .05$. We decided to use the Verbal group children who had answered inconsistently at T1 as the reference group for the Cox regression because theory suggests that this group should have the highest probability of explaining their reasoning with Mass theory. We found that this group had a significantly higher probability of explaining their reasoning with Mass theory than the Free play group children who had answered inconsistently at T1. Furthermore, descriptively, the Verbal group children who had answered inconsistently at T1 had the highest probability of all groups for explaining their reasoning with Mass theory (Figure 6). We found no differences between the Free play group children who had answered consistently at T1 and any of the other groups. Fluid and crystallized intelligence contributed to the consistent use of Mass theory for all groups.

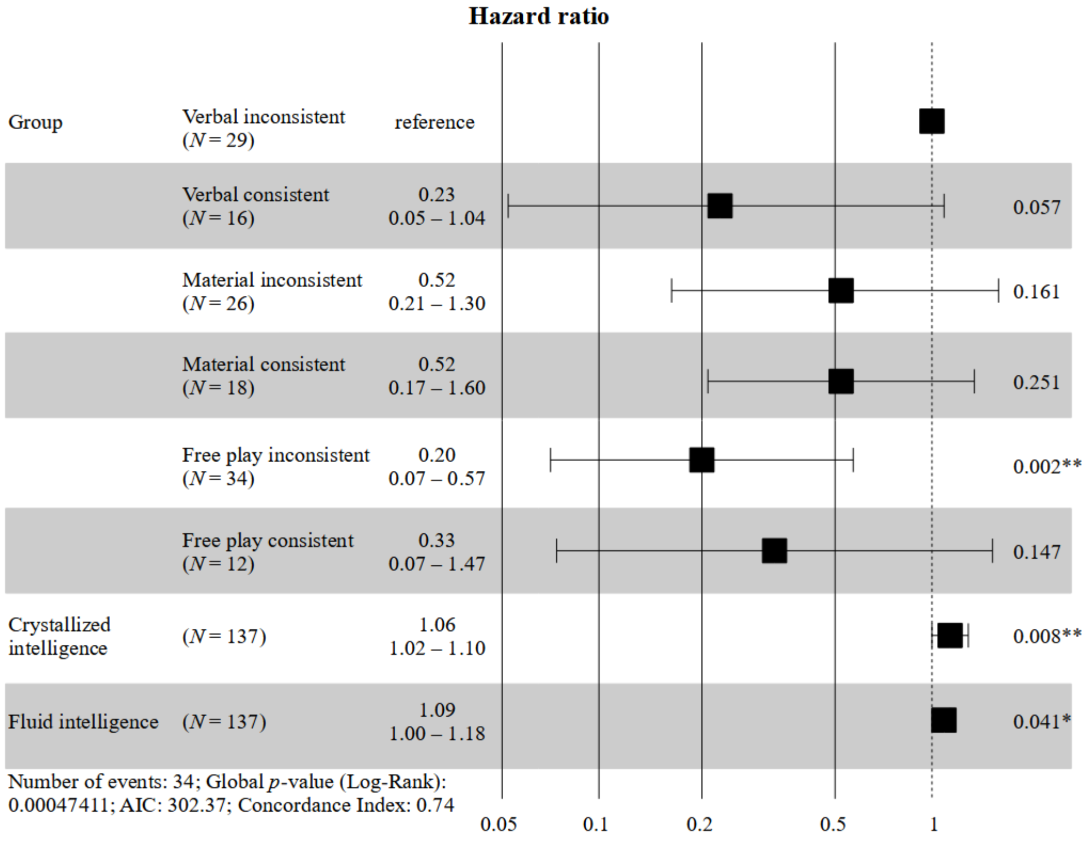


Figure 6. Forest plot with hazard ratios for the intervention groups in consideration of children's prior theories and including T2 and T3

Table 6

Development of Mass theory consistencies between groups (Cox-regression) T3 considering T2

Verbal inconsistent as the reference group

					95% CI	
	b	*SE*	*z*	HR	LL	UL
ΔVerbal inconsistent– Verbal consistent	−1.46	0.77	−1.90	0.23	0.05	1.04
ΔVerbal inconsistent– Material inconsistent	−0.66	0.47	−1.40	0.52	0.21	1.30
ΔVerbal inconsistent– Material consistent	−0.66	0.58	−1.15	0.52	0.17	1.60
ΔVerbal inconsistent– Free play inconsistent	−1.59**	0.53	−3.03	0.20	0.07	0.57
ΔVerbal inconsistent–Free play consistent	−1.11	0.76	−1.45	0.33	0.07	1.47
FI	0.08*	0.04	2.05	1.09	1.00	1.18
CrI	0.06**	0.02	2.67	1.06	1.02	1.10

Free play consistent as the reference group

					95% CI	
	b	*SE*	*z*	HR	LL	UL
ΔFree play consistent–Verbal inconsistent	1.11	0.76	1.45	3.02	0.68	13.46
ΔFree play consistent–Verbal consistent	−0.35	1.01	−0.35	0.70	0.10	5.07
ΔFree play consistent–Material inconsistent	0.45	0.81	0.55	1.56	0.32	7.64
ΔFree play consistent–Material consistent	0.45	0.87	0.51	1.56	0.28	8.65
ΔFree play consistent–Free play inconsistent	−0.49	0.85	−0.58	0.61	0.12	3.22
FI	0.08*	0.04	2.05	1.09	1.00	1.18
CrI	0.06**	0.02	2.67	1.06	1.02	1.10

Notes. FI = fluid intelligence. CrI = crystallized intelligence. Δβ = difference in regression coefficient between two groups. HR = hazard ratio. CI = 95% confidence-interval. LL = lower level. UL = upper level. *p < .05. **p < .01.

4.4 Transfer test

To address the fourth research question concerned with children's performance on the transfer test, we compared children who had explained their reasoning with Mass theory consistently at T2 and T3, *M* = 11.44, *SD* = 3.50, to children who had not explained their reasoning with Mass theory consistently at T2 and T3, *M* = 7.41, *SD* = 4.16. Regardless of the intervention group, those children who had explained their reasoning with Mass theory consistently at T2 and T3 performed better on the transfer test than the other children, $t(65.83) = -5.26$, $p < .001$. A multiple regression analysis showed that neither fluid, $b = 0.05$, $p = .571$, nor crystallized intelligence, $b = 0.04$, $p = .319$, contributed to children's performance on the transfer test beyond the consistent use of Mass theory. In addition, crystallized intelligence did not moderate the consistent use of Mass theory on children's performance on the transfer test, $b = 0.17$, $p = .082$.

Furthermore, we compared the three intervention groups on the transfer test: Verbal group, *M* = 8.17, *SD* = 4.64; Material group, *M* = 8.89, *SD* = 4.19; Free play group, *M* = 8.57, *SD* = 4.25. Crystallized intelligence did not moderate the effect of the intervention group on the children's performance on the transfer test. An ANOVA showed no differences between the groups on performance in the transfer test, $F(2) = 0.27$, $p = .762$. Furthermore, there were no differences between the groups if the consistency at T1 was included, $F(5) = 1.37$, $p = .242$. Crystallized intelligence did not moderate the effect of consistency at T1*intervention group on performance in the transfer test.

5 Discussion

Science learning in early childhood can and should be promoted (Eshach & Fried, 2005; Trundle & Saçkes, 2012). However, studies on early science learning are quite sparse, and it remains unclear how to best support young children with different individual prerequisites.

Therefore, we conducted a study on 5- to 6-year-olds' science learning in the specific domain of statics with regard to their prior intuitive theories and their individual cognitive prerequisites to investigate the effects of different types of play on theory adjustments. First, we were interested whether children explained their reasoning with Mass theory. In accordance with Pine et al. (2007), the children in our study faced problems estimating the stability of asymmetrical constructions and explaining the reasons for these stabilities. Most children provided another explanation and referred to characteristics of building blocks that have nothing to do with the mass or the center. Some children considered the center, and few considered the mass. This is in line with findings from other studies concerning the development of children's knowledge of mass (e.g., Bonawitz et al., 2012; Krist, 2010; Siegler, 1976; Siegler & Chen, 1998; Weber & Leuchter, 2020).

The result that young children do not have a Mass theory, but rather a Center theory or provide answers unconcerned with mass and center can serve as a starting point for designing learning environments that foster children's scientific reasoning, i.e., their theory adjustment, by providing them with perfect evidence for Mass theory (Gopnik & Wellman, 2012; Klahr et al., 2011; Koerber et al., 2005). These learning environments should consider developmentally appropriate practice, i.e., play with scaffolds (Copple & Bredenkamp, 2009; Weisberg et al., 2016). Thus, we investigated whether a playful intervention could support children's theory adjustment from an intuitive prior theory to a Mass theory. Play was implemented in the form of construction play with building blocks with differing amounts of adult guidance. The Free play group played with blocks on their own. The Material group received static material scaffolds prepared by an adult. In both of these groups, the adult only motivated and praised the children's efforts but did not intervene in the play process. The Verbal group received the same material scaffolds as the Material group, and additionally, an adult used verbal scaffolds during the play. Thus, the Verbal group received the highest amount of adult guidance. After the playful intervention as well as after the follow-up test, the Kaplan-Meier analysis showed that the children in the Verbal group were most likely to use Mass theory consistently to explain their reasoning. The Material group was more likely to use Mass theory than the Free play group. Group comparisons with a Cox regression showed that the Verbal group outperformed the Free play group but not the Material group.

The acquisition of Mass theory however, might be dependent on interindividual variables such as intelligence and consistency of prior theory that interact with the degree of scaffolding. Thus, we investigated whether children's prior theories are related to the adjustments of their theories to Mass theory. The link of children's prior theories with theory adjustment to Mass theory seems to be partly dependent on intervention type. The children in the Verbal group who had answered inconsistently during the pretest were most likely to adopt a Mass theory after the playful intervention compared to all other groups.

Last, we investigated whether the use of Mass theory to explain stabilities was related to children's results on a paper-pencil transfer test. Mass theorists performed better on the transfer test than Center theorists and children in the Other category. Our findings contribute to the literature on science education in the kindergarten years and will be discussed following the order of the research questions.

5.1 Children's use of Mass theory

The first research question was concerned with children's explanations of asymmetrical constructions' stabilities before the playful intervention. We investigated whether 5- to 6-year-old children explained asymmetrical constructions' stabilities with Mass theory, Center theory or Other.

Considering the assumptions of theory theory (Gopnik & Wellman, 2012), we found that children's theories about stability encompassed causal relations, as the theories provided explanations for the stability of the constructions and allowed for predictions of whether a construction would be stable or unstable depending on how it is supported. Thus, a child with a Center theory believes that the middle of a construction needs to be supported and that the middle is the cause for a construction's stability. A child with a Mass theory believes that a construction's weight needs to be supported and that the weight distribution is the cause for a construction's stability. A child providing another explanation might have other ideas concerning the causal relationship between support and stability, e.g., the color. Since most children could not explain asymmetrical constructions' stabilities correctly, we can assume that children do not have a Mass theory (e.g., Bonawitz et al., 2012; Krist, 2010; Siegler, 1976; Siegler & Chen, 1998; Weber & Leuchter, 2020).

Nevertheless, the children were more likely to use Center theory and less likely to provide another explanation in the course of the pretest. The reason might be that they received visual confirmation of the Center theory for the symmetrical items. Hence, the children might have inferred and generalized Center theory as an explanation for constructions' stabilities (Bonawitz et al., 2012). Therefore, the children might have acquired a Center theory instead of remaining in the Other category or kept their Center theory instead of adopting Mass theory.

Even though all the children had the opportunity to learn about the mass even during the pretest because they received feedback concerning the constructions' stabilities, the probability that the children would explain their reasoning with Mass theory remained the same across all three items at pretest. This outcome indicates that the presented evidence at pretest might not have been sufficient to acquire Mass theory. Even though the children observed perfect evidence for Mass theory (Koerber et al., 2005), a short presentation and asking for explanations about the stability in a one-to-one setting was not enough to introduce the children to Mass theory.

Since the children seemed to be unable to acquire an understanding of the mass that easily, construction play with varying degrees of structuring seems to be an appropriate approach to investigate whether the children could acquire an understanding of the mass during a play.

5.2 Effects of guided play

The second research question was concerned with the playful interventions' effects on children's consistent use of Mass theory and the possible relationship with intelligence.

From the results it can be concluded that the more support the children received when confronted with evidence, the more likely they were to adjust their Other or Center theory and to explain their reasoning with Mass theory. This result indicates that children need support when learning about stabilities. Guided play with material and verbal scaffolds has been shown to support children's acquisition of Mass theory more than free play (Zosh et al., 2018). In the Material group, the children might have overlooked the evidence, and in the Free play group, the children could only have observed it randomly.

Consistent with the literature on theory theory and theory-evidence coordination (e.g., Gopnik & Wellman, 2012; Koerber et al., 2005), our results indicate that three factors should be fulfilled when supporting young children's learning about science: (1) perfect covariation of the evidence, (2) considering the children's prior theories, and (3) giving the children enough time to explore the evidence. We approached these factors by (1) taking care to present children with perfect covariation. We only used constructions that included asymmetrical features that always confirmed Mass theory but always disconfirmed Center theory. (2) We assessed children's prior theories at pretest and tried to confront them with evidence supporting Mass theory and contradicting their Center theory. (3) The children were free to play with the provided materials for an hour so that they had a sufficient amount of time to explore and play with the materials. Furthermore, we considered the children's developmental constraints and related science to their everyday activities by using different playful activities with building blocks as a learning setting (Copple & Bredenkamp, 2009).

The activities the children engaged in fulfilled the characteristics of play (Daubert et al., 2018; Pellegrini, 2013; Rubin et al., 1983; Trawick-Smith, 2015; Weisberg et al., 2013). The children played voluntarily, and the play was child-directed and contained elements of choice. We did not measure the children's motivation during the play. Therefore, we cannot make a statement about their intrinsic motivation. An indication of their motivation might be that the children could stop playing at any time, but approximately 95% of the children continued to play for the provided time in all groups, as the video recordings for the manipulation check and the experimenters' written records showed. Furthermore, highly inferential analyses of the recordings demonstrated that all of the children in all of the groups showed playful behavior (Bundy et al., 2001). The play was not free of goals because we had a specific learning goal, namely, the acquisition of Mass theory, in mind. However, the play was still process-oriented, as we did not push this goal on the children.

Our playful intervention was based on the continual view postulated by, e.g., Zosh et al. (2018). The free play was free of an adult's guidance, the guided play with material scaffolds was structured and offered children suggestions for playing with blocks, and the guided play with material and verbal scaffolds offered additional verbal guidance. Specifically, when implementing the verbal scaffolds, we *asked for the children's prior knowledge* to allow them to express their presumptions to facilitate the adjustment of their theory (Gurlitt & Renkl, 2010; Mayer, 1997; Weinert & Helmke, 1998). By *asking for reasoning*, the children could be made aware of their theory, and they were supported in structuring their theory (Hsin & Wu, 2011). The *provision of explanations* helped the children organize new knowledge, e.g., knowledge about the mass, and integrate this new knowledge into their theories (Richey & Nokes-Malach, 2013). By *encouraging comparisons*, we tried to support the children in comparing stable and unstable constructions and to generalize the underlying principle, i.e., the mass (Hsin & Wu, 2011; Richey & Nokes-Malach, 2013). Last, *modeling* might have offered the children the possibility for imitation (Hmelo-Silver et al., 2007; Mayer, 2004).

Our study showed that crystallized intelligence had a positive relationship with children's consistent application of Mass theory directly after the intervention. The interaction of crystallized intelligence with the intervention group showed that children with high crystallized intelligence profited more in the Verbal group than did children with high crystallized intelligence in the Material group. Fluid intelligence did not relate to the consistent explanation with Mass theory directly after the intervention. This outcome indicates that children acquired Mass theory regardless of their ability to mentally represent and abstract the spatial features of constructions' stabilities. After including the follow-up, both fluid intelligence and crystallized intelligence related to the consistent explanation with Mass theory over an extended period of time. Crystallized intelligence interacted with the intervention group and related to the consistent application of Mass theory in the Verbal group and the Free play group but not in the Material group. Children with low crystallized intelligence were more likely to adjust their theories in the Material group, while children with high crystallized intelligence were more likely to adjust their theories in the Verbal group.

Language capacity is understood as an indicator of crystallized intelligence (Cattell, 1987; Flynn & Blair, 2013). Thus, our results suggest that the children with a low language capacity profited most from the Material group. Seemingly, the material scaffolds were sufficiently self-explaining so that the children with a low language capacity could observe evidence for Mass theory and adjust their theories. Moreover, the children with a low language capacity in the Verbal group did not profit from the intervention because they may have suffered from a high cognitive load (Kirschner, 2002). They not only needed to process the new information about the learning content provided through the verbal scaffolds but also the language itself. In contrast, the children with a high language capacity profited from the verbal scaffolds that were provided in the Verbal group. Our findings imply that when providing verbal instructions and support, it may be important to consider children's language capacity.

Children with a higher fluid intelligence, i.e., a capacity to represent constructions mentally and abstract important spatial features (Cattell, 1987; Weber & Leuchter, 2020; Weiß & Osterland, 2013), were more likely to acquire Mass theory over the course of 10 weeks. This outcome is in line with studies from developmental psychology showing that children's Mass theory develops between ages five and seven (Bonawitz et al., 2012). In their everyday lives, children have many possibilities to explore stabilities and develop an understanding of the underlying principles. A possible explanation might be that children with a higher fluid intelligence learn these principles faster than children with a lower fluid intelligence (Weber & Leuchter, 2020).

In addition to intelligence, other individual competencies, such as children's prior theories about stabilities, could relate to children's acquisition of Mass theory.

5.3 Relationship of children's prior theories and their consistent use of Mass theory

The third research question was concerned with the role of children's prior theories on their consistent use of Mass theory after the interventions.

The children with inconsistent prior theories who received the highest amount of support (Verbal group) acquired a Mass theory, while those children who received less support (Material group and Free play group) did not acquire a Mass theory. This result indicates that prior theories play a role in theory adjustment, which is in line with findings concerned with Bayesian inference in the context of theory-theory by, e.g., Bonawitz et al. (2012) and Sobel et al. (2004). The children who answered inconsistently at T1 did not have a consistent prior theory and therefore had the highest chance of acquiring a Mass theory. Their theoretical assumptions were inconsistent compared to the theories of children who had explained their reasoning consistently with

Center theory or Other (Gopnik, 2013; Gopnik & Wellman, 2012; Griffiths et al., 2011; Koerber et al., 2005). Since the children in the Verbal group were most likely to acquire a Mass theory, children seem to profit from high amounts of guidance and support; moreover, in our study, only observation seems to be insufficient for the understanding of counterevidence. In sum, children with inconsistent prior theoretical assumptions profit from supported play but do not adjust their theories by playing with blocks freely.

5.4 Transfer test

The fourth research question was concerned with children’s performance on a transfer test at follow-up. Independent of intervention group, we compared children who had used Mass theory consistently after the playful intervention to children who had failed to do so on a transfer test with asymmetrical block constructions. The children who had answered consistently outperformed children who had answered inconsistently. This result indicates that children who explain their reasoning with Mass theory are also more likely to rate asymmetrical constructions’ stabilities correctly, which suggests that the children acquired an understanding of Mass theory.

However, comparing the three intervention groups, children performed equally well on the transfer test, even though we found group differences on the reasoning test. The transfer test, unlike the reasoning test, was a paper-pencil test and according to Karmiloff-Smith (1992) and Pine and Messer (2003) rather tested knowledge that children might not have been able to verbalize. The transfer test indicates that children in all groups had knowledge about stabilities at T3, but only the children in the Verbal group were able to verbalize their reasoning.

The children who had a low language capacity succeeded in the transfer test but not in the reasoning task. They did not have to explain their reasoning in the transfer test; they were only required to decide about the constructions’ stabilities. Although we tried to consider a low language capacity in the reasoning task by counting specific gestures as indicators for Mass theory, e.g., pointing to the Mass, or Center theory, e.g., pointing at the middle, the transfer test was seemingly easier for the children to handle. This is especially meaningful for children with a different native language because these children might face challenges in explaining their reasoning adequately but might be able to show their knowledge about stability with a nonverbal test. Therefore, to offer children the opportunity to show their knowledge about science phenomena such as stability, methods that do not require the children to speak might be helpful.

5.5 Limitations

There are some limitations to this study concerning the implementation and measurement.

Implementation of play. Regarding the implementation of the playful intervention, we compared material scaffolds, material + verbal scaffolds and free play regarding their effects on children’s Mass theory. The effect of verbal scaffolds uncoupled from material scaffolds was not investigated. Future studies could implement a verbal scaffolds group by presenting children with the same unstructured building blocks a free play group receives and adding verbal scaffolds. Moreover, the implementation of a baseline group not receiving any intervention would allow investigating whether free play has an effect on children’s theory adjustment towards Mass theory compared to children’s development.

We videotaped only some of the playful interventions for a manipulation check; as some children or their parents denied permission to videotape, some interventions were only audio-recorded. Moreover, for a few interventions, neither videos nor recordings exist due to technical failures with the equipment. Therefore, children’s behavior during play cannot be analyzed, even though there might be interindividual differences in how children interacted with the experimenter and used the provided materials. For example, some children might have asked for help more often or might have built with the building blocks more actively, while others may have instead watched other children build. Furthermore, the materials provided in the guided play groups served as suggestions, and children in all groups were free to build other buildings. From the existing videos and recordings, we assume that the children in the guided play groups played the suggested activities and used the scaffolding materials. However, some children might have built at a higher pace and thus may have built more of the suggested structures than other children. Last, regarding children’s behavior, the amount of time that the children spent playing on their own or with other children, their manipulation of and their conversations about the building blocks might have contributed to children’s Mass theory after the intervention. These factors should be investigated in a future study.

In this study, we only used a limited set of verbal scaffolds and did not control for the verbal scaffolds’ adaptability. However, the adaptability might have contributed to children’s acquisition of Mass theory. Therefore, children’s and experimenters’ behavior during play should be investigated in the future.

Measures. The children received eight items showing different block constructions, and three asymmetrical items were used to assess children’s theories about stability. The other symmetrical items were used to familiarize the children with the test and motivate them during testing because children find it easier to estimate symmetrical constructions’ stabilities (Krist, 2010). These Center theory-compliant items might have led some children to adopt a Center theory instead of remaining in the Other category, even though the evidence for Center theory was imperfect. The results in this study show that although the children received these Center theory items, many still adopted Mass theory after the playful intervention. Future studies might benefit from the use of more items, which would also prolong the testing time, as more asymmetrical as well as symmetrical items would be needed. This addition could impact the children’s attention capacity and their motivation to participate.

Children received feedback about the constructions’ stabilities during testing because they built the construction and then removed the supporting black block to ascertain whether they had rated the stability correctly. Therefore, children had the opportunity to learn during testing, and the items were dependent on each other. As a result, we could not just sum up the items, and every item was considered a point in time. Thus, we used methods of risk-event analysis to analyze the data. Independent measurements would allow for different statistical approaches, e.g., statistical procedures that refer to the mean. Thus, in future studies, to achieve independent measures, children could not build constructions on their own but only rate and explain stabilities on the basis of photographs so that they do not receive feedback about stability.

Nevertheless, our study indicates that guided play can support young children’s science learning. Differing degrees of scaffolding in guided play can be beneficial for helping children with different prerequisites adjust their theories when observing new evidence.

6 Acknowledgments

We wish to thank Ina Plöger for helping develop the scaffolding materials and our research assistants Laura Löwe, Hanna Fuhrmann, Kristin Hoffmeister, Rebecca Ott and Aliena Eichhorn for helping with data acquisition, data input, transcribing and coding of the interviews. Last but not least, we would like to thank all children, who participated in our research.

7 References

Akerson, V. L., Buck, G. A., Donnelly, L. A., Nargund-Joshi, V., & Weiland, I. S. (2011). The importance of teaching and learning nature of science in the early childhood years. *Journal of Science Education and Technology*, *20*(5), 537–549. https://doi.org/10.1007/s10956-011-9312-5

Alboukadel, K., Kosinski, M., & Biecek, P. (2019). *Survminer: Drawing survival curves using ggplot2* (Version 0.4.6) [Computer software]. https://CRAN.R-project.org/package=survminer

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *10*3(1), 1–18. https://doi.org/10.1037/a0021017

Baillargeon, R., & Hanko-Summers, S. (1990). Is the top object adequately supported by the bottom object? Young infants' understanding of support relations. *Cognitive Development*, *5*(1), 29–53. https://doi.org/10.1016/0885-2014(90)90011-H

Bonawitz, E. B., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Borriello, G. A., & Liben, L. S. (2018). Encouraging maternal guidance of preschoolers' spatial thinking during block play. *Child Development*, *89*(4), 1209–1222. https://doi.org/10.1111/cdev.12779

Bundy, A. C., Nelson, L., Metzger, M., & Bingaman, K. (2001). Validity and reliability of a test of playfulness. *The Occupational Therapy Journal of Research*, *21*(4), 276–292. https://doi.org/10.1177/153944920102100405

Butts, D. P., Hofman, H. M., & Anderson, M. (1994). Is direct experience enough? A study of young children's views of sounds. *Journal of Elementary Science Education*, *6*(1), 1–16. https://doi.org/10.1007/BF03170646

Cattell, R. B. (1987). *Intelligence: It's structure, growth and action*. Elsevier.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*(5), 1098–1120. https://doi.org/10.1111/1467-8624.00081

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching*, *44*(6), 815–843. https://doi.org/10.1002/tea.20171

Cook, C., Goodman, N. D., & Schulz, L. E. (2011). Where science starts: Spontaneous experiments in preschoolers' exploratory play. *Cognition*, *120*(3), 341–349. https://doi.org/10.1016/j.cognition.2011.03.003

Copple, C., & Bredenkamp, J. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*. (3rd ed.). National Association for the Education of Young Children.

Cremin, T., Glauert, E., Craft, A., Compton, A., & Stylianidou, F. (2015). Creative little scientists: Exploring pedagogical synergies between inquiry-based and creative approaches in early years science. *Education 3-13*, *43*(4), 404–419. https://doi.org/10.1080/03004279.2015.1020655

Daubert, E., Ramani, G. B., & Rubin, K. H. (2018). Play-based learning and social development. In M. Boivin, R. D. Peters, & R. Tremblay (Eds.), *Encyclopedia on early childhood development*. Centre of Excellence for Early Childhood Development and the Strategic Knowledge Cluster on Early Child Development.

DeLoache, J. S. (2014). Early development of the understanding of symbolic artifacts. In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 312–336). Wiley-Blackwell.

Dunn, L. M., Dunn, D. M., Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *PPVT 4. Peabody Picture Vocabulary Test*. Pearson Assessment.

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology*, *14*(3), 315–336. https://doi.org/10.1007/s10956-005-7198-9

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block talk: Spatial language during block play. *Mind, Brain, and Education*, *5*(3), 143–151. https://doi.org/10.1111/j.1751-228X.2011.01122.x

Fisher, K. R., Hirsh-Pasek, K., Golinkoff, R. M., Singer, D. G., & Berk, L. E. (2011). Playing around in school: Implications for learning and educational policy. In A. D. Pellegrini (Ed.), *The Oxford handbook of the development of play* (pp. 341–362). Oxford Univ. Press.

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development*, *84*(6), 1872–1878. https://doi.org/10.1111/cdev.12091

Flynn, J. R., & Blair, C. (2013). The history of intelligence: New spectacles for developmental psychology. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 765–790). Oxford Univ. Press.

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*, 150–158.

Gopnik, A. (2013). Causality. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 628–650). Oxford Univ. Press.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. MIT Press.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620–629. https://doi.org/10.1037//0012-1649.37.5.620

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, *35*(8), 1407–1455. https://doi.org/10.1111/j.1551-6709.2011.01203.x

Gurlitt, J., & Renkl, A. (2010). Prior knowledge activation: How different concept mapping tasks lead to substantial differences in cognitive processes, learning outcomes, and perceived self-efficacy. *Instructional Science*, *38*(4), 417–433. https://doi.org/10.1007/s11251-008-9090-5

Haden, C. A. (2010). Talking about science in museums. *Child Development Perspectives*, *4*(1), 62–67. https://doi.org/10.1111/j.1750-8606.2009.00119.x

Hadzigeorgiou, Y. (2002). A study of the development of the concept of mechanical stability in preschool children. *Research in Science Education*, *32*(3), 373–391. https://doi.org/10.1023/A:1020801426075

Hawkins, J., Pea, R. D., Glick, J., & Scribner, S. (1984). "Merds that laugh don't like mushrooms": Evidence for deductive reasoning by preschoolers. *Developmental Psychology*, *20*(4), 584–594. https://doi.org/10.1037/0012-1649.20.4.584

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. G. (2009). *A mandate for playful learning in preschool. presenting the evidence*. University Press.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107.

Hogan, K., & Pressley, M. (1997). Becoming a scaffolder of students' learning. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning* (pp. 185–191). Brookline.

Hsin, C.-T., & Wu, H.-K. (2011). Using scaffolding strategies to promote young children's scientific understandings of floating and sinking. *Journal of Science Education and Technology*, *20*(5), 656–666. https://doi.org/10.1007/s10956-011-9310-7

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

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*, 1–10.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effect of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667. https://doi.org/10.1111/j.0956-7976.2004.00737.x

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*(6045), 971–975. https://doi.org/10.1126/science.1204528

Kleickmann, T., Tröbst, S., Jonen, A., Vehmeyer, J., & Möller, K. (2016). The effects of expert scaffolding in elementary science professional development on teachers' beliefs and motivations, instructional practices, and student achievement. *Journal of Educational Psychology*, *108*(1), 21–42. https://doi.org/10.1037/edu0000041

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*(3), 141–152. https://doi.org/10.1024/1421-0185.64.3.141

Krist, H. (2010). Development of intuitions about support beyond infancy. *Developmental Psychology*, *46*(1), 266–278. https://doi.org/10.1037/a0018040

Krist, H., Atlas, C., Fischer, H., & Wiese, C. (2018). Development of basic intuitions about physical support during early childhood: Evidence from a novel eye-tracking paradigm. *Quarterly Journal of Experimental Psychology*, *2*, 174702181773719. https://doi.org/10.1177/1747021817737196

Krist, H., Horz, H., & Schönfeld, T. (2005). Children's block balancing revisited: No evidence for representational redescription. *Swiss Journal of Psychology*, *64*(3), 183–193. https://doi.org/10.1024/1421-0185.64.3.183

Krist, H., & Krüger, M. (2012). Towards a new method for bridging the gap between "smart" infants and "dumb" preschoolers. *European Journal of Developmental Psychology*, *9*(5), 631–637. https://doi.org/10.1080/17405629.2012.657895

Lazonder, A. W., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction*, *22*(6), 458–464. https://doi.org/10.1016/j.learninstruc.2012.05.005

Leuchter, M., & Naber, B. (2019). Studying children's knowledge base of one-sided levers as force amplifiers. *Journal of Research in Science Teaching*, *56*(1), 91–112. https://doi.org/10.1002/tea.21470

Leuchter, M., Saalbach, H., & Hardy, I. (2014). Designing science learning in the first years of schooling. An intervention study with sequenced learning material on the topic of 'floating and sinking'. *International Journal of Science Education*, *36*(10), 1751–1771.

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2013). Science literacy in school and home contexts: Kindergarteners' science achievement and motivation. *Cognition and Instruction*, *31*(1), 62–119. https://doi.org/10.1080/07370008.2012.742087

Martin, N. D., Dornfeld Tissenbaum, C., Gnesdilow, D., & Puntambekar, S. (2019). Fading distributed scaffolds: The importance of complementarity between teacher and material scaffolds. *Instructional Science*, *47*(1), 69–98. https://doi.org/10.1007/s11251-018-9474-0

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist*, *59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Murphy, N., & Messer, D. (2000). Differential benefits from scaffolding and children working alone. *Educational Psychology*, *20*(1), 17–31. https://doi.org/10.1080/014434100110353

Pellegrini, A. D. (2013). Play. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol 2: Self and other* (pp. 276–299). Oxford Univ. Press.

Pellegrini, A. D., & Gustafson, K. (2005). Boys' and girls' uses of objects for exploration, play, and tools in early childhood. In A. D. Pellegrini & P. Smith (Eds.), *The nature of play: Great apes and humans* (pp. 113–138). Guilford Press.

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, *31*(2), 153–179. https://doi.org/10.1111/j.2044-835X.2012.02082.x

Pine, K. J., Lufkin, N., Kirk, E., & Messer, D. (2007). A microgenetic analysis of the relationship between speech and gesture in children: Evidence for semantic and temporal asynchrony.

Language and Cognitive Processes, *22*(2), 234–246. https://doi.org/10.1080/01690960600630881

Pine, K. J., & Messer, D. (2003). The development of representations as children learn about balancing. *British Journal of Developmental Psychology*, *21*, 285–301.

Pine, K. J., Messer, D., & Godfrey, K. (1999). The teachability of children with naive theories: An exploration of the effects of two teaching methods. *British Journal of Educational Psychology*, *69*, 201–211.

Pyle, A., DeLuca, C., & Danniels, E. (2017). A scoping review of research on play-based pedagogies in kindergarten education. *Review of Education*, *5*(3), 311–351. https://doi.org/10.1002/rev3.3097

R Core Team. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing* (Version 3.6.2) [Computer software]. Vienna, Austria. URL https://www.R-project.org/.

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529–556. https://doi.org/10.1016/S0959-4752(01)00030-5

Reuter, T., & Leuchter, M. (2020). Children's concepts of gears and their promotion through play. *Journal of Research in Science Teaching*, *Advance online publication*. https://doi.org/10.1002/tea.21647

Richey, J. E., & Nokes-Malach, T. J. (2013). How much is too much? Learning and motivation effects of adding instructional explanations to worked examples. *Learning and Instruction*, *25*, 104–124. https://doi.org/10.1016/j.learninstruc.2012.11.006

Riley, W. F., & Sturges, L. D. (1993). *Engineering mechanics: Statics* (2nd ed.). Wiley.

Rubin, K. H., Fein, G. G., & Vandenberg, B. (1983). Play. In E. M. Hetherington (Ed.), *Handbook of child psychology* (pp. 693–774). Wiley.

Rubin, K. H., Watson, K. S., & Jambor, T. W. (1978). Free-Play behaviors in preschool and kindergarten children. *Child Development*, *49*(2), 534–536. https://doi.org/10.2307/1128725

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, *64*(6), 1617–1636. https://doi.org/10.2307/1131459

Sarama, J., & Clements, D. H. (2009). "Concrete" computer manipulatives in mathematics education. *Child Development Perspectives*, *3*(3), 145–150. https://doi.org/10.1111/j.1750-8606.2009.00095.x

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*(5), 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481–520.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analyses. *Cognitive Psychology*, *36*, 273–310.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modelling change and event occurence*. Oxford Univ. Press.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303–333. https://doi.org/10.1016/j.cogsci.2003.11.001

Therneau, T. (2015). *A Package for Survival Analysis in S*. (Version 2.38) [Computer software]. https://CRAN.R-project.org/package=survival

Thomson, D., Casey, B. M., Lombardi, C. M., & Nguyen, H. N. (2018). Quality of fathers' spatial concept support during block building predicts their daughters' early math skills – But not their sons'. *Early Childhood Research Quarterly*. *Advance online publication*. https://doi.org/10.1016/j.ecresq.2018.07.008

Thorsen, C., Gustafsson, J.-E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology, 84*(4), 556–570. https://doi.org/10.1111/bjep.12041

Trawick-Smith, J. (2015). Teacher–child play interactions to achieve learning outcomes. In K. C. Trundle & M. Saçkes (Eds.), *Research in early childhood science education* (pp. 259–277). Springer.

Trundle, K. C., & Saçkes, M. (2012). Science and early education. In R. Pianta (Ed.), *Handbook of early childhood education* (pp. 240–258). Guilford Publications.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

van Schijndel, T. J. P., Visser, I., van Bers, B. M. C. W., & Raijmakers, M. E. J. (2015). Preschoolers perform more informative experiments after observing theory-violating evidence. *Journal of Experimental Child Psychology*, *131*, 104–119. https://doi.org/10.1016/j.jecp.2014.11.008

Verdine, B. N., Zimmermann, L., Foster, L., Marzouk, M. A., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2019). Effects of geometric toy design on parent-child interactions and spatial language. *Early Childhood Research Quarterly*, *46*, 126–141. https://doi.org/10.1016/j.ecresq.2018.03.015

Weber, A. M., & Leuchter, M. (2020). Measuring preschool children's knowledge of the principle of static equilibrium in the context of building blocks: Validation of a test instrument. *British Journal of Educational Psychology*, *90*, 50–74. https://doi.org/10.1111/bjep.12304

Weinert, F.E., & Helmke, A. (1998). The neglected role of individual differences in theoretical models of cognitive development. *Learning and Instruction*, *8*(4), 309–323. https://doi.org/10.1016/S0959-4752(97)00024-8

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play. *Current Directions in Psychological Science*, *25*(3), 177–182. https://doi.org/10.1177/0963721416645512

Weisberg, D. S., Zosh, J. M., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Talking it up: Play, language development, and the role of adult support. *American Journal of Play*, *6*, 39–54.

Weiß, R. H., & Osterland, J. (2013). *CFT 1-R*. Hogrefe.

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Solis, S. L., & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology*, *9*, 1124. https://doi.org/10.3389/fpsyg.2018.01124

Article 3: Construction play promotes change in 5- to 6-year old children's science knowledge about stabilities and science self-concept[9]

Anke Maria Weber and Miriam Leuchter
Institute for Children and Youth Education, University of Koblenz-Landau, Germany

Abstract

Science education in kindergarten should promote children's knowledge acquisition and self-concept, which interact following the reciprocal effects model, and might be implemented through play. We investigated three types of construction play: (a) guided play (verbal and material scaffolds), (b) guided play (material scaffolds), (c) free play. We examined their effects on stability knowledge acquisition and self-concept as well as the reciprocal effects model's fit to kindergarten children. We implemented a pre-post-follow-up design, *N* = 183 5- to 6-year-olds (88 female). Both guided play groups outperformed the free play group in stability knowledge acquisition. Self-concept declined only in the free play group. The reciprocal effects model was not supported. Guided play may be effective in fostering children's stability knowledge and self-concept.

Keywords: guided play; science learning; self-concept; free play; scaffolding

1 Science education in the early years

In recent years, researchers, teachers, and politicians have shown increased interest in improving science education in the early childhood years, as early science learning is believed to influence later science achievement (Dunbar & Klahr, 2012; Trundle, 2015). One purpose of science education is the promotion of children's science knowledge acquisition (OECD, 2014).

Educational contexts that promote science knowledge acquisition integrate domain-specific knowledge and process knowledge (van der Graaf et al., 2020). The search for an appropriate domain to foster science knowledge must consider children's developmental constraints (Copple & Bredenkamp, 2009). Children have intuitive knowledge about stability from an early age on (e.g., Bonawitz et al., 2012). Research shows that kindergarten children consider an object's geometrical center when estimating its stability, thus rating symmetrical objects' stabilities correctly (Krist, 2010). However, kindergarten children face problems with estimating

---

[9] The article was submitted to an American journal and is therefore written in American English.

asymmetrical objects' stabilities (Krist, 2010), overgeneralizing the geometrical center and ignoring an object's center of mass. Thus, the stability domain might be an appropriate context to integrate domain-specific and process knowledge such as observing, testing presumptions, interpreting, generalizing, and reasoning (Klahr et al., 2011).

Bonawitz et al. (2012) asked children between 4 and 7 years to balance different symmetrical and asymmetrical blocks on a beam scale and categorized the children according to their balancing behavior. Children who tried to balance objects at their *center* and failed to do so with asymmetrical objects were categorized as center theorists; children who considered the *mass* and succeeded to balance all objects were categorized as mass theorists; and children who showed an *undifferentiated* pattern were categorized as having no theory. Their findings suggest that many kindergarten children do not consider the mass when balancing objects, which is in line with Krist et al. (2005) and Krist (2010). Additionally, Krist and colleagues found that the form of presentation is irrelevant, as children between 4 and 8 years showed the same performance when actively balancing symmetrical and asymmetrical objects on a beam scale (Krist et al., 2005), or when being presented with photographs (Krist, 2010).

Beyond knowledge acquisition, science education should consider the promotion of motivational components such as science self-concept (Belland et al., 2013; OECD, 2014; Pintrich, 2003). According to the reciprocal effects model (e.g., Marsh & Craven, 2006), academic success and corresponding academic self-concept influence each other. This relation has been identified even in young children (Guay et al., 2003). Therefore, a positive science self-concept may support children's engagement in science activities and their knowledge acquisition and thus result in a sense of achievement in science learning (Marsh et al., 2012).

Effective science education supports children by considering their developmental constraints as well as their motivation to learn about science phenomena by relating science to children's everyday activities, i.e., developmentally appropriate practice (Copple & Bredenkamp, 2009). One such activity that children engage in during free time at home or in kindergarten is construction play, e.g., block play (Borriello & Liben, 2018). Moreover, science learning can be supported through an adult's scaffolding (van de Pol et al., 2010). Scaffolding in early science education encompasses, inter alia, drawing children's attention to phenomena, supporting their active learning, guiding their explorations, and fostering their feeling of competence (Belland et al., 2013; Martin et al., 2019; van de Pol et al., 2010). Thus, play and scaffolding are believed to foster children's science knowledge and their trust in their science abilities.

1.1 Self-concept and its relation to knowledge acquisition

Shavelson et al. (1976) define self-concept as a person’s perception of themselves as shaped by the environment and their experiences. They conceptualize self-concept as hierarchical with academic and non-academic facets (Marsh et al., 1988). Moreover, the academic self-concept has been separated into domain-specific facets. Thus, a person’s science self-concept refers to their trust in their science abilities (Pintrich, 2003).

Academic self-concept has been construed as comprising motivational and competence beliefs subsumed under one respective domain-specific factor (Marsh et al., 2002). However, other theories model motivational beliefs and competence beliefs as separate constructs, which are considered to be related and to influence knowledge acquisition (e.g., Eccles, 2009; Kaplan et al., 2012). Arens et al. (2016) showed that self-concept can be differentiated into a motivational and a competence belief component even for kindergarten children. The motivational belief component can be conceptualized as the value a person attributes to a task (*How much do I enjoy building with blocks?*), while the competence belief component functions as an expectancy (*How good am I at building with blocks?*; see Eccles, 2009). The interaction of motivational and competence beliefs with achievement has been examined for kindergarten children (Arens et al., 2016), elementary school children and adolescents in mathematics (Guo et al., 2015; Lauermann et al., 2017) and STEM (Andersen & Ward, 2014; Ball et al., 2017).

In line with the multifaceted model of self-concept (Marsh et al., 1988; Shavelson et al., 1976), Harter (2015) summarizes that 5- to 7-year-old children typically focus on specific competences (e.g., social skills, cognitive abilities) when asked about their self-concept. Correspondingly, Marsh et al. (2002) found that measuring kindergarten children’s self-concept is possible, as even young children can differentiate between (academic) aspects of the self (e.g., verbal skills versus mathematics). However, Helmke (1999) found that when comparing themselves to their peers, children’s academic self-concept are generally unrealistically high during the kindergarten and early primary school years (Eccles, 2009; Harter, 2015).

Science success and science self-concept are intertwined and influence each other, following the reciprocal effects model (Marsh & Craven, 2006). According to Shavelson et al. (1976), positive experiences can enhance self-concept. This relation has been demonstrated in studies on general school achievement, mathematics, and language, mainly for primary and high school students (e.g., Guay et al., 2003; Marsh & Martin, 2011). However, few studies have examined this relation in the science domain (Jansen, 2017). Some findings indicate a positive relation between physics achievement and physics self-concept for middle school students (Jansen, 2017; Möller et al., 2006). Denissen et al. (2007) investigated the development of the relation between achievement and self-concept in science and mathematics for children and adolescents between the ages 6 to 17 and found that they were related for all ages.

As an increase in academic self-concept can result in an increase in achievement and vice versa, designing interventions that help children engage in science activities and promote positive experiences and a sense of achievement is crucial for young children’s knowledge acquisition. In a meta-analysis of the effects of interventions on children’s self-concept, integrating studies with children up to the age of 18, O'Mara et al. (2006) found that interventions targeting specific skills such as science skills also affected the corresponding self-concept. Marsh et al. (2012) hypothesize that interventions targeting achievement, while undermining academic self-concept, will likely only temporarily affect achievement. Samarapungavan et al. (2011) found that an intervention on science learning with a teacher’s verbal support promoted kindergarten children’s knowledge of biological phenomena and their biology self-concept. Children in the intervention group had gained more science knowledge and had a higher science self-concept compared to a control group that had received their kindergarten’s traditional science education. However, Samarapungavan et al. (2011) did not investigate the relation between science knowledge acquisition and science self-concept. Marsh and Richards (1988) found that an intervention aiming at enhancing adolescents’ math and reading self-concept positively affected both self-concept as well as adolescents’ math and reading achievements. However, Craven et al. (1991) found that feedback on positive abilities and performance enhanced reading and math self-concept in primary school children, but not their achievement. Despite these ambiguous results, these studies indicate that enhancing science self-concept is possible and that interventions targeting science knowledge acquisition might also enhance corresponding self-concept and vice versa (Marsh & Richards, 1988; Samarapungavan et al., 2011).

However, research on the relation between science knowledge and science self-concept and their possible promotion for kindergarten children is sparse, although many psychologists view self-concept as a pillar of children’s development (Marsh et al., 2002; Marsh et al., 2012). Children’s self-concept is fostered through positive experiences that may be gained during play (Trawick-Smith, 2012). Accordingly, a teaching approach that encompasses play is considered to promote academic and motivational development through a positive interaction with an educational content (Zosh et al., 2018).

1.2 Play

Play is understood as one way of implementing developmentally appropriate practice (Copple & Bredenkamp, 2009; Trawick-Smith, 2012). It is considered voluntary, intrinsically motivating, child-directed, process- rather than goal-oriented, and as containing elements of choice (Pellegrini, 2013; Rubin et al., 1983; Trawick-Smith, 2012). Nevertheless, definitions of play comprise elements that both overlap and exclude each other. Some researchers conceive play as a category necessarily comprising attention to a process rather than a goal and not serving a certain purpose (Pellegrini, 2013). Others consider play a continuum in which the above aspects might be realized to a greater or lesser extent (e.g., Borriello & Liben, 2018; Fisher et al., 2013; Rubin et al., 1983; Weisberg et al., 2016).

Pellegrini (2013) claims that block play should not be considered play because construction is goal-oriented and not mainly concerned with the process. However, the view of play as a continuum suggests that an activity might be considered play even if some aspects are only partly fulfilled and thus allows activities such as block play to be considered playful (Zosh et al., 2018). Accordingly, Rubin et al. (1983) questioned if construction play, e.g., block play, is necessarily concerned with the end product rather than the process of building. We follow the continuum definition for the present study and consider block play as play.

Extending the continuum-view, Zosh et al. (2018) define play as a spectrum with different types of play such as free play, which satisfies all characteristics named earlier, and guided play, which is not without goals but can also be process-oriented. Accordingly, guided play can be defined as a playful activity initiated by an adult with a learning goal, but the activity itself is directed by a child. Fisher et al. (2013) showed that guided play as an amalgamation of instructional aspects and free play might be effective for fostering children’s learning. Furthermore, guided play might maintain children’s motivation for learning science and their positive science self-concept (Trawick-Smith, 2012). Guided play allows children to choose their focus and shape the activity, engaging in play voluntarily and directing it (Fisher et al., 2013).

Concerning the adult’s role in guided play, intervening might reduce children’s motivation for play (Bonawitz et al., 2011). However, without interventions, children might not learn (Stipek et al., 1995), raising the question how an adult’s support should be implemented to foster children’s knowledge acquisition during play. According to models of science education (van de Pol et al., 2010), an adult may help children master challenging tasks and acquire new insights through scaffolding, which might also enhance children’s science self-concept (Samarapungavan et al., 2011). Based on the idea of scaffolding as an effective way of support, we focus on two elements: (a) material scaffolds, e.g., in the form of structured learning materials, and (b) scaffolding through verbal support (Martin et al., 2019; van de Pol et al., 2010).

Incorporating scaffolding elements into guided play may promote young children’s learning (Fisher et al., 2013; Weisberg et al., 2016). Moreover, scaffolding may maintain or enhance children's motivational and competence beliefs (Guthrie et al., 2004). Material scaffolds are effective, if they link new content to prior knowledge and draw attention to specific aspects essential for understanding (DeLoache, 2014). Verbal support can promote the learning process (van de Pol et al., 2010) and support motivational and competence beliefs (Belland et al., 2013; Guthrie et al., 2004). It is also important for heterogeneous groups of kindergarten children (Weisberg et al., 2016). Verbal support can be implemented through scaffolding techniques, such as modeling, activation of prior knowledge, explanations, encouraging comparisons, asking for reasoning, and promoting perceptions of challenge, competence, and success (e.g., Belland et al., 2013; van de Pol et al., 2010).

Modeling provides a learning opportunity for young children (Vygotsky, 1978). It refers to a teacher performing certain behaviors and thinking styles that deliberately offer learners an opportunity for imitation. In science education the teacher may model how to observe certain science phenomena (Leuchter & Naber, 2019). *Activation of prior knowledge* is a crucial supporting strategy for children's science learning, targeting higher order cognitive processes (Leuchter & Naber, 2019; Richey & Nokes-Malach, 2013). It describes a teacher referring to knowledge that a learner has already acquired and therefore supports the learner in integrating new aspects into existing schemata. Children's prior knowledge might be activated when testing their presumptions. *Providing explanations* allows learners to coordinate observations with an evidence-based interpretation of a phenomenon. A teacher may provide possible explanations, helping learners structure cognitive processes and organize knowledge (Richey & Nokes-Malach, 2013). *Encouraging comparisons* supports the identification of relational similarities or differences between entities. A teacher may encourage comparisons through presenting different entities that were chosen to highlight certain features to help the learner generalize the underlying concepts (Hsin & Wu, 2011; Richey & Nokes-Malach, 2013). *Asking for the learner's reasoning* allows them to question and structure their prior knowledge and thinking processes (Hsin & Wu, 2011). The teacher may ask for reasoning through questions urging learners to justify their assumptions. Motivational and competence beliefs are supported through scaffolding as well (Belland et al., 2013). *Promoting the perception of challenge* by expressing the belief that learners are able to solve the task may enhance learners' expectancy of success (Britner & Pajares, 2006). *Referring to the learner's competence* highlights their achievements and strategies (Kaplan & Maehr, 2007). *Referring to a learner's success* and encouraging them to explain their strategy may invoke a sense of pride (Pintrich, 2003).

Leuchter and Naber (2019) found that a combination of structured learning materials and verbal scaffolds supported 6- to 7-year-old children's learning in a physics domain better than only materials, only verbal support, or free exploration. Furthermore, scaffolding might help children acquire domain-specific knowledge and enhance their science self-concept during a playful activity such as block play. Block play offers the opportunity to implement a variety of scaffolds (Borriello & Liben, 2018). Studies on block play found that it can foster children's science knowledge. For example, 5- and 6-year-old children were asked to balance symmetrical and asymmetrical blocks on a beam scale and to explain the blocks' behavior (Pine & Messer, 2003). Afterwards, the children played with the blocks freely. A few days later, a posttest was conducted, which showed that the children were able to correctly balance more blocks compared to the pretest.

Furthermore, some research has examined the effects of block play on spatial language. Borriello and Liben (2018) investigated the effects of mothers' spatial talk during block play on their 4- to 6-year-old children's spatial language. The children and their mothers first played with blocks freely and subsequently, they engaged in guided construction play. If the mothers were introduced to spatial talk beforehand, both mothers and their children used more spatial language during the initial free play and the subsequent construction play than if they had not received such introduction. In another study, Ferrara et al. (2011) obtained similar results.

Guided play may thus be an effective way of supporting kindergarten children's knowledge acquisition and their science self-concept through a combination of child-directed activities and an adult's scaffolds (Fisher et al., 2013; Samarapungavan et al., 2011). Therefore, we chose block play in combination with scaffolds to support children's stability knowledge concerning the importance of the mass—as a part of children's science knowledge—and their science self-concept. However, it remains unclear whether guided play is more effective in fostering mass knowledge and science self-concept than free play and whether the effectiveness of guided play varies with the implementation of either material or material and verbal scaffolds. Given this research gap, we focus on studying the impact of play on the acquisition of mass knowledge and changes in science self-concept.

1.3 Research questions

This study investigates three research questions with three different types of construction play: guided construction play with material and verbal scaffolds, guided construction play with material scaffolds, and free construction play. We investigate the effect of construction play on children's mass knowledge acquisition and science self-concept.

Research question 1: Do 5- to 6-year-old children's acquisition of mass knowledge and their science self-concept, i.e., motivational and competence beliefs regarding science, differ between the three different play settings (Borriello & Liben, 2018; Ferrara et al., 2011; Leuchter & Naber, 2019; Marsh & Richards, 1988; Samarapungavan et al., 2011)?

Hypotheses:

(a) Children in the guided construction play group with material and verbal scaffolds have higher mass knowledge gains than children in the guided construction play group without verbal scaffolds.

(b) Children in both guided construction play groups have higher mass knowledge gains than children in the free construction play group.

(c) Children’s motivational beliefs in the guided play group with material and verbal scaffolds are higher than children’s motivational beliefs in the guided play group with material scaffolds.

(d) Children’s motivational beliefs in the guided play group with material and verbal scaffolds are higher than children’s motivational beliefs in the free play group.

(e) Children’s competence beliefs in the guided play group with material and verbal scaffolds are higher than children’s motivational beliefs in the guided play group with material scaffolds.

(f) Children’s competence beliefs in the guided play groups are higher than children’s competence beliefs in the free play group.

Research question 2: Are there reciprocal effects between mass knowledge and motivational or competence beliefs (Denissen et al., 2007; Guay et al., 2003; Jansen, 2017)?

Hypotheses:

(a) There are reciprocal effects between mass knowledge and motivational beliefs.

(b) There are reciprocal effects between mass knowledge and competence beliefs.

(c) There are group differences in the reciprocal effects.

2 Method

2.1 Participants

In total, 183 children from Germany (88 female) aged 5 to 6 years, *M* = 5.55, *SD* = 0.50, participated in the study. A total of 172 children were of European descent, 1 of Central American descent, 2 of African descent, and 9 of Asian descent. The participants visited 23 kindergartens (*N* = 2 to 13 per kindergarten), which were located either in villages (700 to 3,000 inhabitants; *N* = 83 children), small cities (less than 20,000 inhabitants; *N* = 10 children) or medium sized cities (approximately 50,000 inhabitants; *N* = 91 children). The sample was randomly collected and consisted of children, who had not received any formal education. All kindergartens had building blocks that the children were free to play with. All children participated voluntarily and with their parents’ consent, which was obtained in written form.

2.2 Procedure

The study adopted a pre-post-follow-up design with two guided play groups and a free play group. The pretest (T1) was administered approximately two weeks before the one-hour play session and the posttest (T2). The follow-up (T3) occurred approximately ten weeks after the posttest. For each of the three measurement points, the children completed a test for mass knowledge, which was assessed in a group of up to six children; and a single interview assessing science self-concept. The group test lasted 5 minutes and the single interview lasted 10 minutes at each measurement point. For the group procedure, the children were either seated back-to-back, or a screen was placed between them to prevent them from copying from one another. During testing, breaks were permitted whenever a child or the experimenter considered them necessary.

The children were assigned to one of three different intervention groups by parallelizing them according to their language capacity, which had been assessed at pretest, resulting in triplets with the same language capacity. For example, a child with a language capacity of T = 50 was paired with two other children with a language capacity of T = 50 and then each of the children was assigned to one of the play groups. The two guided play groups differed in the scaffolding they received. The *Verbal group* (N = 64, 27 girls) played a guided construction play with provided materials and additionally received verbal scaffolds, the *Material group* (N = 59, 32 girls) played a with the same materials, and the *Free play group* (N = 61, 29 girls) played with blocks freely. In most German kindergartens, kindergarten teachers will not teach learning contents over an extended period of time. Thus, we attempted to achieve ecological validity through implementing the play for each group during approximately one hour. Moreover, the stability domain is rather small and mass knowledge might be acquired in a rather short amount of time (Bonawitz et al., 2012). The guided construction play and the free play were led by one of six female experimenters. To avoid experimenter effects, play groups were varied systematically for all experimenters, who led all intervention groups, i.e., Verbal, Material, and Free play group, and who had been trained accordingly (Weber et al., 2020). In addition, the play sessions were recorded as a manipulation check.

2.3 Construction play

In all three play groups, the experimenter praised children’s efforts and encouraged them to try again when they encountered problems. In the guided play groups, children could choose which construction they wanted to build first and if they wanted to build with a friend or on their own. Furthermore, the children were allowed to stop playing completely or take a break as they desired (Rubin & Smith, 2018; Weisberg et al., 2016).

Children in the Verbal group and the Material group received identical materials for the guided construction play (Martin et al., 2019), photographs of different block constructions, which varied in the number of blocks and complexity. Each photograph went with a small box with the corresponding building blocks inside. The blocks varied in shape (cuboids, triangles, etc.), size, and color (brown, black, yellow, red, and green). Furthermore, by using asymmetrical as well as symmetrical block structures, we aimed at drawing children’s attention to differences in the stability of symmetrical compared to asymmetrical constructions (Hsin & Wu, 2011). The materials were developed prior to the study and tested in play sessions with children to ensure that they could rebuild the structures shown on the photographs and had fun playing with the materials.

Five different activities were played in a standardized order, and the children received the instructions presented in Table 1. The Material group did not receive additional instructions.

The Verbal group received verbal scaffolds to evoke children’s observing, testing of their presumptions, interpreting and generalizing evidence, and reasoning about science phenomena as well as to support their motivational and competence beliefs. These skills were fostered through modeling, activating prior knowledge, encouraging comparisons, providing explanations, and asking questions or asking for reasoning; motivational and competence beliefs were fostered by promoting perceptions of challenge, competence and success (Table 2; Belland et al., 2013; van de Pol et al., 2010). If children asked for help, the experimenter helped with building in the Verbal group through stabilizing the child’s building by holding a block in place. In the other two groups, the experimenter did not assume a teaching role and declined in a friendly manner by stating that she unfortunately could not help the children and suggested that the child could ask another child for help with building. For a complete presentation of all material and verbal scaffolds, please refer to Weber et al. (2020).

The Free play group received the same blocks as the other two groups did; however, the building blocks were unstructured and were provided in a large wooden box. The children only received the instruction to play with the blocks freely. During play, the experimenter did not intervene.

Table 1

Material scaffolds in both guided play groups

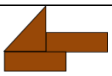
Play	Instruction	Example
Black block (11 photographs)	You can build the building shown on the photograph. Build the building and guess, if the blocks remain stable or tumble.	
Add-a-block (8 photographs)	The blocks on the photos were bewitched so they would remain stable. Can you rebuild the building, so that it is stable? (If a child did not succeed, the experimenter provided a green block:) Look, here is a green block. Try to stabilize the building with it.	
Sliding (9 photographs)	First, you may rebuild the building on the photograph. Then you slide the upper block along the lower one, until it falls (experimenter models it). That makes noise.	
Rebuild (11 photographs)	You can just rebuild the building on these photographs and see how well you are doing. Some buildings are very easy to rebuild; others are more difficult. But every single one will remain stable if built correctly.	
Stable/Tumble (8 photographs)	The buildings on the photographs will remain stable sometimes, but at other times, the blocks were bewitched. Look at the photograph and say "Stable" or "Tumble" and then try out to see whether you were correct.	

Table 2

Scaffolding techniques used in the Verbal group

Technique	Example
Modeling	Look! (Experimenter looks very closely)
Activating prior knowledge	Have you ever seen something like this?
Encouraging comparisons	Your building looks different than [another child’s building], doesn’t it? What is different? Is something similar?
Providing explanations	Well done! If the heavy side of a block hovers in midair, the block will tumble.
Asking for reasons	Can you explain this in more detail, so I can really understand what you think?
Challenge	This is quite difficult to build. I am sure that you can do it!
Competence	That didn’t work out. You need to stabilize this first. How could you do that?
Success	That’s a really good solution! Now the blocks remain in place. Why don’t you tell us how you did this?

2.4 Measures

Mass knowledge. The Center-of-Mass Test (COM Test; Plöger, 2020) assesses children’s mass knowledge with 16 items. The test consists of asymmetrical block structures (Figure 2) that can only be rated correctly by applying mass knowledge. Children with center knowledge will rate all or most items incorrectly, as the geometrical center of the red blocks is always supported, when the center of mass is not supported, and vice versa.

First, the children received a small booklet. The experimenter introduced the test setting by rebuilding the warm-up picture, which was a symmetrical block structure. Next, the response format was introduced to the children, and the children were asked to rate the block constructions’ stabilities by circling a stable structure and crossing out an unstable structure.

Science self-concept. Motivational beliefs for learning how to balance blocks and competence beliefs concerning building block structures were assessed with a standardized single interview adapted from the Young Children’s Science Motivation scale (Y-CSM; Oppermann et al., 2017). Motivational beliefs were assessed with 5 items by asking, e.g., *Do you enjoy learning about building with building blocks or not? Please show me how much you enjoy building with blocks. Not at all, a little, much, very much*. Competence beliefs were assessed with 6 items, e.g., *Do you know much or not so much about building with blocks? Please show me how much you know about building with blocks. Nothing at all, a little, much, very much*. Each of the questions was followed by prompting the children to indicate how much they agreed with each question on a separate sheet of paper showing a diagram of increasing size from 0 (*not at all*) to 3 (*very much*).

Figure 2. Example items of the COM Test. From left to right: Unstable 1a, Unstable 1b, Stable 1a, Stable 1b

2.5 Data analysis

The statistics program R, version 4.0.3 (R Core Team, 2020), was used for data analyses. First, we investigated descriptive statistics and correlation patterns between the three measurement points, and the structure of science self-concept and its motivational and competence facets. Then, we investigated changes in mass knowledge and motivational and competence beliefs as well as possible group differences from T1 to T3 with mixed-effects growth models. Finally, we specified two cross-lagged panel models to examine the longitudinal relations between mass knowledge and motivational and competence beliefs, respectively.

Missing values do not pose a threat to the results of mixed-effects models, which is why we did not take specific measures to deal with missingness for the mixed-effects models (Singer & Willett, 2003). For the CFA and cross-lagged panel models, we used the full information maximum likelihood (FIML) estimation to deal with missingness.

3 Results

3.1 Primary statistical analyses

Descriptive statistics, correlations, and group differences at pretest. Descriptive statistics at each measurement point are presented by condition in Table 3. Cronbach's α was good or satisfactory for all scales.

Correlations at the sample level are presented in Table 4. The motivational component of science self-concept at T1 was negatively correlated with mass knowledge at T3, $r = -.18$, $p = .040$. We found no correlations between mass knowledge and competence beliefs.

Then, we checked for group differences at T1. ANOVAs showed no group differences for any of the measures at T1; mass knowledge, *F*(2, 161) = 1.11, *p* = .331; motivational beliefs, *F*(2, 172) = 0.36, *p* = .698; competence beliefs, *F*(2, 169) = 0.91, *p* = .404.

To investigate whether motivational and competence beliefs can be construed as two different facets of science self-concept, we computed two CFAs with FIML estimation with motivational and competence beliefs at T1 either loading on one or two latent factors. The CFA with a single latent factor showed a poor fit, $\chi^2 = 114.70$, *df* = 44, *p* < .001, CFI = .79, SRMR = .08, RMSEA = .10, *p* < .001. The CFA with two correlated latent factors showed a good fit, $\chi^2 = 68.67$, *df* = 43, *p* = .008, CFI = .92, SRMR = .06, RMSEA = .06, *p* = .256. The latent factors motivational and competence beliefs were correlated, *r* = .60, *p* < .001. A model comparison implied that the model with two latent factors explained the data better than the model with a single latent factor, $\Delta\chi^2 = 40.38$, *df* = 1, *p* < .001. Therefore, motivational and competence beliefs were investigated as two independent constructs.

Table 3

Descriptive statistics by condition

	Verbal			Material			Free play				
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*Range*	α
Mass T1	58	5.84	3.18	52	6.46	3.30	54	5.52	3.42	0–16	.71
Mass T2	55	7.84	4.10	56	7.27	3.53	49	6.14	3.58	0–16	.78
Mass T3	48	8.31	3.70	41	8.39	4.55	47	6.89	3.65	0–16	.80
MB T1	62	2.13	0.76	55	2.24	0.73	58	2.13	0.85	0–3	.75
MB T2	60	2.05	0.93	53	2.14	0.81	52	1.87	0.98	0–3	.83
MB T3	47	2.07	0.88	41	2.28	0.81	49	1.75	0.98	0–3	.82
CB T1	61	2.17	0.65	54	2.15	0.77	57	2.31	0.69	0–3	.78
CB T2	60	2.14	0.75	53	2.14	0.68	53	2.14	0.79	0–3	.78
CB T3	47	2.09	0.78	40	2.30	0.73	48	1.96	0.87	0–3	.84

Notes. α = Cronbach’s α. Mass = mass knowledge. MB = motivational beliefs. CB = competence beliefs.

The descriptive statistics in Table 3 suggest that the Verbal group experienced the highest knowledge gain in mass knowledge and that the motivational and competence beliefs in the Verbal and the Material group remained constant over time, while declining in the Free play group. To investigate this further, we considered a possible multilevel structure of the data.

Table 4

Correlations between the constructs at each time of measurement

	Mass T1	Mass T2	Mass T3	MB T1	MB T2	MB T3	CB T1	CB T2
Mass T2	.22**							
Mass T3	.00	.53***						
MB T1	.01	−.08	−.18*					
MB T2	.06	.00	−.03	.58***				
MB T3	.00	.09	.00	.39***	.62***			
CB T1	.01	−.07	−.10	.48***	.39***	.25**		
CB T2	.06	−.04	.03	.37***	.61***	.48***	.51***	
CB T3	.01	.14	.05	.35***	.53***	.69***	.41***	.61***

Notes. Mass = mass knowledge. MB = motivational beliefs. CB = competence beliefs.
*$p < .05$. **$p < .01$. ***$p < .001$.

3.2 Research question 1: Group differences in changes in mass knowledge and science self-concept

We examined the change in mass knowledge and science self-concept, and possible group differences from T1 to T3. Differences between children explained 22% of the variance in changes in mass knowledge, 52% in motivational beliefs, and 52% in competence beliefs. Kindergarten explained 1% of the variance in changes in mass knowledge, 7% in motivational beliefs, and 5% in competence beliefs. This indicates that the points of measurement are nested in children, but not in the kindergarten that the children visited. Thus, we specified three multilevel models with children on level-2 and included time as a random effect.

Children in both guided play groups gained mass knowledge from T1 to T3, Verbal, $\gamma_{11} = 1.28$, $p < .001$; Material, $\gamma_{12} = 1.02$, $p = .008$. The Free play group did not improve their mass knowledge from T1 to T3, $\gamma_{13} = 0.68$, $p = .065$. Subsequently, we examined hypotheses 1a and 1b concerned with group differences in change. Group differences in change were non-significant, ΔFree Play–Verbal, $p_{\text{one-tailed}} = .121$; ΔFree play–Material, $p_{\text{one-tailed}} = .259$; ΔVerbal–Material, $p_{\text{one-tailed}} = .310$. However, we found group differences at T2, directly after the intervention between the guided play groups and the Free play, ΔFree Play–Verbal, $p_{\text{one-tailed}} = .008$; ΔFree play–Material, $p_{\text{one-tailed}} = .009$, but not between the two guided play groups, ΔVerbal–Material, $p_{\text{one-tailed}} = .493$.

Motivational beliefs remained stable in both guided play groups from T1 to T3, Verbal, $\gamma_{11} = -0.03$, $p = .646$; Material, $\gamma_{12} = -0.05$, $p = .506$, but declined in the Free play group from T1 to T3, $\gamma_{13} = -0.19$, $p = .004$. Hypotheses 1c and 1d are concerned with group differences in change between the intervention groups. The group differences in change in motivational beliefs were significant between the Verbal group and the Free play group, ΔFree Play–Verbal, $p_{\text{one-tailed}} = .040$, but not between the guided play groups, ΔVerbal–Material, $p_{\text{one-tailed}} = .432$, or the Material group and the Free play, ΔFree play–Material, $p_{\text{one-tailed}} = .064$. We found group differences at T2 between the Material group and the Free play group, ΔFree play–Material, $p_{\text{one-tailed}} = .027$, but not between the Verbal group and the Free play group, ΔFree Play–Verbal, $p_{\text{one-tailed}} = .111$, or between the guided play groups, ΔVerbal–Material, $p_{\text{one-tailed}} = .228$.

Competence beliefs remained stable in both guided play groups from T1 to T3, Verbal, $\gamma_{11} = -0.04$, $p = .448$, and Material, $\gamma_{12} = 0.04$, $p = .550$, but declined in the Free play group from T1 to T3, $\gamma_{13} = -0.18$, $p = .002$. Hypotheses 1e and 1f are concerned with group differences in change between the intervention groups. Indeed, the group differences in change in competence beliefs were significant between the guided play groups and the Free play group, ΔFree Play–Verbal, $p_{\text{one-tailed}} = .042$; ΔFree play–Material, $p_{\text{one-tailed}} = .005$, but not between the guided play groups, ΔVerbal–Material, $p_{\text{one-tailed}} = .170$. We found no group differences at T2, ΔFree Play–Verbal, $p_{\text{one-tailed}} = .490$; ΔFree play–Material, $p_{\text{one-tailed}} = .348$; ΔVerbal–Material, $p_{\text{one-tailed}} = .355$.

3.3 Research question 2: Testing reciprocal effects

To address the second research question and hypotheses 2a and 2b concerned with the relation between mass knowledge and science self-concept, we specified two cross-lagged panel models (Figures 3 and 4). A single cross-lagged panel model integrating both components of science self-concept might face problems with multicollinearity and thus results might be unreliable. Since, motivational and competence beliefs can be construed as two distinct facets of science self-concept, we tested the reciprocal effects between mass knowledge and motivational or competence beliefs in separate models following Arens et al. (2016).

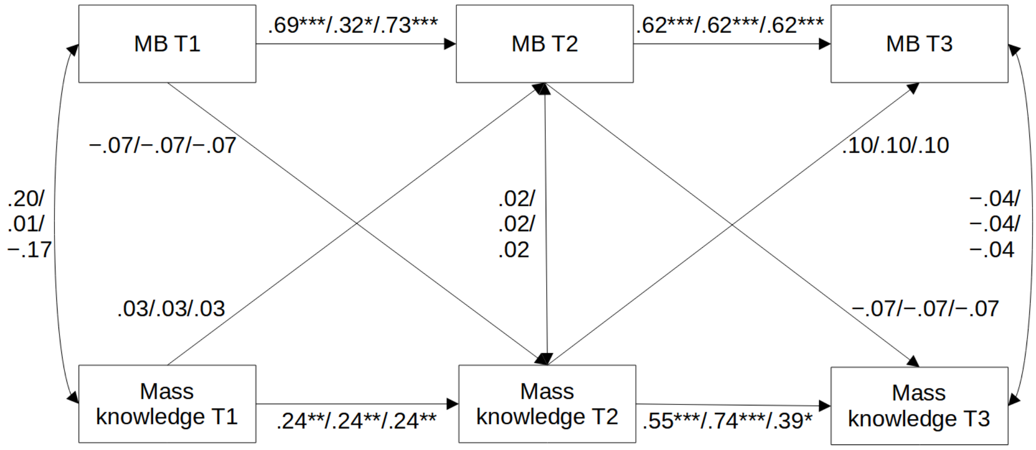


Figure 3. Cross-lagged panel model for the reciprocal effects between mass knowledge and motivational beliefs

Notes. MB = motivational beliefs. Coefficients before the slashes refer to the Verbal group; coefficients between the slashes refer to the Material group; coefficients behind the slashes refer to the Free play group. **p* < .05, ***p* < .01, ****p* < .001.

Hypothesis 2c is integrated in these two models and was concerned with group differences between the relation of mass knowledge and science self-concept. To test for group differences in cross-lagged panel models, the model with the reciprocal and auto-regressive effects constrained across groups was compared to a model with the same parameters estimated freely with a $\Delta\chi^2$-test. This analysis involved two steps. First, autoregressive and reciprocal effects between mass knowledge and motivational and competence beliefs were compared across groups. Then, the model with the best fit to the data was specified. All variables were *z*-standardized.

The two final models are presented in Table 5. The autoregressive effects show high stability estimates for mass knowledge and motivational and competence beliefs for all three play groups. There were no group differences in stability of mass knowledge between T1 and T2, however, the stability between T2 and T3 was lower for the Free play group than the Material group, *p* = .026. Stability of motivational beliefs (Figure 3) from T1 to T2 was lower in the Material than in the Verbal group, *p* = .010, and in the Free play group, *p* = .011. Furthermore, the stability of competence beliefs (Figure 4) from T1 to T2 was lower in the Material groups than in the Free play group, *p* = .027.

The reciprocal analyses indicate no evidence for reciprocal effects between mass knowledge and motivational or competence beliefs, except for the effect of mass knowledge at T2 on competence beliefs at T3 in the Free play group, β = .25, *p* = .044.

Table 5

Standardized path coefficients of the path models

	Motivational beliefs			Competence beliefs		
	Verbal	Material	Free play	Verbal	Material	Free play
Mass knowledge T1 → Mass knowledge T2	.24**	.24**	.24**	.24**	.24**	.24**
Mass knowledge T2 → Mass knowledge T3	.55***	.74***	.39*	.55***	.74***	.39*
Mass knowledge T1 ~ SC T1	.20	.01	−.17	.02	.02	.02
Mass knowledge T1 → SC T2	.03	.03	.03	.03	.03	.03
Mass knowledge T2 ~ SC T2	.02	.02	.02	−.03	−.03	−.03
Mass knowledge T2 → SC T3	.10	.10	.10	.11	.16	.25*
Mass knowledge T3 ~ SC T3	−.04	−.04	−.04	−.05	−.05	−.05
SC T1 → SC T2	.69***	.32*	.73***	.59***	.33**	.71***
SC T2 → SC T3	.62***	.62***	.62***	.65***	.65***	.65***
SC T1 → Mass knowledge T2	−.07	−.07	−.07	−.04	−.04	−.04
SC T2 → Mass knowledge T3	−.07	−.07	−.07	−.01	−.01	−.01

Notes. SC = science self-concept, either motivational or competence beliefs. *p < .05. **p < .01. ***p < .001.

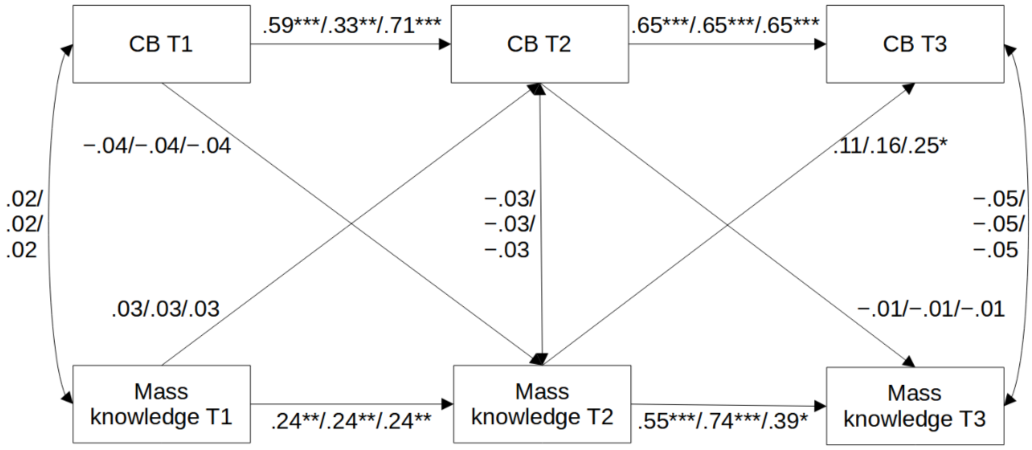


Figure 4. Cross-lagged panel model for the reciprocal effects between mass knowledge and competence beliefs

Notes. CB = competence beliefs. Coefficients before the slashes refer to the Verbal group; coefficients between the slashes refer to the Material group; coefficients behind the slashes refer to the Free play group. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

4 Discussion

Science education should start in the early years of childhood and encompass the promotion of science knowledge with regard to motivational aspects (OECD, 2014; Trundle, 2015). Therefore, it is important to examine whether it is possible to foster both domain-specific knowledge and science self-concept in kindergarten science learning. Studies on promoting science learning in the early years are sparse, and the question concerning the best approach to teaching science in the kindergarten years remains unanswered (Trundle, 2015).

Thus, we conducted this study to investigate whether 5- to 6-year-old children’s knowledge acquisition in a specific science domain, i.e., mass knowledge, as well as children’s science self-concept can be supported through play. We conducted an experiment with building blocks in three groups, which differed in the way play was realized. The Material group engaged in a guided construction play with provided materials, the Verbal group played with the same materials and additionally received verbal scaffolds, and the Free play group played with blocks freely. Our findings contribute to the literature on science education in the kindergarten years.

Research question 1. Over the course of 12 weeks, mass knowledge increased in both guided play groups, but not in the Free play group. Moreover, directly after the intervention, the Free play group showed less gain in mass knowledge than the Verbal and Material groups. Thus, we may assume that the guided play conditions enhanced children’s mass knowledge, while free play did not contribute to a knowledge gain (Stipek et al., 1995). These results show that mass knowledge can be promoted.

In both guided play groups, material scaffolds were employed. For knowledge acquisition after a very brief period, scaffolding materials seem to be crucial. The effect of verbal scaffolds was investigated in combination with material scaffolds. The fact that no additional influence of the verbal scaffolds was discovered invites four possible interpretations. (1) As the play only lasted approximately one hour and only took place once, additional interventions over a longer period could uncover a possible difference between materials and materials with verbal scaffolds (Leuchter & Naber, 2019). (2) The scaffolding was not controlled for its adaptivity, which could be crucial for learning. Longer and more adaptive interventions might improve children’s mass knowledge further and could reveal possible influences of verbal scaffolds, if individual variables are controlled for. (3) The guided play materials might have been so self-explanatory and low threshold that the children did not require an adult’s scaffolds in addition to the materials to gain new knowledge (Martin et al., 2019). (4) We aimed to measure children’s mass knowledge and accordingly measured their knowledge with the COM Test, a non-verbal instrument (Krist, 2010). Therefore, children were not required to explain their reasoning behind their understanding of mass. In a study, in which children were asked about their underlying reasoning for determining objects’ stabilities, the Verbal group had an advantage over the Material group (Weber et al., 2020).

Children’s motivational and competence beliefs at pretest were high in all three groups, which is consistent with previous research on kindergarten children’s academic self-concept (Eccles, 2009; Harter, 2015; Helmke, 1999). However, motivational and competence beliefs decreased in the Free play group, while remaining stable in both guided play groups.

Possibly, the high initial motivational and competence beliefs prevented an increase over the 12-week period. Guided construction play with scaffolds might have stabilized children’s motivational and competence beliefs. This stabilization was presumably supported by the challenges posed by the guided play conditions that allowed children to gain new experiences with block building (Shavelson et al., 1976). Since we made specific efforts to vary the number of blocks used in the scaffolding materials, the children likely had positive mastery experiences. This might have prevented developmentally determined decreases in science self-concept (Marsh et al., 2012; Samarapungavan et al., 2011; Shavelson et al., 1976). In the Free play group, however, the children might have experienced many failures, as their buildings tumbled because they did not know how to stabilize them. Thus, the developmentally determined decrease in children’s motivational and competence beliefs was not prevented (Harter, 2015).

Ideally, science education occurs in contexts that children are familiar with from their everyday lives, e.g., block play (Copple & Bredenkamp, 2009). Our results show that 5- to 6-year-olds can acquire knowledge of familiar science phenomena if they are presented with these phenomena in ways that they can comprehend and process.

In the guided construction play groups, we implemented process knowledge by drawing children’s attention to the mass and by guiding their explorations with material and verbal scaffolds (Martin et al., 2019). Thus, children’s process knowledge, observing, testing their presumptions, interpreting, generalizing and reasoning, might have also been stimulated (van de Pol et al., 2010). However, we did not assess whether children used more process knowledge in the guided play groups compared to the Free play group, as we were primarily interested in domain-specific knowledge acquisition and science self-concept, not in process knowledge. Nonetheless, this question should be investigated in the future. Furthermore, we implemented a variety of verbal scaffolds without comparing the different verbal scaffolds to one another in terms of their effects on mass knowledge and science self-concept, which could be an additional focus of a future study.

The guided play conditions supported children’s mass knowledge acquisition. These results are consistent with those of other studies on guided block play concerned with spatial talk (e.g., Borriello & Liben, 2018; Ferrara et al., 2011) and extend these findings by considering self-related developmental aspects such as science self-concept.

Play is often characterized as containing elements of choice and as being voluntary, child-directed, intrinsically motivating, and process-oriented (Pellegrini, 2013; Rubin et al., 1983). Thus, we discuss whether and how our guided play implementation can be considered play based on (1) the use of building blocks and (2) the use of scaffolds.

(1) Play is sometimes considered categorical, with an activity being either play or not play (Pellegrini, 2013). Pellegrini (2013) does not consider construction play to be free of goals and therefore does not categorize it as play. However, following Borriello and Liben (2018), Ferrara et al. (2011), Fisher et al. (2013), Rubin et al. (1983), and Trawick-Smith (2012), construction play might be considered play because children playing with building blocks seldom seem to have a goal in mind. Even though they articulate goals such as building a tower, their goals may change quickly, e.g., from building a tower to building a house. Therefore, the process of building seems to be of greater importance than achieving a specific goal.

From the continual perspective (Rubin et al., 1983; Rubin & Smith, 2018; Weisberg et al., 2016), all three construction play conditions in our study featured characteristics of play. More specifically, the play involved elements of choice, as children in the Free play condition could build whatever they liked. In the guided play groups, children were free to choose the blocks they played with and could build other constructions with the provided blocks. In all three conditions, children participated voluntarily and could quit playing whenever they wanted or continue playing after a break. Therefore, the play can be considered child-directed. However, we are unable to make any claims concerning the children’s intrinsic motivation during the play, as this variable was not assessed. Consistent with Zosh et al. (2018), the guided play conditions were not free of goals established by an adult, as the children were encouraged to concentrate on the constructions’ stabilities and the reasons for their stability. However, the children were not made directly aware of the learning goal. By contrast, in the Free play condition, play was free of goals set by an adult because the children could build whatever they liked without the experimenter providing suggestions. Nevertheless, in all three conditions, play was initiated by an experimenter through providing materials.

(2) In the Verbal group, the experimenter also provided verbal scaffolds. If parents or teachers play with children freely, they often talk about the play contents and offer explanations on different topics (Franse et al., 2020). Those explanations often are not purposefully implemented as verbal scaffolds, but can nonetheless be understood as such (Vygotsky, 1978). Thus, following Rubin and Smith (2018), implementing verbal scaffolds into guided play might not narrow the playfulness. Therefore, the study design may be understood as based on Zosh et al.’s (2018) suggestions regarding the implementation of guided play, with different types of play meeting all (Free play group) or some characteristics of play (guided play groups). In conclusion, the play implemented in our study can be classified as play under the continual perspective (Rubin et al., 1983; Rubin & Smith, 2018; Zosh et al., 2018).

Research question 2. We examined the relation of mass knowledge with motivational and competence beliefs. We found no indication of reciprocal effects for any of the play groups.

Difficulties in measuring self-concept in early childhood have been addressed by, e.g., Marsh et al. (2002), who developed a test instrument assessing young children’s self-concept in different domains. They found that asking about different domains with specific questions is crucial to obtain a valid measurement of self-concept in early childhood. According to the reciprocal effects model (e.g., Marsh & Craven, 2006), achievement and academic self-concept in a domain are interrelated. Previous studies have found evidence for this relation even for 7-year-old children (Guay et al., 2003). Furthermore, Marsh et al. (2002) found that even 4-year-old children can differentiate between multiple domains of self-concept, therefore measuring domain-specific self-concept is possible. Accordingly, Samarapungavan et al. (2011) assessed kindergarten children’s biology self-concept. The self-concept measure used in their study, which we used as well, showed a good fit with the data, as well as in a study by Oppermann et al. (2017). In our study, the Cronbach's α showed a good fit for all three measurement points and thus provides further evidence of the quality of the measure. Given the aforementioned studies, it can be considered unlikely that unawareness of different self-concept domains or the quality of the measure in general explain the missing correlation between mass knowledge and science self-concept in our study. However, a possible reason for the missing link might be that we assessed science self-concept by asking children how skilled they are at learning about building with blocks and how much they enjoy it. We did not focus on mass knowledge. The children might have focused on different aspects of building with blocks such as building a zoo or a garage. Therefore, our measure might have been too general (Marsh & Martin, 2011). Future research could address this issue by assessing children's science self-concept in terms of mass knowledge more specifically.

Another possible reason we found no support for the reciprocal effects model might be kindergarten children's unrealistically positive academic self-concept (Harter, 2015; Helmke, 1999). Evidence for the model was found for primary school children aged 7 years and older (Guay et al., 2003). However, Arens et al. (2016) failed to provide evidence for reciprocal effects with kindergarten children as well. Helmke (1999) found that children's academic self-concept tends to become more realistic during the primary school years, starting at approximately age 7. A possible reason for this decline in academic self-concept might be the increasing feedback by teachers and peers (Harter, 2015). Concerning block play in the kindergarten years, children are likely mostly left to build freely, and presumably, kindergarten teachers mostly give feedback on the appearance of children's buildings such as "what a beautiful tower" (e.g., Arens et al., 2016). Although feedback may also concern block buildings' stabilities, children might have little opportunity to discuss the underlying reasons for that stability. Therefore, kindergarten children's science self-concept might have little to do with their actual abilities and mass knowledge (Trawick-Smith, 2012). Our results concerning the coupling of mass knowledge and science self-concept support Helmke (1999). The correlation between children's mass knowledge and their science self-concept increased in all three groups. This denotes that children's science self-concept becomes more realistic if they engage in playful activities. Furthermore, this result provides an indication that the development of the reciprocal relation between mass knowledge and science self-concept starts in the kindergarten years.

Concluding, our study confirms the difficulty of obtaining a definitive result concerning the reciprocal effects model for this age group. To our knowledge, we have been the first to study reciprocal effects for kindergarten children in a science domain, in a context that children are familiar with from their everyday lives. Future studies on the relation between kindergarten children’s science knowledge and their science self-concept might produce different results.

4.1 Limitations

Regarding the implementation, play time was relatively brief. The children only played with the blocks for approximately one hour. More interventions over a longer period might enhance children’s mass knowledge further and stabilize their science self-concept to a greater extent. However, it was decided that an hour was sufficient as a first step to achieve ecological validity in the context of German kindergarten practice (Arens et al., 2016). Moreover, mass knowledge and the stability concept in general are rather small topics that can be acquired in a relatively short amount of time. We compared two guided play groups with verbal and material scaffolds and with material scaffolds and a free play group. From our study design, we may only conclude that guided play supports children’s mass knowledge and science self-concept more than free play. To investigate the effects of verbal scaffolds compared to material scaffolds, two other groups could be implemented in future studies. In a verbal scaffolds only group, the children might receive the verbal scaffolds and the same materials as the free play group, i.e., unstructured building blocks. Moreover, verbal scaffolds could be implemented as direct instructions, with the children receiving explanations without play (Fisher et al., 2013). Since some researchers argue that direct instruction is an effective way to teach young children about science (e.g., Dunbar & Klahr, 2012), the effect of direct instruction might be compared with the effects of free and guided play on mass knowledge and science self-concept. Additionally, the study design could be extended by including a baseline group that receives no treatment. This would allow us to study whether free play contributes to the development of mass knowledge and science self-concept compared to receiving no intervention.

The children in our guided play groups were not explicitly made aware of the goal of investigating stabilities. However, they were encouraged to engage in the investigation of stabilities implicitly through material and verbal scaffolds. Thus, a free play group receiving a prompt about the investigation of stabilities could be implemented to exclude possible effects of knowing the goal.

Children’s behavior during play was only partly assessed via manipulation check videos, but not analyzed because some parents or children denied permission to videotape. However, there might have been differences in children’s interaction with the building blocks. Some children might have interacted more actively and more frequently with the blocks, whereas others might have spent more time watching others build. Furthermore, children’s time spent playing alone or with another child and their manipulation of and their conversations about the blocks might be crucial to changes in mass knowledge and science self-concept (Harter, 2015).

Moreover, a multidimensional use of the videos might allow for controlling the scaffolds’ adaptivity (van de Pol et al., 2010). In our study, we used a limited set of verbal scaffolding techniques that were implemented in the Verbal group. Controlling for the adaptivity of the scaffolds could offer insights into individual learning processes, which could help explain learning differences in the Verbal group. Thus, children’s and the experimenter’s behavior during play should be investigated in detail in a future study.

Concerning the measurement of mass knowledge, some limitations can be identified as well. First, we investigated 5- and 6-year-old children at three points over the course of 12 weeks. Tracing developmental trajectories in mass knowledge and science self-concept over a longer period could be valuable to answer questions about possible changes in mass knowledge and science self-concept and their possible interplay. Furthermore, children’s time spent playing with blocks in their everyday lives was not assessed but could affect children’s mass knowledge acquisition (Jirout & Newcombe, 2015).

Science self-concept was assessed with reference to building blocks, but not in direct relation to mass knowledge. Since, contrary to our expectations, children’s mass knowledge was not related to their science self-concept in our study, a more specific measure of science self-concept relating to children’s mass knowledge might have produced different results.

Nevertheless, our study highlights possibilities for supporting children’s knowledge acquisition and science self-concept through incorporating guided play into kindergarten science education. Considering the findings, guided play with or without verbal scaffolds may be an effective way to support children’s knowledge and science self-concept during a brief intervention.

5 References

Andersen, L., & Ward, T. J. (2014). Expectancy-value models for the STEM persistence plans of ninth-grade, high-ability students: A comparison between black, hispanic, and white students. *Science Education*, *98*(2), 216–242. https://doi.org/10.1002/sce.21092

Arens, A. K., Marsh, H. W., Craven, R. G., Yeung, A. S., Randhawa, E., & Hasselhorn, M. (2016). Math self-concept in preschool children: Structure, achievement relations, and generalizability across gender. *Early Childhood Research Quarterly*, *36*, 391–403. https://doi.org/10.1016/j.ecresq.2015.12.024

Ball, C., Huang, K.-T., Cotten, S. R., & Rikard, R. V. (2017). Pressurizing the STEM pipeline: An expectancy-value theory analysis of youths' STEM attitudes. *Journal of Science Education and Technology*, *26*(4), 372–382. https://doi.org/10.1007/s10956-017-9685-1

Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, *48*(4), 243–270. https://doi.org/10.1080/00461520.2013.838920

Bonawitz, E. B., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Borriello, G. A., & Liben, L. S. (2018). Encouraging maternal guidance of preschoolers' spatial thinking during block play. *Child Development*, *89*(4), 1209–1222. https://doi.org/10.1111/cdev.12779

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*(5), 485–499. https://doi.org/10.1002/tea.20131

Copple, C., & Bredenkamp, J. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*. (3rd ed.). National Association for the Education of Young Children.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology*, *83*(1), 17–27. https://doi.org/10.1037/0022-0663.83.1.17

DeLoache, J. S. (2014). Early development of the understanding of symbolic artifacts. In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 312–336). Wiley-Blackwell.

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, *78*(2), 430–447. https://doi.org/10.1111/j.1467-8624.2007.01007.x.

Dunbar, K. N., & Klahr, D. (2012). Scientific thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning*. Oxford handbooks online.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist*, *44*(2), 78–89. https://doi.org/10.1080/00461520902832368

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block talk: Spatial language during block play. *Mind, Brain, and Education*, *5*(3), 143–151. https://doi.org/10.1111/j.1751-228X.2011.01122.x

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development*, *84*(6), 1872–1878. https://doi.org/10.1111/cdev.12091

Franse, R. K., van Schijndel, T. J. P., & Raijmakers, M. E. J. (2020). Parental pre-knowledge enhances guidance during inquiry-based family learning in a museum context: An individual differences perspective. *Frontiers in Psychology*, *11*, 1047. https://doi.org/10.3389/fpsyg.2020.01047

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, *37*, 161–168. https://doi.org/10.1016/j.lindif.2015.01.008

Guthrie, J. T., Wigfield, A., & Perencevich, K. C. (2004). Scaffolding for motivation and engagement in reading. In J. T. Guthrie, A. Wigfield, & K. C. Perencevich (Eds.), *Motivating reading comprehension: Concept-oriented reading instruction* (pp. 55–86). L. Erlbaum Associates.

Harter, S. (2015). *The construction of the self: Developmental and sociocultural foundations* (Second edition). Guilford Press.

Helmke, A. (1999). From optimism to realism? Development of children’s academic self-concept from kindergarten to grade 6. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 198–221). University Press.

Hsin, C.-T., & Wu, H.-K. (2011). Using scaffolding strategies to promote young children's scientific understandings of floating and sinking. J*ournal of Science Education and Technology*, *20*(5), 656–666. https://doi.org/10.1007/s10956-011-9310-7

Jansen, M. M. (2017). Academic self-concept in the sciences: Domain-specific differentiation, gender differences, and dimensional comparison effects. In F. Guay, H. W. Marsh, D. M. McInerney, & R. G. Craven (Eds.), *International advances in self research: Vol. 6. Self. Driving positive psychology and well-being* (pp. 71–112). Information Age Publishing.

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. Sample. *Psychological Science*, *26*(3), 302–310. https://doi.org/10.1177/0956797614563338

Kaplan, A., Katz, I., & Flum, H. (2012). Motivation theory in educational practice: Knowledge claims, challenges, and future directions. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors*. (1st ed., pp. 165–194). American Psychological Association.

Kaplan, A., & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, *19*(2), 141–184. https://doi.org/10.1007/s10648-006-9012-5

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*(6045), 971–975. https://doi.org/10.1126/science.1204528

Krist, H. (2010). Development of intuitions about support beyond infancy. *Developmental Psychology*, *46*(1), 266–278. https://doi.org/10.1037/a0018040

Krist, H., Horz, H., & Schönfeld, T. (2005). Children's block balancing revisited: No evidence for representational redescription. *Swiss Journal of Psychology*, *64*(3), 183–193. https://doi.org/10.1024/1421-0185.64.3.183

Lauermann, F., Tsai, Y.-M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy-value theory of achievement-related behaviors. *Developmental Psychology*, *53*(8), 1540–1559. https://doi.org/10.1037/dev0000367

Leuchter, M., & Naber, B. (2019). Studying children's knowledge base of one-sided levers as force amplifiers. *Journal of Research in Science Teaching*, *56*(1), 91–112. https://doi.org/10.1002/tea.21470

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, *80*(3), 366–380. https://doi.org/10.1037/0022-0663.80.3.366

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Ellis, L. A., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology*, *38*(3), 376–393. https://doi.org/10.1037/0012-1649.38.3.376

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., & Richards, G. (1988). The Outward Bound Bridging Course for low achieving high-school males: Effect on academic achievement and multidimensional self-concepts. *Australian Journal of Psychology*, *40*, 281–298.

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Vol. 1. Theories, constructs, and critical issues* (pp. 427–458). American Psychological Association. https://doi.org/10.1037/13273-015

Martin, N. D., Dornfeld Tissenbaum, C., Gnesdilow, D., & Puntambekar, S. (2019). Fading distributed scaffolds: The importance of complementarity between teacher and material scaffolds. *Instructional Science*, *47*(1), 69–98. https://doi.org/10.1007/s11251-018-9474-0

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education*, *21*(4), 467–487. https://doi.org/10.1007/BF03173515

OECD. (2014). *Measuring innovation in education: A new perspective, educational research and innovation. Educational Research & Innovation*. OECD Publishing. https://doi.org/10.1787/20769679

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, *41*(3), 181–206. https://doi.org/10.1207/s15326985ep4103_4

Oppermann, E., Brunner, M., Eccles, J. S., & Anders, Y. (2017). Uncovering young children's motivational beliefs about learning science. *Journal of Research in Science Teaching*, *24*(2), 195. https://doi.org/10.1002/tea.21424

Pellegrini, A. D. (2013). Play. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol 2: Self and other* (pp. 276–299). Oxford Univ. Press.

Pine, K. J., & Messer, D. (2003). The development of representations as children learn about balancing. *British Journal of Developmental Psychology*, *21*, 285–301.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Plöger, I. (2020). *Statik als Lerngegenstand im Übergang von Kindergarten zur Grundschule [Statics as a learning opportunity at the transition from kindergarten to primary school]*. University of Koblenz-Landau, Landau.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing (Version 4.0.3) [Computer software]. Vienna, Austria. URL https://www.R-project.org/.

Richey, J. E., & Nokes-Malach, T. J. (2013). How much is too much? Learning and motivation effects of adding instructional explanations to worked examples. *Learning and Instruction*, *25*, 104–124. https://doi.org/10.1016/j.learninstruc.2012.11.006

Rubin, K. H., Fein, G. G., & Vandenberg, B. (1983). Play. In E. M. Hetherington (Ed.), *Handbook of child psychology* (pp. 693–774). Wiley.

Rubin, K. H., & Smith, K. (2018). Play in human development. In M. H. Bornstein (Ed.), *The SAGE encyclopedia of lifespan human development*. SAGE Publications, Inc.

Samarapungavan, A., Patrick, H., & Mantzicopoulos, P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cognition and Instruction*, *29*(4), 416–470. https://doi.org/10.1080/07370008.2011.608027

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modelling change and event occurence*. Oxford Univ. Press.

Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development*, *66*(1), 209–223. http://www.jstor.org/stable/1131201

Trawick-Smith, J. (2012). Teacher-child play interactions to achieve learning outcomes. In R. Pianta (Ed.), *Handbook of early childhood education* (pp. 259–277). Guilford Publications.

Trundle, K. C. (2015). The inclusion of science in early childhood classrooms. In K. C. Trundle & M. Saçkes (Eds.), *Research in early childhood science education* (pp. 1–6). Springer.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

van der Graaf, J., Segers, E., & Jong, T. de (2020). Fostering integration of informational texts and virtual labs during inquiry-based learning. *Contemporary Educational Psychology*, *62*, 101890. https://doi.org/10.1016/j.cedpsych.2020.101890

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Harvard Univ. Press.

Weber, A. M., Reuter, T., & Leuchter, M. (2020). The impact of a construction play on 5- to 6-year-old children’s reasoning about balance relationships. *Frontiers in Psychology*, *11*, 1737. https://doi.org/10.3389/fpsyg.2020.01737

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play. *Current Directions in Psychological Science*, *25*(3), 177–182. https://doi.org/10.1177/0963721416645512

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Solis, S. L., & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology*, *9*, 1124. https://doi.org/10.3389/fpsyg.2018.01124

General Discussion

Science education can already be implemented into early childhood education and start as early as kindergarten (OECD, 2014; Trundle, 2015). To implement effective science education in the early childhood years, identifying effective ways to promote kindergarten children's science knowledge and theories is fundamental. However, studies on early science learning and education are sparse. Therefore, it remains unclear how children' science learning may be supported in an effective way (Trundle, 2015).

To identify effective ways of promoting children's science knowledge, e.g., their intuitive knowledge about stabilities, and theories, e.g., their reasoning about stabilities that they can explicitly state, at least three aspects may be considered. These aspects are (1) children's cognitive development, (2) children's motivational prerequisites such as their trust in the science abilities, and (3) other interindividual prerequisites such as intelligence, and interest.

(1) Studies on children's cognitive development have provided evidence indicating that 5- to 6-year-old children are able to use deductive (e.g., Chantal & Markovits, 2017), inductive (e.g., Goddu et al., 2020), and abductive inference (e.g., Bonawitz et al., 2012). Moreover, children are able to generalise (e.g., Gelman & Markman, 1986), categorise (e.g., Gelman & Coley, 1990), form concepts (e.g., Rhodes et al., 2008), and understand causality (e.g., Sobel & Kirkham, 2006). Indeed, causality might facilitate these cognitive processes (Gopnik, 2013; Hayes & Thompson, 2007; Opfer & Bulloch, 2007; Schulz, Bonawitz, & Griffiths, 2007). The role of causality for children's reasoning and their cognitive development is highlighted in theory theory and its Bayesian application, which integrate general, e.g., statistical probabilities, and specific processes, e.g., prior knowledge or theories (Gopnik, 2013; Gopnik & Wellman, 2012). Findings on theory theory and young children's ability to coordinate theory with evidence have yielded results indicating that children can draw inferences from causal relations if they are presented in the form of perfect covariation (Bonawitz et al., 2012; Koerber et al., 2005). Therefore, providing children with causal relations, e.g., *if the mass is supported, the block remains stable*, might facilitate their learning. Moreover, Bayesian inference highlights the role of children's prior theories or knowledge, indicating that science education should refer to them (Gopnik, 2013). Thus, evidence might be most effective for children's science learning if it always contradicts children's intuitive theories, e.g., centre theory, and always supports the more scientifically correct theory, e.g., mass theory.

(2) Studies on children's trust in their own abilities have yielded results indicating that a positive academic self-concept interacts with corresponding achievement (Guo et al., 2015). Moreover, Marsh et al. (2012) suggested that educational interventions aiming at enhancing

knowledge, but undermining children's corresponding self-concept might negatively affect long-term knowledge acquisition. Therefore, the consideration of children's trust in their own abilities in education is demanded (OECD, 2014; Pintrich, 2003).

(3) Other interindividual prerequisites such as fluid and crystallised intelligence, mental rotation ability, and interest in a science content have been suggested as possibly relating to science learning. For example, Flynn and Blair (2013) highlight the role of intelligence in learning. Newcombe et al. (2013) suggest that mental rotation indicates children's spatial abilities, which might relate to their spatial learning, e.g., learning about stabilities. Hidi et al. (2017) underline the role of interest in a specific content for learning about the content.

All of these demands may be met by playful learning (Hassinger-Das et al., 2017; Weisberg et al., 2013; Zosh et al., 2018), which is considered as being developmentally appropriate (Copple & Bredenkamp, 2009). Children may be confronted with perfect covariation relating to their intuitive theories and may have positive experiences in a domain they are interested in. An adult might guide children's learning and refer to their competence and success (Belland et al., 2013; van de Pol et al., 2010). These demands were met in the studies that were conducted for the present dissertation. The next chapter summarises the results of these studies and afterwards, they are discussed in the light of these demands.

1 Summary

The findings of the three articles are presented in Table 1. The first article *Measuring preschool children's knowledge of the principle of static equilibrium in the context of building blocks: Validation of a test instrument* was concerned with the validation of the COM Test for measuring children's stability knowledge. The test validity was examined in a sample of 5- to 6-year-old German preschool children. In Study 1, the construct structure of stability knowledge was tested and results showed that it had two subfacets, knowledge about stable and knowledge about unstable constructions. Moreover, a 1-PL-testlet model with the two mentioned subfacets showed a good fit to data and the COM Test covered the average and high stability knowledge spectrum. Study 2 was concerned with the COM Test's construct validity. Results implied that children's stability knowledge was related to their geometrical-centre knowledge as measured by Krist (2010), their fluid intelligence indicated by figural perception and figural reasoning, and their crystallised intelligence indicated by their language capacity (Flynn & Blair, 2013). However, there was no relation between stability knowledge and mental rotation ability, interest, motivational or competence beliefs.

The second article *The impact of a construction play on 5- to 6-year-old children's reasoning about stability* focused on children's intuitive theories about stabilities and the effects of guided and free play, children's prior theories as well as their intelligence on these intuitive theories. These questions were investigated with a sample of 5- to 6-year-old preschool children with a pre-post-follow-up intervention design. At pretest, few children applied a mass theory when explaining the building blocks' stabilities, most children applied a centre or another theory (Bonawitz et al., 2012). Consistent application of mass theory increased in all three intervention groups, i.e., Verbal, Material, and Free play. Children in the Verbal group, which had received a guided play with material and verbal scaffolds, were most likely to apply mass theory consistently after the intervention. In line with theory theory and its Bayesian extension (Gopnik & Wellman, 2012), consistency of children's prior theories affected their acquisition of mass theory. Children with inconsistent prior theories were more likely to acquire a mass theory than children with consistent prior theories. Specifically, children with inconsistent prior theories in the Verbal group were most likely to apply mass theory consistently at posttest and follow-up. Additionally, both fluid and crystallised intelligence positively affected mass theory acquisition. More precisely, crystallised intelligence interacted with play group on children's consistent application of mass theory. Children with a high crystallised intelligence were most likely to acquire mass theory if they received a high amount of scaffolding, i.e., in the Verbal group. Whereas children with a low crystallised intelligence were most likely to acquire a mass theory if they received a lower amount of scaffolding, i.e., in the Material group.

The third article *Construction play promotes change in 5- to 6-year old children's science knowledge about stabilities and science self-concept* examined the effect of guided and free play on 5- to 6-year-old children's stability knowledge and their corresponding science self-concept, i.e., their motivational and competence beliefs. Furthermore, potential reciprocal effects between stability knowledge and motivational and competence beliefs were investigated in a pre-post-follow-up intervention design. Children's stability knowledge increased in both guided play groups after the playful intervention, but not in the Free play group. Additionally, children's motivational and competence beliefs remained stable in both guided play groups, but decreased in the Free play group. Results showed limited evidence for reciprocal effects between stability knowledge acquisition and motivational or competence beliefs, which is in line with results by Arens et al. (2016).

The results of the three articles are discussed in light of the research questions that guided the present dissertation.

Table 1

Short summary of the key findings of articles 1, 2, and 3

	Article 1 **Measuring preschoolers' knowledge of static equilibrium**	**Article 2** **Play's impact on children's reasoning**	**Article 3** **Play fosters children's knowledge and self-concept**
Aims	(1) Validating a test instrument to measure children's stability knowledge using item response modelling (2) Investigating construct validity	(1) Investigating children's intuitive theories about stability (2) Examining if guided play with material (+ verbal) scaffolds and free play affect the consistent application of theories (3) Investigating the effect of children's prior theories on their consistent use (4) Investigating the relation of theory development and intelligence	(1) Examining if guided play with material (+ verbal) scaffolds and free play affect knowledge about stability and science self-concept (2) Investigating the reciprocal effects between knowledge acquisition and science self-concept
Method	Study 1: Validation study; *N* = 217 5- and 6-year-old children from Germany Study 2: Validation study; *N* = 166 5- and 6-year-old children from Germany	Pre-Post-Follow-Up design with three play groups, guided play with material and verbal scaffolds, guided play with material scaffolds, free play; *N* = 183 5- and 6-year-old children from Germany	Pre-Post-Follow-Up design with three play groups, guided play with material and verbal scaffolds, guided play with material scaffolds, free play; *N* = 183 5- and 6-year-old children from Germany
Analyses	Study 1: Item response modelling Study 2: Moderated regression, correlations, Kruskal-Wallis Tests	*z*-tests of proportions, Kaplan-Meier analyses, Cox-regressions, multiple and moderated regressions, ANOVA	ANOVA, CFA, mixed-effects models, cross-lagged panel models
Findings	Evidence for validity of the test instrument to measure children's stability knowledge Evidence for the latent structure of stability knowledge Correlations of stability knowledge with geometrical-centre knowledge (Krist's Test) and intelligence	Few children applied a mass theory at pretest Children in the guided play group with material and verbal scaffolds were most likely to apply mass theory consistently after the play Consistency of prior theories and intelligence affected acquisition of mass theory	Guided play with material and verbal scaffolds and guided play with material scaffolds enhanced children's stability knowledge and stabilised children's science self-concept No evidence for reciprocal effects between stability knowledge acquisition and science self-concept

2 Question 1: Is knowledge about stability a unidimensional construct?

In order to investigate children's stability knowledge acquisition, the construct structure according to a test assessing stability knowledge, the COM Test, was examined in the first article. The test consisted of photographs of block constructions, which children are familiar with through their own block play (Fisher et al., 2011; Rubin et al., 1978). Results showed that stability knowledge can be construed as a unidimensional construct with the two subfacets knowledge of stable and knowledge of unstable constructions. Moreover, the items showing unstable constructions were easier for the children to solve. This indicates that children's stability concept might include knowledge about a weight destabilising an object earlier than the knowledge that a counterweight can also stabilise an object. From a developmental perspective, tumbling objects might be more memorable because of the noise they make and the emotional response a child might have, e.g., distress (Jiang et al., 2017; Johnson, 2013). The results also implied that the items were suitable to measure children's stability knowledge in the medium to high knowledge range, and are another indication for the COM Test's validity.

It was decided to use photographs in a paper-pencil test to assess children's stability knowledge, as studies by Krist and his colleagues (Krist et al., 2005; Krist, 2010; Krist et al., 2018) showed that different assessment methods (behavioural observation, paper-pencil tests, and eye-tracking) yield the same results concerning children's stability knowledge. The COM Test as a paper-pencil test has three advantages. (1) It only takes a few minutes for children to fill out, and can be conducted in a group of children. Thus, it is much more time-efficient than behavioural observations in a one-on-one setting. (2) Moreover, it is non-verbal, as children do not need to verbalize their knowledge. This is especially important for children, who have a different native language than the language mostly spoken in their country of residence, as they might not speak their second language as well as their native language. By using a non-verbal instrument, these children might find it easier to show their stability knowledge. (3) Additionally, the COM Test consists only of asymmetrical constructions that can only be rated correctly if a child has mass knowledge. Children with centre knowledge will rate the constructions incorrectly, because the geometrical centre is never supported if the block remains stable, and always supported if the construction tumbles. Children with neither mass or centre knowledge can only guess and will rate approximately 50% of the items correctly. In Krist's Test (Krist, 2010) symmetrical as well as asymmetrical constructions were used. The asymmetrical constructions could be rated correctly with mass knowledge, however both children with centre and other knowledge could only guess and would rate 50% of the items correctly, as the blocks were balanced directly at their geometrical centre. Therefore, the COM Test can differentiate between children with mass, centre, and other knowledge, while Krist's Test can simply differentiate between mass knowledge and not mass knowledge.

However, the COM Test only measures children's knowledge, not their intuitive theories about stability. For this purpose, children's balancing behaviour (Bonawitz et al., 2012) or their reasoning about stabilities have to be assessed. Therefore, the second article focused on children's reasoning about stabilities, thus, measuring their intuitive theories. For this, children were first asked to rate a photograph of a construction supported by a yellow and a black block as stable or unstable if the black block was removed. Then they rebuilt the structure and were asked why they thought that the construction would remain stable or tumble. After they had explained their reasoning, children removed the black block and ascertained if they had answered correctly. Children's answers could not be summarised, as they were categorical and children had received feedback between items. Thus, children's answers depended on each other.

Both the first and the second article examined whether children were more likely to consider the mass, the centre, or something else, when rating and explaining stabilities. In the first article, most children applied centre or other knowledge, and very few children applied mass knowledge. When asked about their reasoning, children were still more likely to reason with a centre or another theory, but more children answered with a mass theory for at least one of the items. However, only 7 children, i.e., 4%, applied a mass theory consistently for all three items of the pretest in the second article. In the first article, results showed that 12 children, i.e., 6%, in Study 1 and 6 children, i.e., 4%, in Study 2 were categorised as having mass knowledge. Therefore, the percentage of children considering the mass consistently was approximately the same in both articles, regardless of assessment method. This result further supports the validity of the COM Test. Moreover, results of both articles were in line with Bonawitz et al. (2012), who also found that children apply mass, centre, or another theory. Furthermore, the results indicated that children's stability knowledge is a unidimensional construct. In order to investigate whether children's knowledge and theories can be supported, so that they learn about the importance of the mass for stability, a playful intervention was implemented.

3 Question 2: How do different forms of play affect children's knowledge and theories about stabilities?

To examine whether different forms of play affected children's knowledge acquisition and intuitive theories about stability, three forms of play, guided play with material and verbal scaffolds, guided play with material scaffolds, and free play, were implemented in the second and third articles. The second article was concerned with children's intuitive theories and results implied that a combination of material and verbal scaffolds were suited best to foster young children's mass theory. Children, who had received the material scaffolds only, were less likely to acquire a mass theory than children, who had received a combination of material and verbal scaffolds. However, they were more likely to acquire a mass theory than children, who played with blocks freely. The third article was concerned with children's stability knowledge acquisition and results showed that children in both guided play groups acquired mass knowledge, but the free play group did not.

Children in the free play group might have observed evidence for mass theory, while they were playing with blocks. However, they were not actively confronted with evidence and thus their observations could have only happened randomly. Maybe children ignored or explained away the evidence (Kuhn, 2014). In line with findings on theory theory, children might have to observe evidence contradicting their prior theory—mostly centre or other theory—or their knowledge repeatedly in order to adapt their prior theory or to acquire new knowledge (for an overview see, Gopnik & Wellman, 2012; Koslowski & Masnick, 2014).

Therefore, the results of both articles imply that guided play supports children's learning, which is in line with results from previous studies (Borriello & Liben, 2018; Ferrara et al., 2011; Leuchter & Naber, 2019; Reuter & Leuchter, 2020; Verdine et al., 2019). Moreover, the results also show that children need support in order to learn about science. Krist (2010) has demonstrated that children learn about the importance of the mass for an object's stability during the course of their development. However, in order to promote this development scaffolds are necessary, while only free play might not support children's learning (Weisberg et al., 2016; Zosh et al., 2018).

Nonetheless, the two articles differ in their evidence on the importance of verbal scaffolds for children's learning. While verbal scaffolds were important for children's consistent application of mass theory, they did not contribute to their mass knowledge acquisition beyond the effect of material scaffolds. There are at least three possible explanations for the different results concerning (1) the statistical analyses, (2) the possibility that children acquired new vocabulary, and (3) the possibility that knowledge and reasoning should be supported differently.

(1) Two different statistical analyses were conducted in the second and the third article to investigate change in knowledge and theory, respectively. In the second article, methods of survival analyses were applied in order to investigate children's theories as categorical variables (Singer & Willett, 2003). In the third article, mixed-effects growth models were used to investigate change in mean levels (Singer & Willett, 2003). Survival analyses need a criterium, i.e.,

consistent application of mass theory, that can either be attained or not attained and at different points in time. Reaching the criterium at an earlier time is deemed to imply faster learning. The mixed-effects growth models did not need such a criterium, but relied on linear change. Therefore, the statistical analyses might be partly responsible for the different results.

(2) Another possible reason for the different results concerns the possibility that children in the group with material and verbal scaffolds might have acquired the vocabulary to explain their mass theory. The verbal scaffolds included providing explanations, e.g., *If the heavy side of a block hovers in mid-air, the block will tumble*. Children in the group with only material scaffolds might have not acquired this vocabulary and therefore could not explain their reasoning as well as the guided play group with material and verbal scaffolds. However, particular care was taken to prevent a lack in vocabulary from influencing the rating of children applying mass, centre, or other theory. Consequently, children did not necessarily have to use the words mass or weight in order to be categorised as applying mass theory. Instead, it was deemed sufficient if they gestured to the counterweight stabilising or destabilising the building (Pine et al., 2004), or referred to the counterweight verbally, e.g., *It is stable because of this block*. Therefore, it seems unlikely that a more sophisticated vocabulary was the reason for the different results concerning knowledge and theory.

(3) The results may also be interpreted as evidence that knowledge acquisition and change in theory should be supported differently. For knowledge gain, material scaffolds might be sufficient (Martin et al., 2019), as acquiring specific knowledge might be easier than changing an intuitive theory (Karmiloff-Smith & Inhelder, 1974; Pine et al., 2004). The material scaffolds used in both studies might have been so self-explanatory and low-threshold that simply engaging with the materials fostered stability knowledge. However, the verbal scaffolds might have directly stimulated children’s reasoning according to Kuhn (2013) by asking for children’s justifications, e.g., *How do you know that the blocks will remain stable?*. Therefore, children had to actively think about causes, which are essential for their intuitive theories (Gerstenberg & Tenenbaum, 2017; Gopnik, 2013; Gopnik & Wellman, 2012) and which have been shown to stimulate cognitive processes and theory adaptation (Bonawitz et al., 2012; Hayes & Thompson, 2007; Opfer & Bulloch, 2007; Schulz, Bonawitz, & Griffiths, 2007). Children might have used deduction (*If the mass is supported, the block remains stable.*), induction (*All of these blocks remained stable if their mass was supported. Maybe the mass is important.*), and abduction in combination with their Bayesian priors and their ability to coordinate perfect covariation with their theory (*I thought the centre was important, but I saw that the centre was unsupported and still the blocks remained stable. Something else must be important. Maybe it’s*

the mass.). Moreover, the search for causality might have facilitated the generalisation of the importance of the mass (Hayes & Thompson, 2007). Children may have started to view support of the mass as essential for stability (Gopnik & Sobel, 2000; Sobel & Kirkham, 2006), thus influencing their categorisation of block constructions as stable or unstable. Therefore, children in the group with material and verbal scaffolds might have actively searched for causes and adapted their theories. Thus, they did not only acquire stability knowledge, but an explicit theory that they can explain, because they were encouraged to reason about underlying causal relations and verbalize them. Children in the group with material scaffolds might have acquired knowledge, but not an explicit theory that they were able to explain (Karmiloff-Smith & Inhelder, 1974).

Concluding, in order to foster children's knowledge, material scaffolds may be sufficient. However, to support children's theory adaptation, their reasoning should be encouraged with verbal scaffolds. Moreover, according to theory theory, children's prior knowledge or theories might affect their knowledge acquisition and theory adaptation as well.

4 Question 3: What is the effect of prior knowledge or theories on knowledge acquisition or theory adaptation in the science domain of stability?

Both the second and third article investigated whether prior knowledge or theories affected children's knowledge acquisition or theory adaptation. Results implied that prior knowledge or theories influenced children's mass knowledge and mass theory acquisition.

Results of the second article produced evidence for Bayesian inference, because consistency of prior theory affected children's adaptation to mass theory (Bonawitz et al., 2012; Gopnik, 2013; Gopnik & Wellman, 2012; Griffiths et al., 2011). Bayesian inference suggests that the consistency of children's prior theories indicates how much they believe their intuitive theory to be true (Gopnik, 2013; Schulz, Gopnik, & Glymour, 2007). If a child applies centre theory consistently, they are assumed to view the importance of the centre as extremely likely and to dismiss evidence to the contrary. Children with inconsistent prior theories are considered to perceive their intuitive theory as less likely. Therefore, children who apply different theories, might be unsure of their reasoning about stabilities and might be more likely to accept counterevidence for centre theory. The results of the second article provided evidence for these assumptions. Children with inconsistent prior theories might have been more likely to accept evidence for other theories, since they perceived their prior theories as unlikely, whereas children with consistent prior theories were probably quite convinced of their intuitive theories. Presumably, the theories have been valuable in explaining stability in the past and were thus perceived as likely being true. Therefore, children might have needed to observe even more evidence contradicting their prior theory and confirming mass theory. The children might have ignored or dismissed the evidence they observed (Griffiths et al., 2011; Kuhn, 2013).

Moreover, consistency of prior theory interacted with play group. Children with inconsistent prior theories were most likely to acquire a mass theory if they received material and verbal support, while children with consistent prior theories were less likely to acquire a mass theory. All children were encouraged to reason about stabilities in the guided play group with material and verbal scaffolds. This seemingly promoted only children with inconsistent prior theories, as they thought about possible causes and concluded that the mass must be a deciding factor (Bonawitz et al., 2012). Children with consistent theories might have dismissed the counterevidence completely, or reasoned that their intuitive theory, i.e., centre or other theory, explained the evidence sufficiently and ignored evidence to the contrary (Bonawitz et al., 2012).

In the guided play group with material scaffolds or in the free play group, consistency of prior theory did not affect children's acquisition of a mass theory. Children in these groups were not encouraged to think about underlying causes for stability. Therefore, children might not have reasoned about underlying causes, and dismissed counterevidence (Griffiths et al., 2011; Kuhn, 2013). Concluding, the second article yielded evidence for theory theory and its Bayesian application. The results indicate that children apply both general processes, as they presumably interpreted perfect covariance, as well as specific processes, as they had formed theories about stability and interpreted evidence according to their prior theory.

While not the primary focus of the third article, results of the cross-lagged panel model indicated that prior mass knowledge affected mass knowledge at posttest in all three groups. This effect was equally large in all three play groups. Again, the two articles came to different results concerning group differences. The reason for the different results may be that the consistency of children's prior knowledge was not analysed in the third article, while it was the consistency of prior theories that affected children's mass theory in the second article.

The studies have provided evidence that children's science learning can be supported through guided play. Pintrich (2003) and Marsh et al. (2012) postulate that children's trust in their own abilities, i.e., their science self-concept, should be considered as well.

5 Question 4: How do different forms of play affect children's science self-concept?

The third article focused on the effects of different forms of play on young children's science self-concept. Jansen (2017) and Mantzicopoulos et al. (2013) report that young children and adolescents perceive science as hard and often have a lower science self-concept compared to their language self-concept. Patrick and Mantzicopoulos (2015) suggest that this low science self-concept might stem from the inadequate science education in early education. Since Marsh, Lüdtke, et al. (2015) and Murayama et al. (2013) found that motivational and competence self-concept are fundamental predictors of educational outcomes, fostering children's science self-concept at an early age is crucial for children's science learning. Moreover, Marsh et al. (2012) claim that interventions trying to enhance achievement, without regarding positive self-beliefs are likely to undermine achievement as well. Therefore, potential effects of interventions targeting specific knowledge on children's corresponding self-concept should be investigated. Thus, the different play forms were examined concerning their effects on science self-concept.

To investigate this question, the structure of science self-concept was examined first, and science self-concept was divided into a motivational and a competence subfacet in line with Arens et al. (2016). The motivational subfacet indicated children's joy in learning about science, and the competence subfacet indicated children's perceived competence in science (Trautwein et al., 2013; Wigfield & Eccles, 2000). Therefore, science self-concept was found to be hierarchically organised (Marsh & Shavelson, 1985; Shavelson et al., 1976), since motivational and competence beliefs each constituted one facet of children's science self-concept. Moreover, the correlation of the two facets indicates that they share a common higher order factor, i.e., science self-concept. Furthermore, the results of the third article imply that science self-concept can be measured reliably in early childhood (Harter, 2015; Jansen, 2017; Marsh et al., 2002; Oppermann et al., 2017; Samarapungavan et al., 2011).

Moreover, the results showed that children had both high motivational and competence beliefs at pretest. This is in line with previous studies on kindergarten children's academic self-concept (Helmke, 1999). Young children at around 5 to 6 years of age often engage in all-or-none thinking and thus might have considered themselves as *all good* and *having much fun*.

Furthermore, results implied that both children's motivational and their competence beliefs remained stable in the guided play groups, but decreased in the free play group. For the guided play groups, a further increase was unlikely, because the motivational and competence beliefs at pretest were already quite high. However, the guided play might have forestalled the developmentally determined decrease (Harter, 2015; Helmke, 1999; Mantzicopoulos et al., 2013). In the guided play groups the children probably had challenging experiences that were still suitable for young children and might have allowed them to gain new positive experiences and successes with block play and stability (Shavelson et al., 1976). In the free play group, children played with blocks freely and might have experienced more failures such as tumbling block towers, and they might not have tried to rebuild their previously collapsed building, but moved on to building something else. Therefore, the sense of achievement might not have formed. Children in all groups were encouraged to try again and praised for their effort. However, seemingly praise and encouragement alone were insufficient to foster their science self-concept. Accordingly, the developmentally determined decrease in self-concept might not have been prevented in the free play group (Harter, 2015).

Even though specific verbal scaffolds were implemented to support children's science self-concept (Belland et al., 2013), results showed no difference between the guided play group with material and verbal scaffolds and the guided play group with material scaffolds. Therefore, verbal scaffolds seemingly did not strengthen children's science self-concept, but material scaffolds during play were sufficient. There are at least two possible reasons that the verbal scaffolds did not affect science self-concept, (1) the importance of positive experiences compared to feedback, and (2) the short intervention time.

(1) Positive experiences might be more important for young children than feedback by an adult, especially by an adult they hardly know (Harter, 2015). Studies found that feedback from attachment figures, such as parents, siblings, friends, or teachers, influence children's academic self-concept (Frome & Eccles, 1998; Helmke & van Aken, 1995). Feedback becomes increasingly important during the primary school years, because children are confronted with teacher and peer feedback, i.e., attachment figures, on a daily basis (Helmke, 1999). In the third study, children received scaffolds by an experimenter they had only met once or twice before. The scaffolds provided by the experimenter probably did not affect children's motivational and competence beliefs, as the children might not have perceived the experimenter as an attachment figure. The verbal scaffolds might have affected children's motivational and competence beliefs if they had been provided by their kindergarten teacher or their parents.

(2) The second possible reason goes hand in hand with the first. Starting in primary school, children are confronted with feedback daily, and studies on enhancing children's and adolescents' academic self-concept implemented interventions that spanned at least multiple weeks (Craven et al., 1991; Marsh & Richards, 1988; Patrick et al., 2009; Samarapungavan et al., 2011). Therefore, the time frame for the verbal scaffolds to take effect might have been too short as the intervention only lasted for one hour.

Nevertheless, guided play is an effective way to support and stabilise children's science self-concept. Even after a very short intervention, an effect of the play on self-concept could be detected. Moreover, it is also of interest whether there were any reciprocal effects between children's achievement and their motivational and competence beliefs, and if these effects differed between the different play groups.

6 Question 5: Are there reciprocal effects between young children’s science self-concept and their stability knowledge acquisition?

Another main focus of the third article was the relation between children’s mass knowledge and their motivational and competence beliefs. As suggested in the reciprocal effects model (Marsh & Craven, 2006), mass knowledge and science self-concept, i.e., motivational and competence beliefs, should influence each other over time. However, similar to the first article, mass knowledge and motivational or competence beliefs were not related at pretest. Moreover, there was no relation between mass knowledge and motivational or competence beliefs at posttest or follow-up. The only reciprocal effect that was uncovered was an effect of mass knowledge at posttest on children’s competence beliefs at follow-up in the free play group. No other reciprocal effects were found. Concerning the autoregressive effects, mass knowledge, motivational and competence beliefs showed a high stability over all three points of measurement. The results are in line with findings by Arens et al. (2016) who also found little support for reciprocal effects in math self-concept for kindergarten children, but found high stability estimates for math achievement and math self-concept. At least three different reasons for the missing reciprocal effects seem plausible. These reasons concern (1) the measurement method of children’s motivational and competence beliefs, (2) children’s unrealistically positive self-beliefs, and (3) the effect of kindergarten education.

(1) Three reasons might have contributed to the missing reciprocal effects concerning the measurement method. First, the science self-concept measure itself may have been responsible for the missing reciprocal relation between mass knowledge and motivational or competence beliefs. However, taking a closer look at the measure this seems unlikely. Children’s science self-concept was assessed with an adaptation of the Young Children’s Science Motivation Scale (Oppermann et al., 2017; Samarapungavan et al., 2011). Both studies as well as the results of the third article revealed that the scale was a reliable measure for children’s motivational and competence beliefs. Second, children might have been unaware of different self-concept domains. Yet, studies have found that 5- and 6-year-old children are aware of different domains and self-concept can be measured domain-specifically (Marsh et al., 2002; Samarapungavan et al., 2011). Therefore, neither the quality of the scale itself or children’s unawareness of different self-concept domains are likely to be the reasons for the missing reciprocal relation. Third, and more likely, the assessment of science self-concept related to block play instead of stability knowledge might be responsible for the missing reciprocal effects. Accordingly, Marsh et al. (2002) claim that assessing academic self-concept as accurately and domain-specific as possible is important especially for young children. The measure in the third article referred to children’s

motivational and competence beliefs in learning about building with blocks, not to mass knowledge in specific. It was decided to ask about block building, because it is a familiar activity for young children that they engage in often (Borriello & Liben, 2018). Therefore, children probably have already formed motivational and competence beliefs for this specific activity. Nevertheless, asking children about their science self-concept related to stabilities might have led to different results concerning the reciprocal effects with mass knowledge. Nonetheless, Arens et al. (2016) did not find results in support of the reciprocal effects model for kindergarten children either, even though they used a different self-concept measure. Moreover, the relation between mass knowledge and science self-concept increased over time in the third article. Both results indicate that the way science self-concept was measured, i.e., in relation to block play instead of stability knowledge, might be a contributing factor, but not the only reason for the missing reciprocal effects.

(2) From a developmental psychological perspective, young children often believe that they are either all good or all bad. Accordingly, studies have uncovered that kindergarten children have very positive and unrealistic motivational and competence beliefs (Harter, 2015; Helmke, 1999). Therefore, they might not have related their science self-concept to their actual science achievement, yet.

(3) The German kindergarten system might have affected the missing reciprocal effects between mass knowledge and motivational and competence beliefs as well. Helmke (1999) related children’s decrease in academic self-concept in primary school to the feedback provided by teachers and peers, and accordingly Guay et al. (2003) found reciprocal effects for 7-year-old primary school children. Concerning kindergarten education, Arens et al. (2016) suggest that the education in kindergarten is mainly implemented through free play, and kindergarten teachers’ feedback is mostly encouraging and concerned with surface features. Concerning block play, kindergarten teachers may refer to a specific building a child has built, e.g., *What a beautiful building you made*. While they might also talk about stabilities, children might have little opportunity to think and talk about the reasons for stability. This also suggests that children might not have related their actual abilities to their motivational and competence beliefs (Trawick-Smith, 2012).

Nevertheless, the results of the third article also suggested that the relation between mass knowledge and motivational and competence beliefs increased over time. The relation even became significant for the free play group for the effect of mass knowledge at posttest on competence beliefs at follow-up. This finding might be interpreted as an indicator that the reciprocal relation between children's knowledge in a science domain and their science self-concept starts to form by the age of 5 to 6 years, shortly before children start primary school. This relation might have been supported by the playful intervention. Through playing with building blocks children's motivational and competence beliefs might have become more realistic, as the children received feedback, while their mass knowledge increased as well. Therefore, playful learning might enhance the relationship between children's knowledge and academic self-concept in a domain.

Besides the relation of mass knowledge and motivational and competence beliefs, other interindividual prerequisites might affect children's knowledge acquisition and theory adaptation. Therefore, the relation of mass knowledge and mass theory acquisition with intelligence, mental rotation ability, and interest in block play was investigated.

7 Question 6: Is children's knowledge about stability related to interindividual prerequisites such as fluid and crystallised intelligence, mental rotation ability, and interest in block play?

The first and the second article examined whether different interindividual prerequisites were related to children's knowledge about stabilities concerning the mass and their mass theory acquisition over three points of measurement. In both articles, results showed positive relations of mass knowledge and mass theory with fluid and crystallised intelligence. The relations between mass knowledge and mental rotation ability, and interest in block play were only investigated in the first article, but results showed no evidence for relations between the constructs.

7.1 Fluid intelligence

Figural perception and figural reasoning are indicators of fluid intelligence and were assessed in both studies. They represent children's abilities to mentally represent and identify spatial regularities (Cattell, 1987; Weiß & Osterland, 2013). These abilities might have been underlying processes for correctly estimating stability, because they might be necessary to identify the position of the blocks and the amount of contact with a supporting surface. Moreover, fluid intelligence also facilitates discovering new principles, such as the importance of the mass, and applying them to solve problems (Flynn & Blair, 2013). In the first article, children's fluid intelligence was related to their mass knowledge. Therefore, children with a higher fluid intelligence might have determined the mass as the deciding factor for the block constructions' stabilities, and thus estimated the constructions' stabilities correctly.

Concerning children's mass theory acquisition, as investigated in the second article, fluid intelligence did not relate to the consistent explanation of the constructions' stabilities with mass theory at posttest. This indicates that, directly after the playful intervention, children's fluid intelligence did not contribute to their mass theory acquisition. Children's mass knowledge was related to fluid intelligence in the first article, in which no intervention was implemented. Maybe fluid intelligence supports the development of mass knowledge from a developmental perspective. However, the material and verbal scaffolds implemented during the playful intervention in the second article might have supported children with varying degrees of fluid intelligence equally. Therefore, fluid intelligence might not have had an effect on children's mass theory directly after the intervention beyond the effect of the scaffolds. Moreover, after the follow-up was included, fluid intelligence positively related to children's consistent application of mass theory. This indicates that the ability to mentally represent and identify spatial features contributed to children's mass theory acquisition over an extended period of time (Flynn & Blair, 2013). This might be the case, because the follow-up took place approximately ten weeks after the playful intervention, which might have allowed for fluid intelligence to support the development of mass theory once more.

7.2 Crystallised intelligence

Language capacity was assessed as an indicator for crystallised intelligence and affected mass knowledge and mass theory in both the first and the second article, respectively (Cattell, 1987; Weiß & Osterland, 2013). Crystallised intelligence has been found to influence knowledge acquisition, which is underlined by the results of both articles (Brydges et al., 2012; Flynn & Blair, 2013; Thorsen et al., 2014). It was related to children's knowledge that they had already acquired, as indicated by the results of the first article, and children found it easier to integrate this knowledge into their theories (Gopnik & Wellman, 2012), as indicated by the findings of the second article.

Moreover, crystallised intelligence moderated the effect of the playgroup on acquisition of a mass theory. A high crystallised intelligence affected the acquisition of a mass theory in the guided play group with material and verbal scaffolds as well as in the free play group. Children with a low crystallised intelligence were more likely to adjust their theory to a mass theory in the guided play group with material scaffolds. These results indicate that children with a high crystallised intelligence learn under two different circumstances, (1) if they receive a high amount of scaffolding or (2) if they play freely. (1) Concerning the role of the verbal scaffolds—activating prior knowledge, asking for reasoning, providing explanations, encouraging comparisons, and modelling—children with a higher crystallised intelligence may be able to integrate a lot of information at the same time. Therefore, a combination of material and verbal scaffolds may have supported their learning more than just material scaffolds. Through the verbal scaffolds children were encouraged to think about underlying causal relations for stabilities and they also received explanations. The guided play with material and verbal scaffolds might have been a stimulating environment for children with a higher crystallised intelligence, while the guided play with material scaffolds might have not been complex enough. (2) If children have a higher crystallised intelligence, they may engage with materials in a free play differently than children with a lower crystallised intelligence. In the free play group implemented in the second article, children might have interpreted evidence for mass theory that they randomly observed, and integrated it into their prior intuitive theories. Maybe the children even investigated stabilities during their free play (Schulz & Bonawitz, 2007).

For children with a lower crystallised intelligence, the guided play group with material scaffolds best supported their acquisition of a mass theory. Two possible reasons might explain this finding, (1) children with a lower crystallised intelligence might need support, but (2) a high amount of support might overexert them leading to an increased cognitive load (Kirschner, 2002), as the scaffolds might have triggered complex reasoning processes (van de Pol et al., 2010). (1) Compared to the free play group, the children in the guided play group with material scaffolds received structured materials that guided their play. Thus, they were encouraged to engage with constructions' stabilities. However, the children with a lower crystallised intelligence in the free play group might not have been concerned with stabilities (Thorsen et al., 2014). (2) Concerning the verbal scaffolds, children with a lower crystallised intelligence in the guided play group with material and verbal scaffolds might have had a high cognitive load (Kirschner, 2002). They had to process and integrate the evidence they observed through the material scaffolds with the new information about stabilities provided by the verbal scaffolds. Moreover, as children's crystallised intelligence was assessed through their language capacity, children with a higher language capacity might have had an easier time understanding the verbal scaffolds.

Summarising, the results imply that playful learning should consider children's crystallised intelligence indicated by their language capacity. Children with a high crystallised intelligence profit from free play and from a stimulating environment integrating information about causes and encouraging the children to reflect about causal relations themselves. Children with a low crystallised intelligence profit from a structured guided play with material scaffolds that does not overly increase their cognitive load.

7.3 Mental rotation ability

Mental rotation ability and mass knowledge were not related in the first article. Since mental rotation is the ability to mentally represent an object from different perspectives (Newcombe et al., 2013; Shepard & Metzler, 1971), it was assumed that mental rotation ability and mass knowledge might have the same underlying mental processes. It was presumed that children need to mentally represent and abstract spatial features of blocks in order to mentally rotate them and to rate their stability. However, this does not seem to be the case. Interestingly, fluid intelligence was related to mass knowledge and the acquisition of mass theory, indicating that the ability to mentally represent objects and abstract spatial features is of importance. Mentally representing the object from different perspectives, however, does not seem to be important for the correct estimation of stabilities.

7.4 Interest

Interest in block play was not related to children's mass knowledge either. Interest supports engagement in the content a child is interested in and thus affects learning (Crouch et al., 2018; Renninger & Hidi, 2016). Therefore, it was assumed that children's interest in block play might increase their engagement with block building and might have increased their knowledge about stabilities. However, seemingly this was not the case. Possibly children's prior experiences with block play did not relate to their stability knowledge. Yet, their time spent playing with blocks or whether children had blocks at home were not assessed, even though it might have been related to their stability knowledge (Jirout & Newcombe, 2015).

The present dissertation contributed to the investigation of the role of interindividual prerequisites for science learning. Results imply that fluid and crystallised intelligence affect science learning and crystallised intelligence interacts with play form. It is worthwhile to discuss the implementation of the playful interventions, the scaffolds applied, and the consideration of children's cognitive development.

8 Method discussion

In the present dissertation, play was characterized as being voluntary, joyful, without external goals, child-directed, and containing elements of choice. These characteristics are based on the literature on play (Daubert et al., 2018; Fisher et al., 2013; Pellegrini, 2013; Rubin et al., 1983; Rubin & Smith, 2018; Trawick-Smith, 2012; Weisberg et al., 2016; Zosh et al., 2018).The playful interventions in the present dissertation can be discussed in the light of (1) the playfulness of the block building activities, (2) the implementation of the verbal scaffolds, and (3) children's cognitive development.

8.1 Playfulness of the block building activities

Regarding the playfulness of the block building activities in the three play conditions—guided play with material and verbal scaffolds, guided play with material scaffolds, and free play—the fulfilment of the characteristics of play may be reflected. From a sociological and evolutionary perspective, play is often viewed as a category and, therefore an activity is either considered as play or not play (Bateson, 2011; Burghardt, 2011; Pellegrini & Pellegrini, 2012). Following this perspective, Pellegrini (2013) questioned whether block play should be considered play, because children might have goals, such as building a tower, a zoo, etc. However, according to Borriello and Liben (2018), Ferrara et al. (2011), and Rubin et al. (1983), block play can be recognised as play for two reasons. First, children often do not have a specific goal in mind. They might articulate goals, but they will quickly dismiss one goal and set another. Thus, the process of building seems to be more important to them than achieving a goal. Second, if children have a goal, the goal is self-imposed. Thus, the children are free to disregard their goal and move on to something else. This view on play draws from the continuum perspective that is often adopted by developmental psychologists (Daubert et al., 2018; Fisher et al., 2013; Rubin et al., 1983; Rubin & Smith, 2018; Trawick-Smith, 2012; Weisberg et al., 2016; Zosh et al., 2018). According to this perspective, play can be construed as a continuum with activities having different degrees of playfulness. This dissertation follows the continuum perspective. Correspondingly, all three intervention forms can be regarded as play.

The children played *voluntarily,* because participation in the guided and free play was voluntary. Moreover, during the play, children were always free to take breaks or stop playing altogether. Additionally, the play seemed to be *joyful*. Even though, enjoyment was not measured directly, it was indicated by children's persistence that the video recordings and the experimenters' written statements suggested. Approximately 95% of children continued playing for the provided time in every intervention group. Furthermore, a highly inferential rating of the video recordings showed that all children showed playful behaviour during the intervention (Bundy et al., 2001). The children joked around, sang, laughed, etc. Concerning the goals, the guided play had a *goal* as the children were encouraged to investigate stabilities. Although, the children were not actively made aware of this goal, it is possible that they inferred the goal or even merely the existence of it. Therefore, the guided play is not free of external goals, which is in line with Zosh et al. (2018). The free play was free of goals set by an adult, since the children were completely free to build and the experimenter did not make any suggestions. As proposed by Rubin et al. (1983), the children might have had self-imposed goals, such as building a high tower etc. All three activities were *child-directed*, because the children in the guided play groups were free to build different constructions and could choose the material they wanted to play with, while the free play group was completely free in their play. Therefore, the playful activities also contained *elements of choice*.

Concluding, all three play groups met characteristics of play as defined in the continuum view following Zosh et al. (2018). Concerning the guided play group with material and verbal scaffolds, the way in which the verbal support was realised can be discussed.

8.2 The implementation of the verbal scaffolds

Regarding the implementation of the verbal scaffolds for supporting mass knowledge or mass theory, five scaffolding techniques were implemented, i.e., activating prior knowledge, asking for reasoning, providing explanations, encouraging comparisons, and modelling (van de Pol et al., 2010). For the support of science self-concept, three scaffolding techniques were applied, i.e., referring to challenge, competence, and success (Belland et al., 2013).

Children’s *prior knowledge was activated* through asking questions and specifically asking for children’s prior knowledge about building with blocks and stabilities. This might have facilitated the integration of mass knowledge or mass theory into children’s prior knowledge or theories (Gurlitt & Renkl, 2010; Mayer, 1997; Piaget, 1950; Weinert & Helmke, 1998). Moreover, children were *asked about their reasoning* concerning the block constructions’ stabilities (Hsin & Wu, 2011). This might have encouraged them to think about underlying causal relations such as the importance of the mass for stability (Bonawitz et al., 2012; Muentener & Bonawitz, 2017). Furthermore, the experimenters *provided explanations* about stabilities to help children integrate the evidence for the mass into their prior knowledge or their intuitive theories (Murphy & Messer, 2000; Renkl, 2002). Moreover, children were *encouraged to compare* different block constructions with each other regarding their stabilities (Hsin & Wu, 2011; Richey & Nokes-Malach, 2013). Children might have compared symmetrical and asymmetrical constructions’ stabilities, which might have supported children’s mass knowledge or mass theory acquisition. Last, *modelling* was implemented in the form of goal-directed behaviour such as closely looking at a block construction, for children to imitate (Hmelo-Silver et al., 2007; Mayer, 2004).

Concerning children's science self-concept, the experimenters *referred to challenge* in order to enhance children's expectancy of success (Britner & Pajares, 2006). They also *referred to the children's competence* to reassure the children of the competences they have already acquired or might acquire through play (Kaplan & Maehr, 2007). Last, the experimenters *referred to the children's successes* and encouraged them to explain their strategy for building certain block buildings or for estimating stabilities. This might have prompted children to take a closer look at the content and invoke a sense of pride concerning their own achievements (Pintrich, 2003; Schunk et al., 2008).

Regarding the effect of the verbal scaffolds on the playfulness of the interventions, Rubin and Smith (2018) and Franse et al. (2020) suggest that parents talk to their children as well if they play with them and offer explanations on different topics. Those explanations may not be purposefully chosen scaffolds, nevertheless, they may be understood as scaffolding (Vygotsky, 1978). Therefore, the implementation of verbal scaffolds might not narrow the playfulness.

8.3 Children's cognitive development

When developing interventions to foster children's knowledge and intuitive theories, children's cognitive development should be considered (Trundle, 2015). Drawing on findings on 5- to 6-year-old children's reasoning, the playful intervention was developed with regard to (1) the general and specific perspective on cognitive development and (2) deduction, induction, and abduction. Moreover, the cognitive processes influenced by the above-mentioned were considered, i.e., (3) generalisation, (4) categorisation, (5) concepts, and (6) causality. Last, (7) theory theory with Bayesian inference and results on theory-evidence coordination were used to connect these findings.

8.3.1 General and specific perspective

Children's cognitive development is regarded from two different perspectives, the general and the specific perspective (Gelman, 2013; Rakison & Lawson, 2013). The general perspective is mainly concerned with children's learning through general mechanisms such as learning through covariation (Sloutsky & Fisher, 2004), while the specific perspective suggests that children have prior knowledge and theories that influence their learning (Hayes & Thompson, 2007). There is research providing evidence for both perspectives and for the interaction of general and specific learning (Bonawitz et al., 2012; Opfer & Bulloch, 2007).

In the guided play, both perspectives were regarded. Children received evidence in the form of perfect covariation that always supported the mass. As studies suggest that children mainly regard surface features for covariation, unless they have prior knowledge or theories (specific perspective; e.g., Namy & Gentner, 2002), the play was concerned with stabilities, because children have prior knowledge in this domain (Bonawitz et al., 2012; Krist, 2010). Therefore, children likely knew that the size or the colour of a block are not important for stability, but other features such as the centre and the mass are. Thus, children's prior knowledge or theories might have facilitated their learning.

8.3.2 Deduction, induction, and abduction

Children's reasoning can be deductive, inductive, and abductive and is involved in children's science knowledge acquisition (Leuchter & Hardy, 2020). Children's developmental constraints regarding these processes were considered in the guided play.

Children might have used deductive reasoning when they reasoned about stabilities, e.g., *This side is heavier, therefore, it will tumble*. Studies found that children are more likely to use deductive inference when they are prompted to think about alternatives (Chantal & Markovits, 2017). Therefore, children were encouraged to place blocks at different positions on the supporting blocks and estimate the stability or compare the stabilities of different block constructions with each other.

Concerning inductive inference, children might have applied induction as they reasoned that if an asymmetrical construction remained stable even though the centre was unsupported, another construction might remain stable as well. Induction is influenced by prior knowledge (Johnson-Laird & Khemlani, 2017). Accordingly, the stability domain was chosen as a suitable domain, because children have prior knowledge (Bonawitz et al., 2012; Krist, 2010). Moreover, induction is facilitated by similarity, but children fail to acknowledge the role of diversity for their inductive inferences (Lawson & Fisher, 2011). Therefore, stability of symmetrical and asymmetrical block constructions was chosen as the target content. This content is quite small, allowed the limitation of the problem space, and did not need to include many diverse examples. Moreover, according to Johnson (2013), children use monotonicity for their inferences. Thus, children were presented with many examples showing evidence for the importance of the mass, which might have facilitated their induction. Furthermore, causal relations support inductive reasoning (Hayes & Thompson, 2007; Opfer & Bulloch, 2007). Accordingly, children were encouraged to think about underlying relations by asking them for their reasoning. This might have supported children in identifying the mass as an important factor for stability.

Last, children might have also applied abductive inference if they observed an event that violated their prior theory, e.g., their centre theory. They might have searched for other explanations and thus inferred that the mass must be of importance (see Lombrozo & Vasilyeva, 2017). Abductive inference can take the form of Bayesian priors underlining the role of children's prior theories (Gopnik & Wellman, 2012) and theory-evidence coordination highlighting children's ability to draw conclusions from perfect covariance (Koerber et al., 2005). Both were considered, as children's prior theories were regarded. The evidence children observed during the guided play always confirmed mass theory and always disconfirmed centre theory. Therefore, children were confronted with perfect covariance. Moreover, abductive inference may take the form of hypothesising. As Ruffman et al. (1993) found that children are able to form hypothesis if they are confronted with perfect covariation. Although not a primary target of the guided play, children were asked about their hypotheses about stabilities. As already noted, the covariation children observed was perfect. Therefore, this might have also allowed children to form hypotheses. Last, abduction involves variable control, which was not a focus of the present dissertation, as teaching young children the control-of-variables strategy is complex and might need specifically designed interventions (van der Graaf et al., 2015).

Therefore, children's deductive, inductive, and abductive inferences were facilitated through the design of the guided play. As deduction, induction, and abduction as well as the general and the specific perspective inform other cognitive processes, such as generalisation, categorisation, concepts, and causality, these processes were investigated closer for an implementation of the guided play that is suitable for young children.

8.3.3 Generalisation

According to Rehder (2017b), generalisation can take place in the form of (1) object to category, e.g., during the guided play, children might have inferred that if one asymmetrical construction remains stable if its mass is supported, all asymmetrical constructions might remain stable if their masses are supported; (2) category to category, e.g., children might have inferred that if all asymmetrical constructions remain stable if their mass is supported, the same may be true for symmetrical constructions; or (3) object to object, e.g., children might have inferred that block construction A remained stable, because the mass was supported, the same might be true for block construction B. Thus, children might have generalised rules about stability.

All of these generalisations are facilitated by similarity, monotonicity, and causal relations (Hayes & Heit, 2013; Lawson & Fisher, 2011; Rehder, 2017b). Thus, children were confronted with many similar examples, all containing block constructions, i.e., similarity and monotonicity. Moreover, children were encouraged to think about underlying causal relations and experimenters provided causal explanations themselves such as *The block tumbles, because the heavy side hovers in mid-air*. This might have supported generalisation of the mass and children's acquisition of mass knowledge and mass theory. Moreover, concerning causalities, Gelman and Markman (1986) and Hayes and Thompson (2007) found that children recognise relevant and irrelevant features for underlying causalities if they have prior knowledge about a domain. Therefore, children might have identified the mass as a relevant characteristic, but not the colour of the blocks.

8.3.4 Categorisation

Categorisation concerns the grouping of entities and the finding of a superordinate characteristic that all group members share (Rehder, 2017a). Concerning stability, children might have categorised constructions whose mass was supported as stable and constructions whose mass was unsupported as unstable. Moreover, children might have generalised the importance of the mass to different constructions (Rehder, 2017a) and determined that all constructions remain stable if their mass is supported.

Gelman and Markman (1986), Hayes et al. (2008), and Rakison and Hahn (2004) found that children use perceptual similarities as well as non-obvious features for categorisation, because they have theories about what features are relevant for a category. A child might have prior knowledge or a theory that certain features, such as the centre or the mass might be relevant for stabilities, whereas others, such as colour or shape of certain blocks might not. As already mentioned, the stability domain is suitable, because children have prior knowledge (Bonawitz et al., 2012; Krist, 2010).

Moreover, causality plays an important role for categorisation in the form of (1) coherence, (2) explaining away evidence, and (3) essentialism (Gelman, 2013; Rehder, 2017b). (1) Coherence indicates that underlying causal relations between entities or between a new entity and a category facilitate categorisation. Children in the guided play group with material and verbal scaffolds were encouraged to compare different constructions and think about similarities. Moreover, the support of the constructions' mass was the reason for all constructions' stabilities. This might have facilitated children's categorisation of constructions as stable or unstable. (2) Children are more likely to explain away evidence if they perceive other causes as equally likely (Kuhn, 2014; Simons, 2000). Therefore, particular care was taken to disconfirm the importance of the constructions’ centres and highlighting the importance of the mass for stability. Children become less likely to explain away evidence if they observe a lot of counterevidence that forms a pattern (Koslowski & Masnick, 2014). Therefore, children received a lot of evidence for the relevance of the mass. (3) Accordingly, as evidence supported mass knowledge or mass theory, children might have come to consider the support of the mass as essential for stability.

As categorisation is intertwined with concept formation and because concepts may include mental representations of categories, concepts are an important part for children’s cognitive development as well. Therefore, they were considered for the implementation of the guided play.

8.3.5 Concepts

According to Rakison and Lawson (2013), concepts are mental representations of features and objects. They are parsimonious and may contain lists about defining characteristics. Therefore, children might have a concept about constructions’ stabilities that includes reasons for the stability such as support of the centre. Concepts can develop and new information can be included (Gelman, 2013). Children may have concluded that the mass is the deciding factor for stability and expanded their concept of stability during or after the guided play.

In order to foster children’s concept of stability, the characteristics of concepts were considered. First, the structure of a concept is informed by general principles such as surface features, but a concept’s content is informed by children’s specific prior knowledge or prior theories such as centre theory (Bhatt & Quinn, 2011; Gelman, 2013; Schulz, Bonawitz, & Griffiths, 2007). Therefore, the integration of general principles such as visual covariation, e.g., *if the middle is unsupported, the block still remains stable*, and specific principles by considering children’s prior knowledge and theories (Bonawitz et al., 2012; Krist, 2010), might have supported children’s concepts of stability. Second, children actively construct concepts, but are also influenced by their experiences (Gelman, 2013). Accordingly, the guided play confronted children with targeted evidence for mass theory, so that they could experience the importance of the mass during their play. Third, children’s concepts are informed by statistical frequencies as well as their theories (Gelman, 2013). Therefore, the decision to present children with evidence in the form of covariation that contradicts their prior centre theory might support children’s concepts of stability as well. Moreover, Gelman (2013) specifies that surface features often correlate with causal relations and are indicative of theories. This might also be true for stability and children might actively search for causal relations such as support of the mass if they observe that asymmetrical constructions tumble when the heavier side hovers in mid-air, although its centre is supported.

Causality plays an important role in children's cognitive development, as it informs other cognitive processes such as generalisation, categorisation, and concepts (Gopnik, 2013). Thus, causality was considered for the implementation of the guided play as well.

8.3.6 Causality

Gopnik (2013) claims that causality is at the core of both concepts and intuitive theories. Young children might have reasoned that if the centre is supported, the blocks will remain stable. During the guided play, they observed contradicting evidence, which might have encouraged a search for other causes (Bright & Feeney, 2014; Gopnik & Sobel, 2000; Griffiths et al., 2011; Schulz, Bonawitz, & Griffiths, 2007). This search for causal relations is influenced by children's prior knowledge or theories, since they affect which probable causes a child might even consider (Griffiths, 2017). Moreover, children's search for causes is also informed by statistical regularities such as covariation (Kushnir & Gopnik, 2005). Therefore, children might have determined that the colour is unlikely to be the reason for the stability, as both their prior knowledge or theory as well as the evidence they observed during the play (the blocks had different colours) suggested this. Moreover, the children might have inferred that the mass might be important, as it is a likely cause and explains the observed evidence.

Studies on children's cognitive development suggest that children learn through a combination of covariation patterns and their prior knowledge/theories. A theory that is concerned with this relation is theory theory. Therefore, theory theory was applied in the second article and it served as the basis for the implementation of the guided play.

8.3.7 Theory theory

Theory theory with its Bayesian application suggests that children form intuitive theories about their surrounding world that are informed by statistical regularities (Carey, 2009; Gerstenberg & Tenenbaum, 2017; Gopnik, 2013; Gopnik & Wellman, 2012). Thus, theory theory combines the results presented above and integrates cognitive processes, such as generalisation, categorisation, concept formation, and causality, because children generalise evidence, categorise it, and form concepts on the basis of causality (Gopnik, 2013).

Following the implications of theory theory, Bayesian inference and the results on theory-evidence coordination (Gopnik & Wellman, 2012; Koerber et al., 2005), the guided play fulfilled five characteristics to facilitate children's learning, (1) familiarity with the content, (2) perfect covariation of the evidence, (3) the consideration of children's prior theories, (4) confronting children with a sufficient amount of evidence, and (5) giving the children enough time to explore the evidence. The guided play realised these five characteristics by (1) taking a content that children are already familiar with. Children possess prior knowledge and theories about stabilities, which might have facilitated their reasoning (Muentener & Bonawitz, 2017). (2) Moreover, children observed evidence that always supported the importance of the mass for stability and always contradicted the importance of the centre. Thus, children received evidence in the form of perfect covariation that they can interpret and relate to their knowledge and theories (Koerber et al., 2005). (3) Children's prior knowledge and theories were considered by relating them to the evidence they observed during the guided play. Children's knowledge and theories were assessed at pretest and many children had a centre theory. Thus, the provided evidence was designed to contradict centre theory and confirm mass theory. (4) Children with consistent prior theories might have considered their prior theory to be very probable and might have explained away contradicting evidence (Gopnik & Wellman, 2012). As explaining away the evidence becomes less likely if children observe a lot of evidence (Koslowski & Masnick, 2014), children observed a high amount of evidence during the guided play. (5) Last, the children could play with the guided play materials for an hour. Therefore, they had enough time to explore the materials and play with them. Moreover, the content was implemented into block play, which is an every-day activity for young children that they are familiar with from their home or their kindergarten. This might have further supported their knowledge acquisition and theory adaptation.

Concluding, the present dissertation implemented different forms of play based on the suggestions by Zosh et al. (2018). The free play met all characteristics of play, while the guided play met some of them. Children's cognitive development was considered for the implementation of the guided play to support children's knowledge acquisition and theory adaptation in the best possible way. The findings of the present dissertation have implications for science education in kindergarten.

9 Implications for science education in kindergarten years

The results of the present dissertation have implications for science education in the kindergarten years concerning (1) the effects of different play forms on science knowledge, science theories, and science self-concept, (2) the importance of interindividual prerequisites, mainly intelligence, and (3) the consideration of children's cognitive development.

9.1 The effects of play

The results of the second and the third article have implications regarding the implementation of play in kindergarten science education concerning the support of science knowledge, science theories, and children's science self-concept. In most kindergartens in Germany, education is implemented during free play (Arens et al., 2016). However, the studies in the present dissertation indicated that free play does not support children's science knowledge or theories and even decreased their science self-concept, which might undermine children's science achievement in the long run (Marsh et al., 2012). Therefore, free play seems to be insufficient as a means of science education.

Guided play that is still playful but implements an adult's scaffolding through material or verbal scaffolds, seems to be a good alternative to free play if the goal is to teach children about science principles (see also Borriello & Liben, 2018; Ferrara et al., 2011; Verdine et al., 2019). The third article demonstrated that children's science knowledge can be supported through material scaffolds and that verbal scaffolds are not necessary. However, as indicated by the second article, intuitive theories are best supported through a combination of material and verbal scaffolds. Probably, verbal scaffolds are important to trigger children's reasoning about underlying causal relations. In order to support children in gaining theories that they can explain, a combination of material and verbal scaffolds during play seems to be the most beneficial.

Moreover, the results of the third article regarding children' science self-concept highlight the importance of experiences for children (Shavelson et al., 1976). In the guided play group with material and verbal scaffolds and in the guided play group with material scaffolds only, children's science self-concept remained stable. In the free play group, science self-concept decreased. In the guided play groups, the material scaffolds probably provided the children with positive experiences that invoked a sense of achievement, while this did not happen in the free play group. This indicates that the experiences during a playful intervention might be of utmost importance for young children.

Regarding the effect of verbal scaffolds, results of the third article implied that they did not support children's science self-concept in addition to the material scaffolds. The reason might be that the verbal scaffolds were provided by a relative stranger. As prior research suggests that self-concept is influenced by the feedback of attachment figures (Harter, 2015; Helmke, 1999), the verbal scaffolds might have affected children's science self-concept if they had been provided by a kindergarten teacher. This should be considered in interventions aiming at enhancing children's science self-concept.

Concluding, the results of the present dissertation suggest that guided play is a promising way to support children's science learning, while also maintaining their trust in their own science abilities. The present dissertation is also concerned with effects of interindividual prerequisites and the results offer suggestions for educational practice.

9.2 The importance of interindividual prerequisites

The results of the second article allow insights into the relation of intuitive science theories to crystallised intelligence. The study highlighted that children with different levels of crystallised intelligence, indicated by language capacity, profited differentially from different forms of play. Children with lower crystallised intelligence profited most from guided play with material scaffolds, while children with higher crystallised intelligence profited most from guided play with material and verbal scaffolds as well as from free play. This suggests that to support children with lower levels of crystallised intelligence, scaffolds may be applied, however, a high amount of scaffolding might lead to an increased cognitive load (Kirschner, 2002) and thus hinder children's theory acquisition. Children with a high crystallised intelligence may profit from many scaffolds, as they encourage them to think about underlying relations without overly increasing their cognitive load, thus, supporting their theory acquisition.

Therefore, children might be divided into groups with varying levels of support through scaffolding. Additionally, science education might start with only material scaffolds to help children familiarise with the underlying science principles and then introduce verbal scaffolds to foster children's theory adaptation.

9.3 Consideration of children's cognitive development

The literature on children's cognitive development and the results of the second and the third article provide suggestions how guided play should be implemented into science education in kindergarten. Five factors should be considered, (1) children should be familiar with the science content, (2) evidence should be presented in the form of perfect covariation, (3) children's prior knowledge and theories should be regarded, (4) children should observe a lot of evidence, and (5) they should have enough time to explore the materials.

(1) Studies demonstrated that familiarity with a content facilitates children's reasoning and, thus their learning (Croker & Buchanan, 2011; Johnson-Laird & Khemlani, 2017; Namy & Gentner, 2002). The second and third article provided similar results. Therefore, science education should relate to children's every-day life, e.g., stability investigated during block play. (2) Concerning the presentation of evidence, children face problems with imperfect, but not

with perfect covariation (Koerber et al., 2005). Therefore, evidence should be presented in the form of perfect covariation that always confirms the educational content, e.g., the importance of the mass for stability. (3) Moreover, children's prior knowledge and theories influence how children evaluate evidence (Schulz, Bonawitz, & Griffiths, 2007; Schulz, Gopnik, & Glymour, 2007). Therefore, children's prior knowledge and theories should be diagnosed by a kindergarten teacher and the guided play should relate to them. For example, evidence might contradict children's prior knowledge and theories in the way the evidence in the second and third article contradicted the importance of the centre for stability. (4) Children might explain away evidence, but become less likely to do so if they observe a lot of evidence that forms a pattern indicating a systematic cause (Koslowski & Masnick, 2014; Kuhn, 2014). Therefore, children should be able to observe a lot of evidence that allows the inference of underlying causal relations. For example, children who took part in the guided play implemented in the present dissertation had the opportunity to observe a lot of evidence concerning the importance of the mass for stability. (5) Concerning the presentation of a high amount of evidence, children should receive enough time to play with the provided materials and explore them, e.g., a kindergarten teacher might create time slots in which they co-play with the children with the provided materials.

Thus, the present dissertation offers implications for educational practice highlighting the role of guided play for children's knowledge acquisition, theory adaptation, and science self-concept. Moreover, it underlines the importance of considering interindividual prerequisites and children's cognitive development. However, there are certain limitations to the present dissertation.

10 Limitations and future research directions

Regarding the presented studies, several limitations concerning the measurement and the implementation of the playful intervention can be addressed.

10.1 Limitations and future research regarding the measurement

Concerning the measurement, limitations mainly regard the assessment of (1) children's mass knowledge, (2) their intuitive theories, (3) their science self-concept, and (4) other variables of interest such as process knowledge.

(1) The measurement of children's mass knowledge has two possible limitations. First, the COM Test consisted of photographs of different block constructions including red blocks. As red-green deficiency is not an uncommon visual impairment affecting approximately 8% of

the male population in European countries (Birch, 2012), some children in the study samples were likely affected as well. These children might have had a disadvantage. This could be prevented in future studies by asking children to point at the red block and comparing children with a red-green impairment to children without that impairment. If there are differences, children with a red-green impairment should be dropped from data analyses for the COM Test.

Second, the question *What do you think will happen if the black block is removed? Will the red blocks remain stable or will they tumble?* is rather complex. Children need to mentally represent the object depicted on the photograph, mentally remove the black block and imagine what might happen. This process requires abstraction capability and inhibitory control (Carlson et al., 2013). Both might be confounding variables that could be investigated in future studies to further validate the COM Test.

(2) Regarding the measurement of children's intuitive theories, they were assessed with a single interview, in which children were asked to estimate and explain asymmetrical block constructions' stabilities. The children received feedback between the items, since they were allowed to test their estimation of each of the constructions' stabilities. Results showed that children did not acquire a mass theory during testing, however, the items were dependent on each other. As a result, the items could not be summed up, but were treated as single points in time. Therefore, statistical methods referring to the mean could not be conducted. As children's intuitive theories are categorical in nature, mean procedures might have been inappropriate to investigate theory adaptation and procedures for categorical variables were also applied in other studies on children's intuitive theories (Bonawitz et al., 2012).

A second limitation regards the presentation of the symmetrical items in addition to the asymmetrical constructions. Even though only the three asymmetrical items were considered in the second article, children received eight items in total, of which one asymmetrical item was excluded, because it was too difficult. The four remaining items were symmetrical items that were presented to familiarise children with the procedure and to motivate them during testing. However, these items supported centre theory, because for symmetrical items the consideration of the centre is sufficient to rate their stabilities (Krist, 2010). Therefore, some children might have acquired a centre theory instead of remaining with their other theory, even though evidence for centre theory was imperfect (cf. Koerber et al., 2005). This could be tested in a future study by comparing the intuitive theories of children who received all eight items with children who only received the three asymmetrical items.

(3) The limitation regarding the measurement of science self-concept has already been addressed in the discussion of research question 5 concerned with the reciprocal effects between

mass knowledge and science self-concept. The limitation regarding the reciprocal effects was concerned with the science self-concept content, since science self-concept was assessed in relation to block play, not stability knowledge. Asking the children about their science self-concept in the stability domain might have produced different results. Therefore, assessing children's science self-concept in relation to their stability knowledge and relating it to children's mass knowledge could be a focus of a future study.

(4) The fourth limitation regards the missing measurement of process knowledge. The focus of the present dissertation was the investigation of children's specific knowledge about stability, which is an important part of science education (Klahr et al., 2011; van der Graaf et al., 2020). Besides specific knowledge, process knowledge is an important part of science education as well (Klahr et al., 2011; van der Graaf et al., 2020). The guided play might have also fostered children's process knowledge, such as observing, testing presumptions, interpreting, and generalizing. Whether guided play fosters process knowledge in addition to specific knowledge could be investigated in a future study.

10.2 Limitations and future research regarding the implementation of the playful intervention

Other limitations of the present dissertation are concerned with the implementation of the playful interventions. These limitations involve (1) the time spent playing, (2) the implementation of other intervention groups, (3) the comparisons of the verbal scaffolds, (4) children's behaviour during the play, (5) the experimenters' behaviour during the play, and (6) possible effects of knowing the goal.

(1) Regarding the time of the intervention, each playful intervention took approximately an hour, which is quite brief. Many other intervention studies aiming at enhancing children's knowledge, intuitive theories, and self-concept implemented their interventions over an extended period of time spanning from multiple weeks (Craven et al., 1991; Marsh & Richards, 1988) to multiple years (Mantzicopoulos et al., 2013; Samarapungavan et al., 2011). However, for this dissertation, it was decided to only play with the children for an hour for three reasons. First and foremost, playing for an hour was a first step to achieve ecological validity. In Germany, most kindergarten teachers will not teach a subject for an extended period of time (Arens et al., 2016). Since the playful intervention was intended to be as ecologically valid as possible, an hour of playtime was deemed sufficient. Second, the stability domain is a rather small topic, which may be acquired in a relatively short time. Third, the playtime in other studies on guided play with building blocks was equally short or even shorter. Nevertheless, the researchers found

results on the respective outcomes (Borriello & Liben, 2018; Ferrara et al., 2011). In line with these findings, the second and the third article also found that even an hour of guided play supports science learning and stabilises science self-concept. These enhancements were even long lasting, as indicated by the results of the follow-up test. Therefore, the present dissertation underlines and expands earlier findings that even a relatively short guided play session can support science learning and simultaneously stabilise science self-concept. Nevertheless, more interventions implemented over a longer period of time with more points of measurement might support children's science learning and their science self-concept even further, and might allow insights into children's science learning trajectories.

(2) A second limitation regards the possible implementation of other intervention groups. In the present dissertation, three play groups with a different amount of guidance were compared, guided play with material and verbal scaffolds, guided play with material scaffolds, and free play. Following Zosh et al. (2018) and Dunbar and Klahr (2012), other intervention forms such as guided discovery, co-opted play, playful instruction and direct instruction, might also support children's learning. However, these interventions are not considered as play, since they differ in the amount that the five characteristics of play—voluntariness, joyfulness, no external goals, child-directedness, elements of choice—are fulfilled (see Table 1 in the General Introduction). By comparing these different intervention forms on their effect on science learning and science self-concept, the relevance of the characteristics for both of the outcomes could be investigated. Additionally, the study design could be extended to include a baseline group, in which children receive no treatment. This would allow for the comparison of the different intervention forms to receiving no treatment. All of these intervention forms could be investigated and compared in future studies.

(3) A third limitation of the playful intervention involves the closer investigation of the verbal scaffolds that might be relevant in three ways. First, to investigate the effects of the verbal scaffolds in comparison to the material scaffolds, two other intervention groups might be implemented. In one group, children could play with the blocks freely and receive the same verbal scaffolds that the guided play group with material and verbal scaffolds received in the present dissertation. In a second group, the verbal scaffolds could be implemented in the form of direct instruction. Thus, the children would receive explanations without play (Fisher et al., 2013). Both of these group would allow the comparison of the verbal scaffolds to the material scaffolds in a playful or an instruction setting, and might be examined in a future study.

Second, in the present dissertation possible differences of the provided verbal scaffolds—activating prior knowledge, asking for reasoning, providing explanations, encouraging comparisons, and modelling—regarding their effects on stability knowledge or children’s intuitive theories were not investigated. For example, asking for children’s reasoning might have supported children’s causal reasoning and thus their acquisition of mass theory in a different way than modelling. The same applies to the verbal scaffolds targeting children’s science self-concept—referring to challenge, competence, and success. Although the verbal scaffolds did not affect children’s science self-concept in addition to the material scaffolds, results might be different if an attachment figure such as a kindergarten teacher or a parent had provided these verbal scaffolds (Harter, 2015). Investigating whether the different verbal scaffolds provided by a significant person have different effects on motivational and competence beliefs might be of interest.

Third, the adaptability of the verbal scaffolds was not investigated, but might have affected children’s knowledge or theory acquisition and their science self-concept. Investigating the verbal scaffolds’ adaptivity might have yielded insights into children’s individual learning and developmental processes. Each of the three factors might be of interest in future studies.

(4) A fourth limitation regards children’s behaviour during play. For manipulation check, the playful interventions were partly videotaped if the children and their parents had consented to it. As some children or their parents had denied permission, not all play sessions were recorded and thus the videos were not analysed. Nevertheless, a detailed analysis of the available videos concerning the children’s behaviour might have supported the analysis of children’s mass knowledge or mass theory acquisition. Regarding children’s behaviour, children might have differed in the amount of time spent building or interacting with each other. Some children might have spent more time building, while others might have watched others build. Moreover, time spent playing with each other or alone might have had an influence on mass knowledge or theory acquisition as well, because children might have talked about the stabilities and might have supported each other while building. The material scaffolds provided in the guided play groups served as suggestions for building, but the children were free to build something else. Drawing from the available videos, children used the material scaffolds for their play. However, some children might have built more buildings at a higher pace than others possibly affecting their learning. Moreover, concerning children’s science self-concept, peer feedback could be investigated further (Harter, 2015). Maybe some children were cheered on by their peers, while others failed to finish buildings and needed support from others, and others might have supported their peers. All these possibilities might have affected children’s science self-concept and could be investigated in the future using detailed video recordings of playful interventions targeting children’s science self-concept.

(5) A fifth limitation regards the investigation of the experimenters' behaviour. All experimenters followed a script for each of the play groups. However, experimenters might have talked and supported some children more than others. Investigating the effect of experimenter behaviour might be of interest in a future study.

(6) The last limitation regards a possible effect of knowing the goal (Lin et al., 2010). The participating children were not actively made aware of the goal of the playful intervention, i.e., acquiring mass knowledge or mass theory. However, the children in the guided play groups were actively encouraged to investigate stabilities through the material and verbal scaffolds, whereas the children in the free play condition were not. Therefore, children in the guided play groups might have inferred the goal of the playful intervention. Whether knowing the goal has an effect on the acquisition of mass knowledge or mass theory, could be investigated by implementing a free play group that is instructed to investigate stabilities during their block play at the beginning of the play session. Afterwards, the children would play freely with the blocks just as they did in the free play condition of the present dissertation.

Despite these limitations, the present dissertation contributes to the research on young children's science education and their science self-concept. The three articles highlighted ways to investigate children's domain-specific science knowledge, their intuitive theories and their science self-concept, as well as the relations with other interindividual prerequisites. Moreover, guided play may support children's science knowledge and intuitive theories, and maintain the trust in their abilities and in relation with children's cognitive prerequisites.

11 Conclusion

The aim of the present dissertation was to contribute to the research on children's science knowledge, their science theories, and science self-concept and how they may be fostered through play. Moreover, the present dissertation relied on research on children's cognitive development for the implementation of the guided play. The three articles revealed that children's science knowledge in the domain of stability can be measured reliably and is a suitable educational science content for young children. Moreover, both knowledge and theories can be fostered through guided play that maintains children's science self-concept and considers children's cognitive development. Findings of the present dissertation offer educational implications how to implement guided play in kindergarten science education. Therefore, it contributes to the research on science education in the early years of childhood.

References

Andersen, L., & Ward, T. J. (2014). Expectancy-value models for the STEM persistence plans of ninth-grade, high-ability students: A comparison between black, hispanic, and white students. *Science Education*, *98*(2), 216–242. https://doi.org/10.1002/sce.21092

Arens, A. K., & Hasselhorn, M. (2015). Differentiation of competence and affect self-perceptions in elementary school students: Extending empirical evidence. *European Journal of Psychology of Education*, *30*(4), 405–419. https://doi.org/10.1007/s10212-015-0247-8

Arens, A. K., Marsh, H. W., Craven, R. G., Yeung, A. S., Randhawa, E., & Hasselhorn, M. (2016). Math self-concept in preschool children: Structure, achievement relations, and generalizability across gender. *Early Childhood Research Quarterly*, *36*, 391–403. https://doi.org/10.1016/j.ecresq.2015.12.024

Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept: Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, *103*(4), 970–981. https://doi.org/10.1037/a0025047

Baddeley, A. (1986). *Working memory*. Oxford University Press.

Baillargeon, R., & Hanko-Summers, S. (1990). Is the top object adequately supported by the bottom object? Young infants' understanding of support relations. *Cognitive Development*, *5*(1), 29–53. https://doi.org/10.1016/0885-2014(90)90011-H

Baillargeon, R., Li, J., Gertner, Y., & Di Wu. (2014). How do infants reason about physical events? In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 11–48). Wiley-Blackwell.

Baillargeon, R., Needham, A., & Devos, J. (1992). The development of young infants' intuitions about support. *Early Development and Parenting*, *1*(2), 69–78. https://doi.org/10.1002/edp.2430010203

Ball, C., Huang, K.-T., Cotten, S. R., & Rikard, R. V. (2017). Pressurizing the STEM pipeline: An expectancy-value theory analysis of youths' STEM attitudes. *Journal of Science Education and Technology*, *26*(4), 372–382. https://doi.org/10.1007/s10956-017-9685-1

Bateson, P. (2011). Theories of play. In A. D. Pellegrini (Ed.), *The Oxford handbook of the development of play* (pp. 41–47). Oxford Univ. Press.

Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, *48*(4), 243–270. https://doi.org/10.1080/00461520.2013.838920

Bhatt, R. S., & Quinn, P. C. (2011). How does learning impact development in infancy? The case of perceptual organization. *Infancy*, *16*(1), 2–38. https://doi.org/10.1111/j.1532-7078.2010.00048.x.

Birch, J. (2012). Worldwide prevalence of red-green color deficiency. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *29*(3), 313–320. https://doi.org/10.1364/JOSAA.29.000313

Bjorklund, D. F. (2013). Cognitive development: An overview. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 447–476). Oxford Univ. Press.

Bonawitz, E. B., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Bonawitz, E. B., Ullman, T. D., Bridgers, S., Gopnik, A., & Tenenbaum, J. B. (2019). Sticking to the evidence? A behavioral and computational case study of micro-theory change in the domain of magnetism. *Cognitive Science*, *43*(8), e12765. https://doi.org/10.1111/cogs.12765

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Borriello, G. A., & Liben, L. S. (2018). Encouraging maternal guidance of preschoolers' spatial thinking during block play. *Child Development*, *89*(4), 1209–1222. https://doi.org/10.1111/cdev.12779

Bright, A. K., & Feeney, A. (2014). Causal knowledge and the development of inductive reasoning. *Journal of Experimental Child Psychology*, *122*, 48–61. https://doi.org/10.1016/j.jecp.2013.11.015

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*(5), 485–499. https://doi.org/10.1002/tea.20131

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, *20*(4), 493–523. https://doi.org/10.1016/0010-0285(88)90014-X

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist*, *19*(1), 1–15. https://doi.org/10.1037/h0044160

Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth*. Wiley.

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, *40*(5), 458–469. https://doi.org/10.1016/j.intell.2012.05.006

Bundy, A. C., Nelson, L., Metzger, M., & Bingaman, K. (2001). Validity and reliability of a test of playfulness. *The Occupational Therapy Journal of Research*, *21*(4), 276–292. https://doi.org/10.1177/153944920102100405

Burghardt, G. M. (2011). Defining and recognizing play. In A. D. Pellegrini (Ed.), *The Oxford handbook of the development of play* (pp. 9–18). Oxford Univ. Press.

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 706–743). Oxford Univ. Press.

Case, R. (1985). *Intellectual Development: Birth to adulthood*. Academic Press. https://doi.org/10.1111/j.2044-8279.1986.tb02666.x

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Cattell, R. B. (1987). *Intelligence: It's structure, growth and action*. Elsevier.

Chantal, P.-L. de, & Markovits, H. (2017). The capacity to generate alternative ideas is more important than inhibition for logical reasoning in preschool-age children. *Memory & Cognition*, *45*(2), 208–220. https://doi.org/10.3758/s13421-016-0653-4

Chen, Z., Sanchez, R., & Campbell, T. (1997). From beyond to within their grasp: The rudiments of analogical problem solving in 10- to 13-month-olds. *Developmental Psychology*, *33*, 790–801.

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching*, *44*(6), 815–843. https://doi.org/10.1002/tea.20171

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development*, *21*(4), 713–731. https://doi.org/10.1016/S0163-6383(98)90040-6

Copple, C., & Bredenkamp, J. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8.* (3rd ed.). National Association for the Education of Young Children.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology*, *83*(1), 17–27. https://doi.org/10.1037/0022-0663.83.1.17

Croker, S., & Buchanan, H. (2011). Scientific reasoning in a real-world context: The effect of prior belief and outcome on children's hypothesis-testing strategies. *The British Journal of Developmental Psychology*, *29*(3), 409–424. https://doi.org/10.1348/026151010X496906

Crouch, C. H., Wisittanawat, P., Cai, M., & Renninger, K. A. (2018). Life science students' attitudes, interest, and performance in introductory physics for life sciences: An exploratory study. *Physical Review Physics Education Research*, *14*(1). https://doi.org/10.1103/PhysRevPhysEducRes.14.010111

Daubert, E., Ramani, G. B., & Rubin, K. H. (2018). Play-based learning and social development. In M. Boivin, R. D. Peters, & R. Tremblay (Eds.), *Encyclopedia on early childhood development.* Centre of Excellence for Early Childhood Development and the Strategic Knowledge Cluster on Early Child Development.

DeLoache, J. S. (2014). Early development of the understanding of symbolic artifacts. In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 312–336). Wiley-Blackwell.

Demetriou, A., Christou, C., Spanoudis, G., & Platsidou, M. (2002). The development of mental processing: Efficiency, working memory, and thinking. *Monographs of the Society for Research in Child Development*, *57*.

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, *78*(2), 430–447. https://doi.org/10.1111/j.1467-8624.2007.01007.x.

Dunbar, K. N., & Klahr, D. (2012). Scientific thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning.* Oxford handbooks online.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist*, *44*(2), 78–89. https://doi.org/10.1080/00461520902832368

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). W. H. Freeman.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, *64*(3), 830. https://doi.org/10.2307/1131221

Ehm, J.-H., Hasselhorn, M., & Schmiedek, F. (2019). Analyzing the developmental relation of academic self-concept and achievement in elementary school children: Alternative models point to different results. *Developmental Psychology*, *55*(11), 2336–2351. https://doi.org/10.1037/dev0000796

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology*, *14*(3), 315–336. https://doi.org/10.1007/s10956-005-7198-9

Evans, J. St. B. T. (2013). Reasoning. In D. Reisberg (Ed.), *Oxford library of psychology. The Oxford handbook of cognitive psychology* (pp. 635–649). Oxford Univiversity Press.

Feeney, A., & Heit, E. (2011). Properties of the diversity effect in category-based inductive reasoning. *Thinking & Reasoning*, *17*(2), 156–181. https://doi.org/10.1080/13546783.2011.566703

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block talk: Spatial language during block play. *Mind, Brain, and Education*, *5*(3), 143–151. https://doi.org/10.1111/j.1751-228X.2011.01122.x

Fisher, K. R., Hirsh-Pasek, K., Golinkoff, R. M., Singer, D. G., & Berk, L. E. (2011). Playing around in school: Implications for learning and educational policy. In A. D. Pellegrini (Ed.), *The Oxford handbook of the development of play* (pp. 341–362). Oxford Univ. Press.

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development*, *84*(6), 1872–1878. https://doi.org/10.1111/cdev.12091

Flynn, J. R., & Blair, C. (2013). The history of intelligence: New spectacles for developmental psychology. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 765–790). Oxford Univ. Press.

Franse, R. K., van Schijndel, T. J. P., & Raijmakers, M. E. J. (2020). Parental pre-knowledge enhances guidance during inquiry-based family learning in a museum context: An individual differences perspective. *Frontiers in Psychology*, *11*, 1047. https://doi.org/10.3389/fpsyg.2020.01047

Friston, K. (2012). A free energy principle for biological systems. *Entropy*, *14*(11), 2100–2121. https://doi.org/10.3390/e14112100

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology*, *74*(2), 435–452.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, *14*(1), 79–106. https://doi.org/10.1207/s15516709cog1401_5

Gelman, S. A. (2013). Concepts in development. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 542–563). Oxford Univ. Press.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*, *26*(5), 796–804.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, *58*(6), 1532–1541. https://www.jstor.org/stable/1130693

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*(4), 487–513. https://doi.org/10.1016/S0885-2014(99)00016-7

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (pp. 515–548). Oxford University Press.

Goddu, M. K., Lombrozo, T., & Gopnik, A. (2020). Transformations and transfer: Preschool children understand abstract relations and reason analogically in a causal task. *Child Development.* Advance online publication. https://doi.org/10.1111/cdev.13412

Golinkoff, R. M., Hirsh-Pasek, K., & Singer, D. G. (Eds.). (2006a). *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth*. Oxford University Press.

Golinkoff, R. M., Hirsh-Pasek, K., & Singer, D. G. (2006b). Why play = learning: A challenge for parents and educators. In R. M. Golinkoff, K. Hirsh-Pasek, & D. G. Singer (Eds.), *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth* (pp. 3–12). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195304381.003.0001

Gopnik, A. (2013). Causality. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 628–650). Oxford Univ. Press.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. MIT Press.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development, 71*(5), 1205–1222. https://doi.org/10.1111/1467-8624.00224

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37*(5), 620–629. https://doi.org/10.1037//0012-1649.37.5.620

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085–1108. https://doi.org/10.1037/a0028044

Goswami, U. (2014). Inductive and deductive reasoning. In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 399–419). Wiley-Blackwell.

Grégoire, É. (2001). Generalization v. (Induction v. Abduction). In *Proceedings of the International Conference on Artificial Intelligence (IC-CAI'2001),* Las Vegas, USA.

Griffin, S. (1992). Young children's awareness of their inner world: A neo-structured analysis of the development of intrapersonal intelligence. In R. Case (Ed.), *The mind's staircase* (pp. 189–206). Erlbaum.

Griffiths, T. L. (2017). Formalizing prior knowledge in causal induction. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (pp. 115–126). Oxford University Press.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, *35*(8), 1407–1455. https://doi.org/10.1111/j.1551-6709.2011.01203.x

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: A ten-year longitudinal study. *Self and Identity, 3*, 53–68.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, *20*(6), 644–653. https://doi.org/10.1016/j.lindif.2010.08.001

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, *37*, 161–168. https://doi.org/10.1016/j.lindif.2015.01.008

Gurlitt, J., & Renkl, A. (2010). Prior knowledge activation: How different concept mapping tasks lead to substantial differences in cognitive processes, learning outcomes, and perceived self-efficacy. *Instructional Science*, *38*(4), 417–433. https://doi.org/10.1007/s11251-008-9090-5

Guthrie, J. T., Wigfield, A., & Perencevich, K. C. (2004). Scaffolding for motivation and engagement in reading. In J. T. Guthrie, A. Wigfield, & K. C. Perencevich (Eds.), *Motivating reading comprehension: Concept-oriented reading instruction* (pp. 55–86). L. Erlbaum Associates.

Haden, C. A. (2010). Talking about science in museums. *Child Development Perspectives*, *4*(1), 62–67. https://doi.org/10.1111/j.1750-8606.2009.00119.x

Hadzigeorgiou, Y. (2002). A study of the development of the concept of mechanical stability in preschool children. *Research in Science Education*, *32*(3), 373–391. https://doi.org/10.1023/A:1020801426075

Halford, G. S., & Andrews, G. (2006). Reasoning and problem solving. In D. Kuhn, R. S. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 2 Cognition, perception, and language* (6th ed., pp. 557–608). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0213

Harris, P. L., & Nuntez, M. (1996). Understanding of permission rules by preschool children. *Child Development*, *67*(4), 1572–1591. https://doi.org/10.1111/j.1467-8624.1996.tb01815.x

Harter, S. (1999). *The construction of the self: A developmental perspective.* Guilford Press.

Harter, S. (2015). *The construction of the self: Developmental and sociocultural foundations* (Second edition, paperback edition). Guilford Press.

Hassinger-Das, B., Toub, T. S., Zosh, J. M., Michnick, J., Golinkoff, R. M., & Hirsh-Pasek, K. (2017). More than just fun: a place for games in playful learning / Más que diversión: El lugar de los juegos reglados en el aprendizaje lúdico. *Infancia Y Aprendizaje*, *40*(2), 191–218. https://doi.org/10.1080/02103702.2017.1292684

Hawkins, J., Pea, R. D., Glick, J., & Scribner, S. (1984). "Merds that laugh don't like mushrooms": Evidence for deductive reasoning by preschoolers. *Developmental Psychology*, *20*(4), 584–594. https://doi.org/10.1037/0012-1649.20.4.584

Hayes, B. K., & Heit, E. (2013). Induction. In D. Reisberg (Ed.), *Oxford library of psychology. The Oxford handbook of cognitive psychology* (pp. 618–634). Oxford Univiversity Press.

Hayes, B. K., McKinnon, R., & Sweller, N. (2008). The development of category-based induction: Reexamining conclusions from the induction then recognition (ITR) paradigm. *Developmental Psychology*, *44*(5), 1430–1441. https://doi.org/10.1037/0012-1649.44.5.1430

Hayes, B. K., & Rehder, B. (2012). The development of causal categorization. *Cognitive Science*, *36*(6), 1102–1128. https://doi.org/10.1111/j.1551-6709.2012.01244.x

Hayes, B. K., & Thompson, S. P. (2007). Causal relations and feature similarity in children's inductive reasoning. *Journal of Experimental Psychology: General*, *136*(3), 470–484. https://doi.org/10.1037/0096-3445.136.3.470

Helmke, A. (1999). From optimism to realism? Development of children’s academic self-concept from kindergarten to grade 6. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 198–221). University Press.

Helmke, A., & van Aken, M. A. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, *87*(4), 624–637.

Hidi, S. E., Renninger, K. A., & Northoff, G. (2017). The development of interest and self-related processing. In F. Guay, H. W. Marsh, D. M. McInerney, & R. G. Craven (Eds.), *International advances in self research: Vol. 6. Self. Driving positive psychology and well-being* (pp. 51–70). Information Age Publishing.

Hinkley, T., Brown, H., Carson, V., & Teychenne, M. (2018). Cross sectional associations of screen time and outdoor play with social skills in preschool children. *PloS One*, *13*(4), e0193700. https://doi.org/10.1371/journal.pone.0193700

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. G. (2009). *A mandate for playful learning in preschool: Presenting the evidence*. University Press.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107.

Hogan, K., & Pressley, M. (1997). Becoming a scaffolder of students' learning. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning* (pp. 185–191). Brookline.

Hsin, C.-T., & Wu, H.-K. (2011). Using scaffolding strategies to promote young children’s scientific understandings of floating and sinking. *Journal of Science Education and Technology*, *20*(5), 656–666. https://doi.org/10.1007/s10956-011-9310-7

Jäncke, L., Kleinschmidt, A., Mirzazade, S., Shah, N. J., & Freund, H. J. (2001). The role of the inferior parietal cortex in linking the tactile perception and manual construction of object shapes. *Cerebral Cortex*, *11*(2), 114–121. https://doi.org/10.1093/cercor/11.2.114

Jansen, M. M. (2017). Academic self-concept in the sciences: Domain-specific differentiation, gender differences, and dimensional comparison effects. In F. Guay, H. W. Marsh, D. M. McInerney, & R. G. Craven (Eds.), *International advances in self research: Vol. 6. Self. Driving positive psychology and well-being* (pp. 71–112). Information Age Publishing.

Jansen, M. M., Lüdtke, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology*, *46*, 116–127. https://doi.org/10.1016/j.cedpsych.2016.05.004

Jansen, M. M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30,* 11–21. https://doi.org/10.1016/j.lindif.2013.12.003

Jansen, M. M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and assimilation effects of dimensional comparisons in five subjects: An extension of the I/E model. *Journal of Educational Psychology*, *107*(4), 1086–1101. https://doi.org/10.1037/edu0000021

Jiang, Z., Waters, A. C., Liu, Y., Li, W., & Yang, L. (2017). Event-related theta oscillatory substrates for facilitation and interference effects of negative emotion on children's cognition. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *116*, 26–31. https://doi.org/10.1016/j.ijpsycho.2017.02.012

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. Sample. *Psychological Science*, *26*(3), 302–310. https://doi.org/10.1177/0956797614563338

Johnson, S. P. (2013). Object perception. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 371–379). Oxford Univ. Press.

Johnson-Laird, P. N., & Khemlani, S. S. (2017). Mental models and causation. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (pp. 147–168). Oxford University Press.

Kaplan, A., Katz, I., & Flum, H. (2012). Motivation theory in educational practice: Knowledge claims, challenges, and future directions. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *Apa educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* (1st ed., pp. 165–194). American Psychological Association.

Kaplan, A., & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review, 19*(2), 141–184. https://doi.org/10.1007/s10648-006-9012-5

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition, 3*(3), 195–212. https://doi.org/10.1016/0010-0277(74)90008-0

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*, 1–10.

Klahr, D., & Li, J. (2005). Cognitive Research and Elementary Science Instruction: From the Laboratory, to the Classroom, and Back. *Journal of Science Education and Technology, 14*(2), 217–238. https://doi.org/10.1007/s10956-005-4423-5

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*(6045), 971–975. https://doi.org/10.1126/science.1204528

Kleickmann, T., Tröbst, S., Jonen, A., Vehmeyer, J., & Möller, K. (2016). The effects of expert scaffolding in elementary science professional development on teachers' beliefs and motivations, instructional practices, and student achievement. *Journal of Educational Psychology, 108*(1), 21–42. https://doi.org/10.1037/edu0000041

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*(3), 141–152. https://doi.org/10.1024/1421-0185.64.3.141

Köksal-Tuncer, Ö., & Sodian, B. (2018). The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cognitive Development, 48*, 135–145. https://doi.org/10.1016/j.cogdev.2018.06.011

Koslowski, B., & Masnick, A. (2014). Causal reasoning and explanation. In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 377–398). Wiley-Blackwell.

Kotovsky, L., & Baillargeon, R. (2000). Reasoning about collisions involving inert objects in 7.5-month-old infants. *Developmental Science*, *3*(3), 344–359. https://doi.org/10.1111/1467-7687.00129

Krist, H. (2010). Development of intuitions about support beyond infancy. *Developmental Psychology*, *46*(1), 266–278. https://doi.org/10.1037/a0018040

Krist, H. (2013). Development of intuitive statics. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *45*(1), 27–33. https://doi.org/10.1026/0049-8637/a000078

Krist, H., Atlas, C., Fischer, H., & Wiese, C. (2018). Development of basic intuitions about physical support during early childhood: Evidence from a novel eye-tracking paradigm. *Quarterly Journal of Experimental Psychology*, *2*, 174702181773719. https://doi.org/10.1177/1747021817737196

Krist, H., Horz, H., & Schönfeld, T. (2005). Children’s block balancing revisited: No evidence for representational redescription. *Swiss Journal of Psychology*, *64*(3), 183–193. https://doi.org/10.1024/1421-0185.64.3.183

Kuhn, D. (2013). Reasoning. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 744–764). Oxford Univ. Press.

Kuhn, D. (2014). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 497–523). Wiley-Blackwell.

Kuhn, D., & Ho, V. (1980). Self-directed activity and cognitive development. *Journal of Applied Developmental Psychology*, *1*(2), 119–133. https://doi.org/10.1016/0193-3973(80)90003-9

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*(9), 678–683. https://doi.org/10.1111/j.1467-9280.2005.01595.x

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology*, *43*(1), 186–196. https://doi.org/10.1037/0012-1649.43.1.186

Lauermann, F., Tsai, Y.-M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy-value theory of achievement-related behaviors. *Developmental Psychology*, *53*(8), 1540–1559. https://doi.org/10.1037/dev0000367

Lawson, A. E. (2010). Basic inferences of scientific reasoning, argumentation, and discovery. *Science Education*, *94*(2), 336-364. https://doi.org/10.1002/sce.20357

Lawson, C. A., & Fisher, A. V. (2011). It's in the sample: The effects of sample size and sample diversity on the breadth of inductive generalization. *Journal of Experimental Child Psychology*, *110*(4), 499–519. https://doi.org/10.1016/j.jecp.2011.07.001

Legare, C. H., & Gelman, S. A. (2008). Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science*, *32*(4), 607–642. https://doi.org/10.1080/03640210802066766

Leibham, M. B., Alexander, J. M., & Johnson, K. E. (2013). Science interests in preschool boys and girls: relations to later self-concept and science achievement. *Science Education*, *97*(4), 574–593. https://doi.org/10.1002/sce.21066

Leuchter, M., & Hardy, I. (2020). Wissenschaftliches Denken und Argumentieren [Scientific thinking and argumentation]. *Manuscript Submitted for Publication*.

Leuchter, M., & Naber, B. (2019). Studying children's knowledge base of one-sided levers as force amplifiers. *Journal of Research in Science Teaching*, *56*(1), 91–112. https://doi.org/10.1002/tea.21470

Leuchter, M., Saalbach, H., & Hardy, I. (2014). Designing science learning in the first years of schooling: An intervention study with sequenced learning material on the topic of ‘floating and sinking'. *International Journal of Science Education*, *36*(10), 1751–1771.

Lin, X., Siegler, R. S., & Sullivan, F. R. (2010). Students' goals influence their learning. In D. D. Preiss & R. J. Sternberg (Eds.), *Innovations in Educational Psychology: Perspectives on learning, teaching and human development* (pp. 79–105). Springer.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Lombrozo, T., & Vasilyeva, M. (2017). Causal explanation. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (pp. 415–432). Oxford University Press.

Love, B. C. (2017). Concepts, meaning, and conceptual relationships. In S. E. F. Chipman (Ed.), *The Oxford handbook of cognitve science* (pp. 137–150). Oxford University Press.

Mandler, J. M. (1992). How to build a baby: Ii. Conceptual primitives. *Psychological Review*, *99*(4), 587–604. https://doi.org/10.1037/0033-295X.99.4.587

Mandler, J. M. (2003). Conceputal categorization. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 103–131). Oxford University Press.

Mantzicopoulos, P., & Patrick, H. (2010). “The seesaw is a machine that goes up and down”: Young children's narrative responses to science-related informational text. *Early Education and Development*, *21*(3), 412–444. https://doi.org/10.1080/10409281003701994

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2013). Science literacy in school and home contexts: Kindergarteners’ science achievement and motivation. *Cognition and Instruction*, *31*(1), 62–119. https://doi.org/10.1080/07370008.2012.742087

Markman, A. B., & Rein, J. R. (2013). The nature of mental concepts. In D. Reisberg (Ed.), *Oxford library of psychology. The Oxford handbook of cognitive psychology* (pp. 321–329). Oxford Univiversity Press.

Markovits, H., & Thompson, V. (2008). Different developmental patterns of simple deductive and probabilistic inferential reasoning. *Memory & Cognition*, *36*(6), 1066–1078. https://doi.org/10.3758/MC.36.6.1066

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*, 129–149.

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. British Psychological Society.

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, *80*(3), 366–380. https://doi.org/10.1037/0022-0663.80.3.366

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Ellis, L. A., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology*, *38*(3), 376–393. https://doi.org/10.1037/0012-1649.38.3.376

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology*, *39*, 326–341.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction*, *35*, 16–32. https://doi.org/10.1016/j.learninstruc.2014.08.005

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Parker, P. D., & Craven, R. G. (2015). Dimensional comparison theory: An extension of the internal/external frame of reference model. In F. Guay, H. W. Marsh, D.

M. McInerney, & R. G. Craven (Eds.), *Self-concept, motivation, and identity: Underpinning success with research and practice* (pp. 115–150). Information Age Publishing.

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, *54*(2), 263–280. https://doi.org/10.1037/dev0000393

Marsh, H. W., & Richards, G. (1988). The Outward Bound Bridging Course for low achieving high-school males: Effect on academic achievement and multidimensional self-concepts. *Australian Journal of Psychology*, *40*, 281–298.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*, 107–125.

Marsh, H. W., Trautwein, U., Ludtke, O., Koller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Vol. 1. Theories, constructs, and critical issues* (pp. 427–458). American Psychological Association. https://doi.org/10.1037/13273-015

Martin, N. D., Dornfeld Tissenbaum, C., Gnesdilow, D., & Puntambekar, S. (2019). Fading distributed scaffolds: The importance of complementarity between teacher and material scaffolds. *Instructional Science*, *47*(1), 69–98. https://doi.org/10.1007/s11251-018-9474-0

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist*, *59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838

Miller, R. S., & Wang, M.-T. (2019). Cultivating adolescents' academic identity: Ascertaining the mediating effects of motivational beliefs between classroom practices and mathematics identity. *Journal of Youth and Adolescence*, *48*(10), 2038–2050. https://doi.org/10.1007/s10964-019-01115-x

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education*, *21*(4), 467–487. https://doi.org/10.1007/BF03173515

Muentener, P., & Bonawitz, E. B. (2017). The development of causal reasoning. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (pp. 677–698). Oxford University Press.

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86. https://doi.org/10.1016/j.cogpsych.2010.02.001

Murayama, K., Pekrun, R., Lichtenfeld, S., & vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development*, *84*(4), 1475–1490. https://doi.org/10.1111/cdev.12036

Murphy, N., & Messer, D. (2000). Differential benefits from scaffolding and children working alone. *Educational Psychology*, *20*(1), 17–31. https://doi.org/10.1080/014434100110353

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*(1), 5–15. https://doi.org/10.1037//0096-3445.131.1.5

Nayfeld, I., Brenneman, K., & Gelman, R. (2011). Science in the classroom: Finding a balance between autonomous exploration and teacher-led instruction in preschool settings. *Early Education and Development*, *22*(6), 970–988. https://doi.org/10.1080/10409289.2010.507496

Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, *47*, 121–148.

Newcombe, N. S., Uttal, David, H., & Sauter, M. (2013). Spatial development. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 564–590). Oxford Univ. Press.

OECD. (2014). *Measuring innovation in education: A new perspective, educational research and innovation*. *Educational Research & Innovation*. OECD Publishing. https://doi.org/10.1787/20769679

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, *41*(3), 181–206. https://doi.org/10.1207/s15326985ep4103_4

Opfer, J. E., & Bulloch, M. J. (2007). Causal relations drive young children's induction, naming, and categorization. *Cognition*, *105*(1), 206–217. https://doi.org/10.1016/j.cognition.2006.08.006

Oppermann, E., Brunner, M., Eccles, J. S., & Anders, Y. (2017). Uncovering young children's motivational beliefs about learning science. *Journal of Research in Science Teaching, 24*(2), 195. https://doi.org/10.1002/tea.21424

Patrick, H., & Mantzicopoulos, P. (2015). Young children's motivation for learning science. In K. C. Trundle & M. Saçkes (Eds.), *Research in early childhood science education* (pp. 7–34). Springer.

Patrick, H., Mantzicopoulos, P., & Samarapungavan, A. (2009). Motivation for learning science in kindergarten: Is there a gender gap and does integrated inquiry and literacy instruction make a difference. *Journal of Research in Science Teaching*, *46*(2), 166–191. https://doi.org/10.1002/tea.20276

Pellegrini, A. D. (2013). Play. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol 2: Self and other* (pp. 276–299). Oxford Univ. Press.

Pellegrini, A. D., & Bjorklund, D. F. (2004). The ontogeny and phylogeny of children's object and fantasy play. *Human Nature*, *15*(1), 23–43. https://doi.org/10.1007/s12110-004-1002-z

Pellegrini, A. D., & Pellegrini, A. F. A. (2012). Play, plasticity, and ontogeny in childhood. In D. Narvaez, J. Panksepp, A. N. Schore, & T. R. Gleason (Eds.), *Evolution, early experience and human development* (pp. 339–351). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199755059.003.0021

Pellegrini, A. D., & Smith, P. K. (1998). Physical activity play: The nature and function of a neglected aspect of play. *Child Development*, *69*(3), 557–598.

Piaget, J. (1950). *The psychology of intelligence*. Harcourt Brace Jovanovich.

Piaget, J. (1952). *Play, dreams, and imitation in childhood*. Norton.

Piekny, J., Grube, D., & Maehler, C. (2014). The Development of Experimentation and Evidence Evaluation Skills at Preschool Age. *International Journal of Science Education*, *36*(2), 334–354. https://doi.org/10.1080/09500693.2013.776192

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, *31*(Pt 2), 153–179. https://doi.org/10.1111/j.2044-835X.2012.02082.x

Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology, 40*(6), 1059–1067. https://doi.org/10.1037/0012-1649.40.6.1059

Pine, K. J., & Messer, D. (1998). Group collaboration effects and the explicitness of childrens' knowledge. *Cognitive Development, 13*, 109–126.

Pine, K. J., & Messer, D. J. (2000). The effect of explaining another's actions on children's implicit theories of balance. *Cognition and Instruction*, *18*(1), 35–51. https://doi.org/10.1207/S1532690XCI1801_02

Pine, K. J., & Messer, D. (2003). The development of representations as children learn about balancing. *British Journal of Developmental Psychology*, *21*, 285–301.

Pine, K. J., Messer, D., & Godfrey, K. (1999). The teachability of children with naive theories: An exploration of the effects of two teaching methods. *British Journal of Educational Psychology*, *69*, 201–211.

Pine, K. J., Messer, D., & St. John, K. (2002). Children’s learning from contrast modelling. *Cognitive Development*, *17*, 1249–1263.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Quaiser-Pohl, C. (2003). The mental cutting test "Schnitte" and the Picture Rotation Test—two new measures to assess spatial ability. *International Journal of Testing*, *3*(3), 219–231.

Quaiser-Pohl, C., Rohe, A. M., & Amberger, T. (2010). The solution strategy as an indicator of the developmental stage of preschool children’s mental-rotation ability. *Journal of Individual Differences*, *31*(2), 95–100. https://doi.org/10.1027/1614-0001/a000017

Quinn, P. C., Bhatt, R. S., & Hayden, A. (2008). What goes with what? Development of perceptual grouping in infancy. *Psychology of Learning and Motivation*, *49*, 105–146. https://doi.org/10.1016/S0079-7421(08)00003-0

Quinn, P. C., & Eimas, P. D. (2000). The emergence of category representations during infancy: Are separate perceptual and conceptual processes required? *Journal of Cognition and Development*, *1*(1), 55–61. https://doi.org/10.1207/S15327647JCD0101N_6

Rakison, D. H., & Hahn, E. R. (2004). The mechanisms of early categorization and induction: Smart or dumb infants? In R. V. Vail (Ed.), *Advances in child development and behavior* (Vol. 32, pp. 281–322). Elsevier Academic Press. https://doi.org/10.1016/S0065-2407(04)80010-7

Rakison, D. H., & Lawson, C. A. (2013). Categorization. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol. 1: Body and mind* (pp. 591–627). Oxford Univ. Press.

Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development*, *73*(1), vii, 1-110. https://doi.org/10.1111/j.1540-5834.2008.00454.x

Rehder, B. (2017a). Concepts as causal models: Categorization. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (pp. 347–375). Oxford University Press.

Rehder, B. (2017b). Concepts as causal models: Induction. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (pp. 377–413). Oxford University Press.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529–556. https://doi.org/10.1016/S0959-4752(01)00030-5

Renninger, K. A., & Hidi, S. (2016). *The power of interest for motivation and engagement*. Routledge.

Reuter, T., & Leuchter, M. (2020). Children's concepts of gears and their promotion through play. *Journal of Research in Science Teaching*, *Advance online publication.* https://doi.org/10.1002/tea.21647

Reynolds, E., Stagnitti, K., & Kidd, E. (2011). Play, language and social skills of children attending a play-based curriculum school and a traditionally structured classroom curriculum school in low socioeconomic areas. *Australasian Journal of Early Childhood*, *36*(4), 120–130. https://doi.org/10.1177/183693911103600416

Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition*, *108*(2), 543–556. https://doi.org/10.1016/j.cognition.2008.03.002

Richey, J. E., & Nokes-Malach, T. J. (2013). How much is too much? Learning and motivation effects of adding instructional explanations to worked examples. *Learning and Instruction*, *25*, 104–124. https://doi.org/10.1016/j.learninstruc.2012.11.006

Riley, W. F., & Sturges, L. D. (1993). *Engineering mechanics: Statics* (2nd ed.). Wiley.

Rochat, P. (2013). Self-conceptualizing in development. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Vol 2: Self and other* (pp. 378–397). Oxford Univ. Press.

Rubin, K. H., Fein, G. G., & Vandenberg, B. (1983). Play. In E. M. Hetherington (Ed.), *Handbook of child psychology* (pp. 693–774). Wiley.

Rubin, K. H., & Smith, K. (2018). Play in human development. In M. H. Bornstein (Ed.), *The SAGE encyclopedia of lifespan human development.* SAGE Publications, Inc.

Rubin, K. H., Watson, K. S., & Jambor, T. W. (1978). Free-Play behaviors in preschool and kindergarten children. *Child Development*, *49*(2), 534–536. https://doi.org/10.2307/1128725

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, *64*(6), 1617–1636. https://doi.org/10.2307/1131459

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Saçkes, M., Trundle, K. C., Bell, R. L., & O'Connell, A. A. (2011). The influence of early science experience in kindergarten on children's immediate and later science achievement: Evidence from the early childhood longitudinal study. *Journal of Research in Science Teaching*, *48*(2), 217–235. https://doi.org/10.1002/tea.20395

Samarapungavan, A., Patrick, H., & Mantzicopoulos, P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cognition and Instruction*, *29*(4), 416–470. https://doi.org/10.1080/07370008.2011.608027

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology*, *43*(5), 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162–176. https://doi.org/10.1037/0012-1649.40.2.162

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science*, *10*(3), 322–332. https://doi.org/10.1111/j.1467-7687.2007.00587.x

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research, and applications* (3. ed.). Pearson/Merrill Prentice Hall.

Schwartenbeck, P., Fitzgerald, T., Dolan, R. J., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Frontiers in Psychology*, *4*, 710. https://doi.org/10.3389/fpsyg.2013.00710

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science (New York, N.Y.)*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481–520.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analyses. *Cognitive Psychology*, *36*, 273–310.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, *4*(4), 147–155. https://doi.org/10.1016/s1364-6613(00)01455-8

Simons, D. J., & Keil, F. C. (1995). An abstract to concrete shift in the development of biological thought: The insides story. *Cognition*, *56*(2), 129–163.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modelling change and event occurence*. Oxford Univ. Press.

Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education*, *67*(2), 135–149. https://doi.org/10.1080/00220979909598349

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166–188. https://doi.org/10.1037/0096-3445.133.2.166

Sobel, D. M., & Buchanan, D. W. (2009). Bridging the gap: Causality-at-a-distance in children's categorization and inferences about internal properties. *Cognitive Development*, *24*(3), 274–283. https://doi.org/10.1016/j.cogdev.2009.03.003

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*(6), 1103–1115. https://doi.org/10.1037/0012-1649.42.6.1103

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental Science*, *10*(3), 298–306. https://doi.org/10.1111/j.1467-7687.2007.00589.x

Sobel, D. M., & Munro, S. E. (2009). Domain generality and specificity in children's causal inference about ambiguous data. *Developmental Psychology*, *45*(2), 511–524. https://doi.org/10.1037/a0014944

Sobel, D. M., & Sommerville, J. A. (2009). Rationales in children’s causal learning from others’ actions. *Cognitive Development*, *24*(1), 70–79. https://doi.org/10.1016/j.cogdev.2008.08.003

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, *28*(3), 303–333. https://doi.org/10.1016/j.cogsci.2003.11.001

Sodian, B. (2018). The development of scientific thinking in preschool and elementary school age: A conceptual model. In F. Fischer, C. A. Chinn, K. F. Engelmann, & J. Osborne (Eds.), *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge* (pp. 227–250). Routledge.

Sylva, K., Taggart, B., Siraj-Blatchford, I., Totsika, V., Ereky-Stevens, K., Gilden, R., & et al. (2007). Curricular quality and day-to-day learning activities in pre-school. *International Journal of Early Years Education*, *15*(1), 49–65.

Thorsen, C., Gustafsson, J.-E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology*, *84*(4), 556–570. https://doi.org/10.1111/bjep.12041

Trautwein, U., Nagengast, B., Marsh, H. W., Gaspard, H., Dicke, A.-L., Lüdtke, O., & Jonkmann, K. (2013). Expectancy-value theory revisited: From expectancy-value theory to expactancy-values theory? In D. M. McInerney, H. W. Marsh, & R. G. Craven (Eds.), *International advances in self research. Theory driving research: New wave perspectives on self-processes and human development* (pp. 233–249). Information Age Publishing Inc.

Trawick-Smith, J. (2012). Teacher-child play interactions to achieve learning outcomes. In R. Pianta (Ed.), *Handbook of early childhood education* (pp. 259–277). Guilford Publications.

Trawick-Smith, J. (2015). Teacher–child play interactions to achieve learning outcomes. In K. C. Trundle & M. Saçkes (Eds.), *Research in early childhood science education* (pp. 259–277). Springer.

Trundle, K. C. (2015). The inclusion of science in early childhood classrooms. In K. C. Trundle & M. Saçkes (Eds.), *Research in early childhood science education* (pp. 1–6). Springer.

Tytler, R., & Prain, V. (2010). A framework for re-thinking learning in science from recent cognitive science perspectives. *International Journal of Science Education*, *32*(15), 2055–2078. https://doi.org/10.1080/09500690903334849

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

van der Graaf, J., Segers, E., & Jong, T. de (2020). Fostering integration of informational texts and virtual labs during inquiry-based learning. *Contemporary Educational Psychology*, *62*, 101890. https://doi.org/10.1016/j.cedpsych.2020.101890

van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: dynamic assessment of the control of variables strategy. *Instructional Science*, *43*(3), 381–400. https://doi.org/10.1007/s11251-015-9344-y

Verdine, B. N., Zimmermann, L., Foster, L., Marzouk, M. A., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2019). Effects of geometric toy design on parent-child interactions and spatial language. *Early Childhood Research Quarterly*, *46*, 126–141. https://doi.org/10.1016/j.ecresq.2018.03.015

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Harvard Univ. Press.

Waldmann, M. R. (2017). Causal reasoning: An introduction. In Waldmann, M., R. (Ed.), *Oxford handbook of causal reasoning* (Vol. 1, pp. 1–9). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.1

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*(2), 343–357. https://doi.org/10.1016/j.cognition.2014.07.008

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420. https://doi.org/10.1037/0022-0663.99.2.397

Weber, A. M., & Leuchter, M. (2020). Measuring preschool children's knowledge of the principle of static equilibrium in the context of building blocks: Validation of a test instrument. *The British Journal of Educational Psychology*, *90*, 50–74. https://doi.org/10.1111/bjep.12304

Weber, A. M., Reuter, T., & Leuchter, M. (2020). The impact of a construction play on 5- to 6-year-old children’s reasoning about balance relationships. *Frontiers in Psychology*, *11*, 1737. https://doi.org/10.3389/fpsyg.2020.01737

Weinert, F. E., & Helmke, A. (1998). The neglected role of individual differences in theoretical models of cognitive development. *Learning and Instruction*, *8*(4), 309–323. https://doi.org/10.1016/S0959-4752(97)00024-8

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play. *Current Directions in Psychological Science*, *25*(3), 177–182. https://doi.org/10.1177/0963721416645512

Weisberg, D. S., Zosh, J. M., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Talking it up: Play, language development, and the role of adult support. *American Journal of Play*, *6*, 39–54.

Weiß, R. H., & Osterland, J. (2013). *CFT 1-R*. Hogrefe.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*(3), 451–469.

Wigfield, A., Tonks, S., & Klauda, S. L. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55–75). Routledge.

Wilkening, F., & Cacchione, T. (2014). Children's intuitive physics. In U. Goswami (Ed.), *Wiley-Blackwell handbooks of developmental psychology. The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 473–496). Wiley-Blackwell.

Wood, D., Brunner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17,* 89–100.

Woodward, A., Phillips, A., & Spelke, E. (1993). Infants' expectations about the motions of inanimate vs. animate objects. In *Proceedings of the Cognitive Science Society.* Erlbaum.

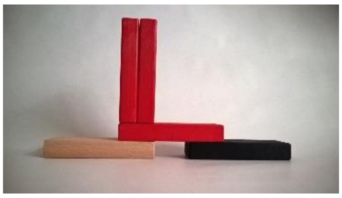
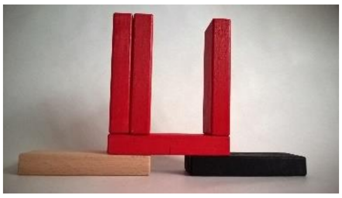
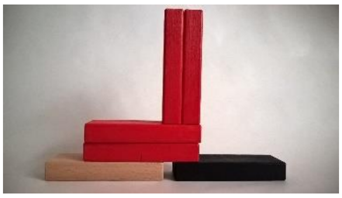
Zhang, L. (2016). Is inquiry-based science teaching worth the effort? *Science & Education, 25*(7-8), 897–915. https://doi.org/10.1007/s11191-016-9856-0

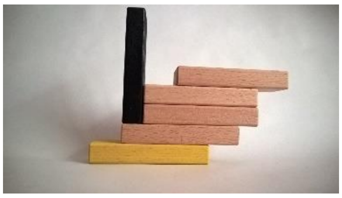
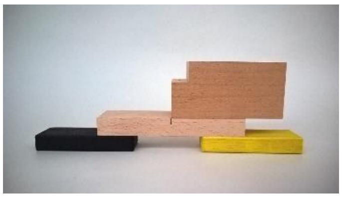
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Solis, S. L., & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology, 9*, 1124. https://doi.org/10.3389/fpsyg.2018.01124

Appendix A: Supplementary Material 1. Material scaffolds for all five activities played during the intervention in the Material group and the Verbal group.

1. Black block:

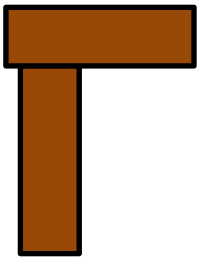



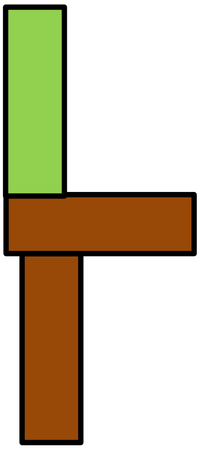


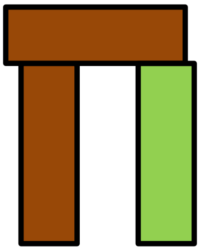

2. Add-a-block:

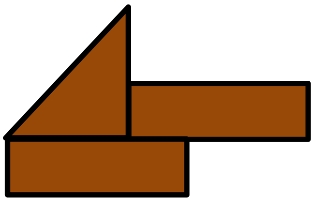
Picture children received:

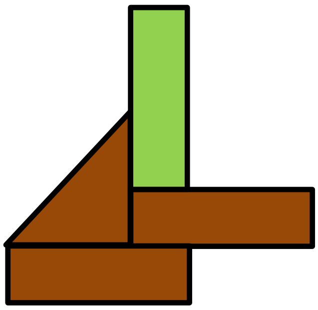
Possible solutions for the experimenters only:

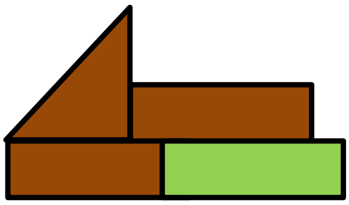


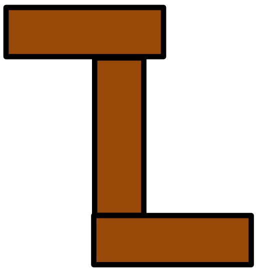


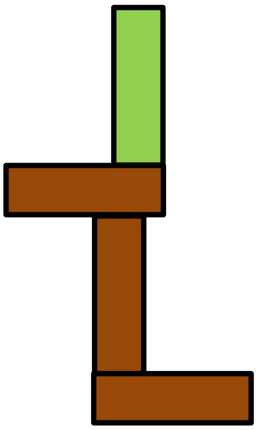


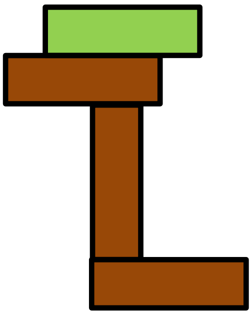


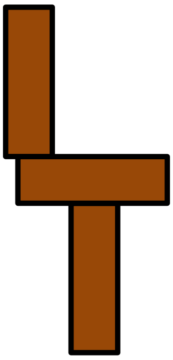


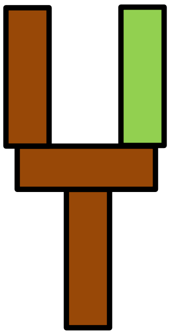


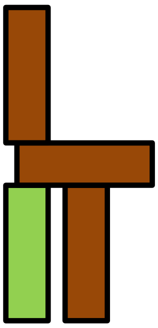




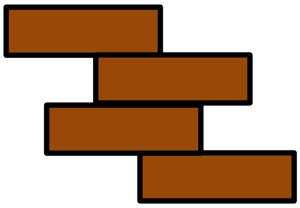


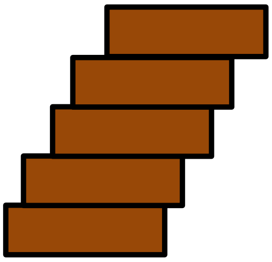


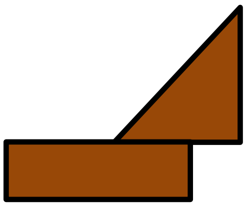


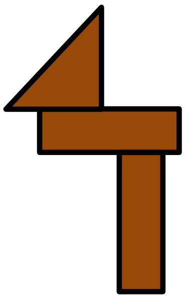


Picture children received:

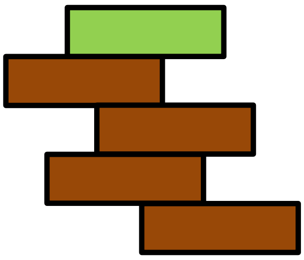


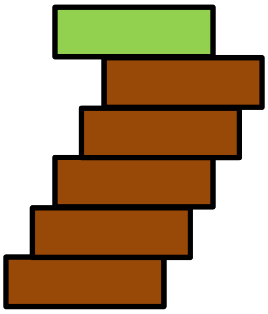


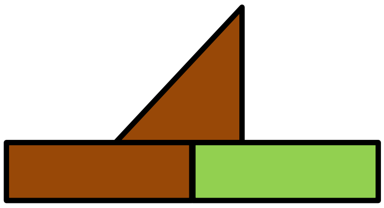


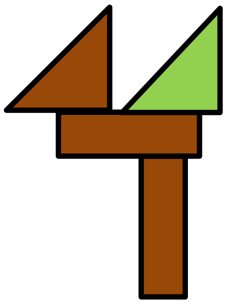


Possible solutions for the experimenters only:

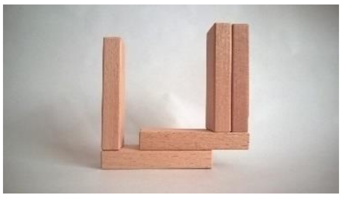


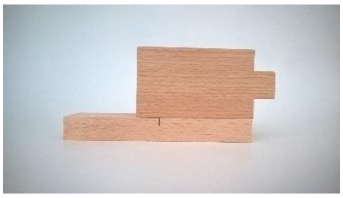
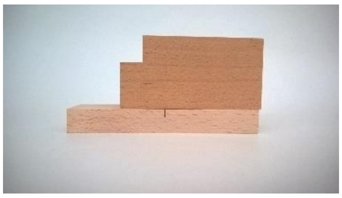
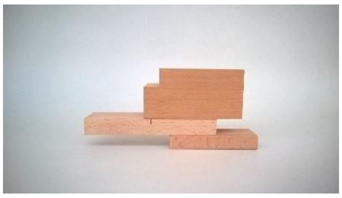


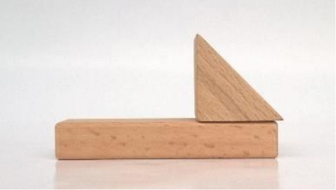
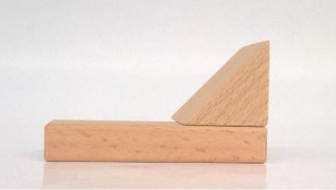




3.Sliding:



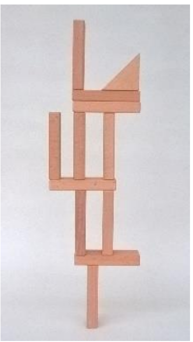

4. Rebuild:



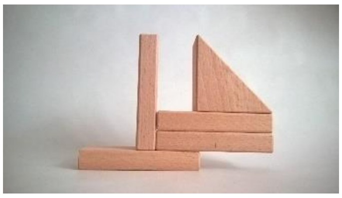
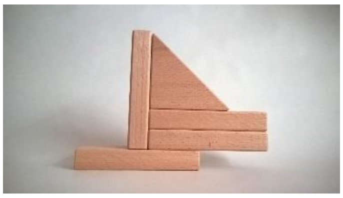
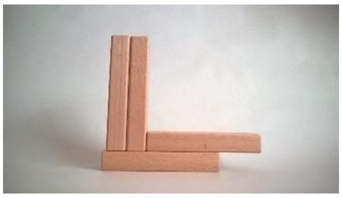


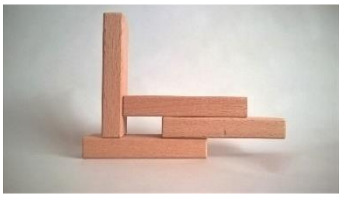




5. Stable/Tumble:


Appendix B: Supplementary Material 2. Script for the verbal scaffolds used during the intervention in the Verbal group.

Activating prior knowledge	Have you ever seen something like this?
	What do you think? Will this remain stable or will it tumble over?
	Do you remember the games we played last time? Now you can try it out yourself.
	When you play with blocks, sometimes blocks tumble over. How high did you build without them tumbling?
	Can you slide these blocks very carefully? Let go of them when they are just supported on the other block.
Encouraging comparisons	Look! What is the difference?
	This building looks different than this one, doesn't it? What is different?
	Look, x built this and y that. Try sliding your blocks along so your building looks like x's.
	Show me, how you have to slide these blocks along.
	How can you compare …?
	What is heavier/lighter/wider/narrower/etc.?
Asking for reasons	Why do you think that?
	What do you mean?
	Why?
	What did you find out?
	Why does it tumble/stand?
	Can you explain this to (finger puppet)?
	Can you explain this to me more clearly?
	Why is that so?
	How did you do this?
	Where are more blocks/is it heavier? On the brown block or on the black block?
	How can you balance this out?
Providing explanations	Exactly! It depends on how the building blocks stand, how much they touch each other.

	Exactly! The building blocks don't always have to stand on the middle to remain standing.
	The building will only remain intact if there are more blocks on the brown block/it is heavier on the brown block.
	If the side with more blocks/the heavier side is dangling in the air, everything tumbles over.
Modelling	Look at these black lines on the blocks. They indicate the block's middle. How do you have to slide these blocks along so that they remain standing or tumble?
	Look at this! (Experimenter looks closely)
	(Experimenter points)
	(Experimenter guides child's gestures)
	(Experimenter turns building blocks with lines and shows how to use them)
	Observe closely.

Appendix C: Supplementary Material 3. Example excerpts from the playful activities (translated from German).

Verbal group, 1 girl and 2 boys from a kindergarten in the periphery:

The children are all building and loudly singing a funny children's song, occasionally they look at each other and chuckle.

Child 1: I will show you that this will remain standing.

Experimenter (to child 1): Look! Do you see this little line (on the picture provided)? There is also a line on this (shows child 1 a line drawn on a building block). This line shows you where you need to place the block's middle.

Child 1: Like this.

Child 2 (motions to his building and looks at child 1 excitedly): Will it tumble or not, child 1?

Child 1: Show me. That will not tumble.

Child 2: I agree.

Experimenter: Child 3, what do you think?

Child 3: It tumbles…?

Child 2 starts to remove the black block.

Experimenter (to child 2): Wait for child 3.

Child 3: Doesn't tumble.

Experimenter: Why not?

Child 1: Here is more burden again (motions to the weight).

Child 3: I don't know.

Experimenter: Okay, but maybe what child 1 said is correct.

Child 2 (excited): Because there is more space, because there is more space!

Child 1 (leans over the table to child 2 and pointing at the building explains to him): There is quite a lot more here (motions to heavier side). There and there is only a bit.

Child 2 removes the black block, smiles and waves his arms: It remains standing!

Experimenter: Super! Ace!

Child 2 smiles proudly.

Experimenter: Because there are more blocks on this side, it remains standing.

Child 2: Yeah.

Material group, 2 girl and 2 boys from a kindergarten in a village:

The children are building in pairs.

Child 1 (excitedly to child 2): That will stand, I know that.

Child 2: I agree.

Child 3 (to child 3): I know that this will stand. And you?

Child 4: Hm, that has to be farther away from another.

Child 3: Yeah.

Child 1 (smiling proudly at experimenter): Yes, it falls. We did know that.

Experimenter nods.

Child 1: Should we build that next? (pulls blocks and picture to himself and child 2)

Experimenter: Sure.

Child 1: I know that this will tumble.

Child 3 and 4 have finished building, they smile at each other and then at the experimenter proudly.

Child 4: Will it stand or tumble?

Child 3: I think that it won’t stand.

Child 1 starts singing a made-up song.

Child 4 removes black block, while child 3 stabilizes building. She lets go, building tumbles. Both children smile a bit sadly at each other.

Child 3 (to child 4 in a comforting voice): Oh well. That’s not that bad.

Child 1 restarts singing his made-up song.

Child 3 and 4 choose another picture and the blocks.

Child 3: Eh? What do we have to do there? Do we have to put the black block over here and on there?

Experimenter: Exactly.

Free play group, 2 boys from a kindergarten in a city:

Child 1: What should we be building?

They both build in silence.

Child 1: Oh, I need that one.

Child 2: I need this one.

Child 1 accidentally shakes the table, while he reaches for more blocks. His building tumbles, he looks at child 2 accusingly. Child 2 starts to laugh and child 1 joins in. Both resume building.

Child 1 starts singing while building.

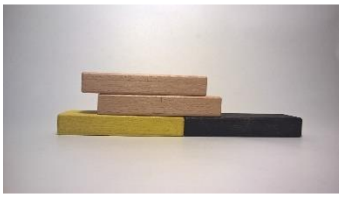
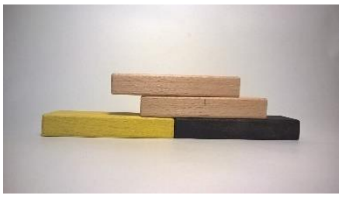
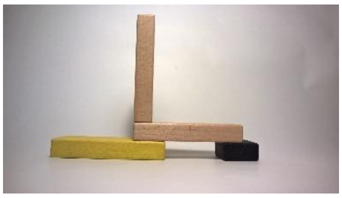
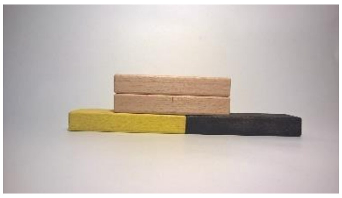
Child 2 reaches for the blocks and accidentally shakes the table, his building and child 1’s building both tumble. Child 2 laughs.

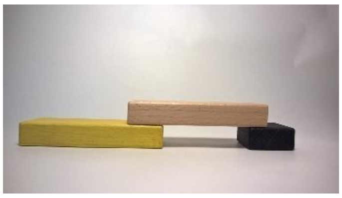
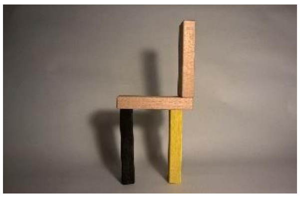
Child 1: Child 2, you shake everything!

Child 2: No, I don’t.

Child 1 starts to build something different: Look, I’ll build it like that.

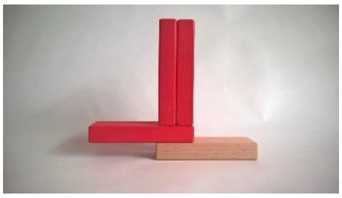
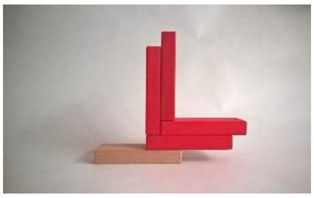
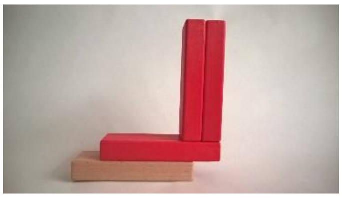
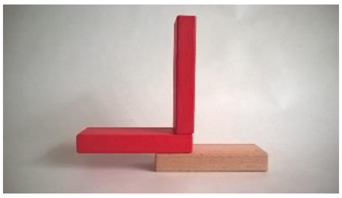
Appendix D: Supplementary Material 4. Items of the reasoning test.

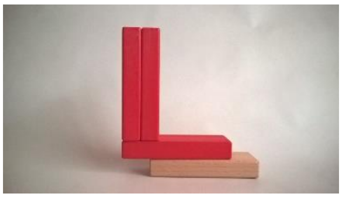





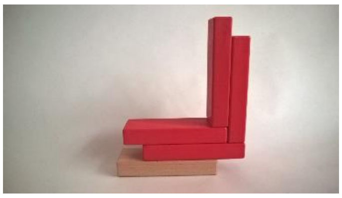




Appendix E: Supplementary Material 5. Items of the transfer test.



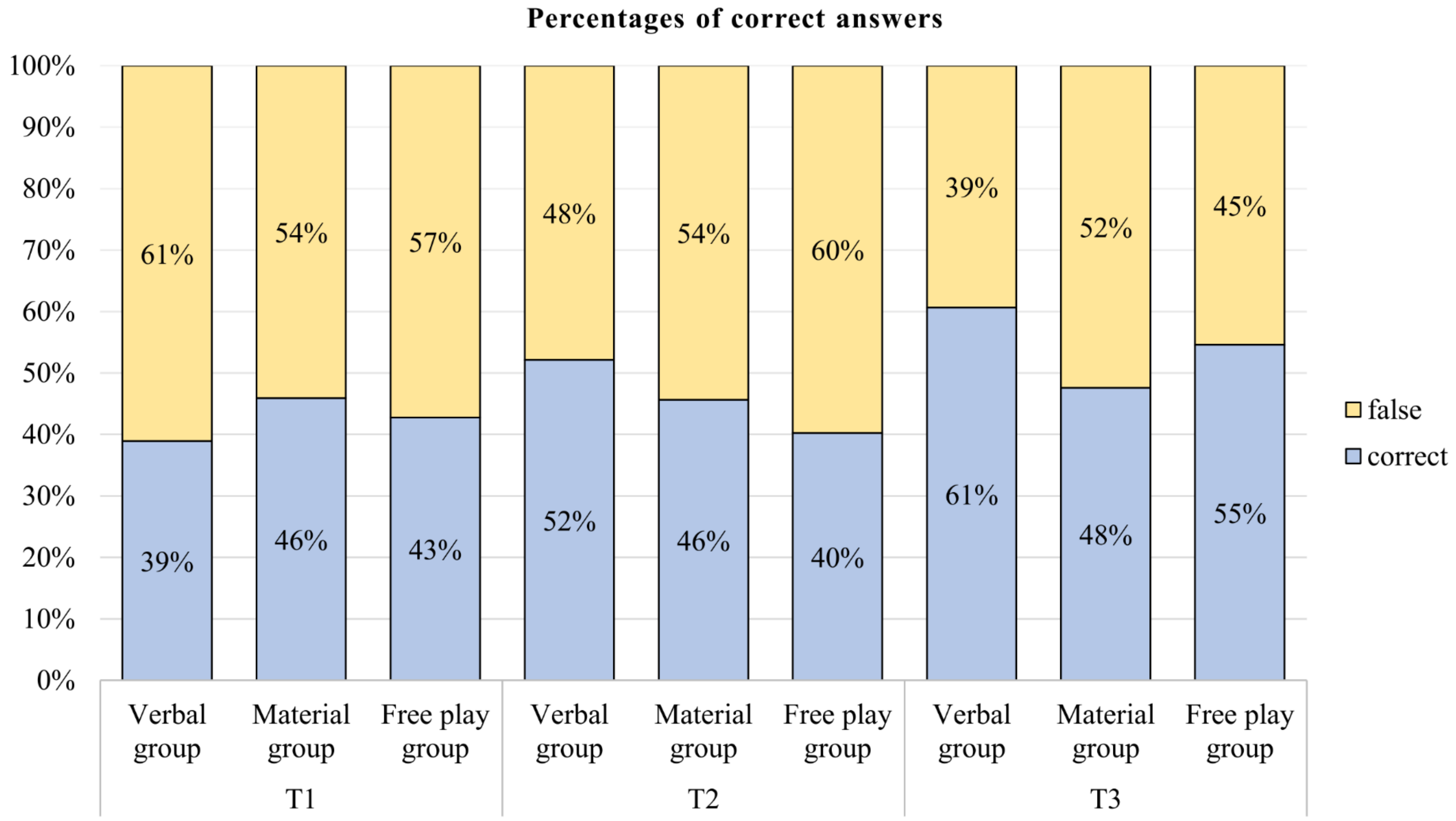










Appendix F: Supplementary Material 6. Percentages of correct answers on the reasoning test.



Appendix G: Supplementary Material 7. Results of the Kaplan-Meier analyses.

Table

Kaplan-Meier analysis for children's acquisition of Mass theory at T2

Survival function as estimated by the Kaplan-Meier analysis for children's acquisition of Mass theory at T2

	Verbal group						Material group						Free play group					
					95% CI						95% CI						95% CI	
time	n.risk	n.event	surv	*SE*	LL	UL	n.risk	n.event	surv	*SE*	LL	UL	n.risk	n.event	surv	*SE*	LL	UL
3	53	12	.77	.06	.67	.89	47	4	.91	.04	.84	1.00	51	3	.94	.03	.88	1.00

Survival function as estimated by the Kaplan-Meier analysis for children's acquisition of Mass theory at T2 and T3

	Verbal group						Material group						Free play group					
					95% CI						95% CI						95% CI	
time	n.risk	n.event	surv	*SE*	LL	UL	n.risk	n.event	surv	*SE*	LL	UL	n.risk	n.event	surv	*SE*	LL	UL
4	47	11	.77	.06	.65	.90	43	2	.95	.03	.89	1.00	47	3	.94	.04	.87	1.00
5	36	4	.68	.07	.56	.83	41	3	.88	.05	.79	.99	44	2	.89	.05	.81	.99
6	32	4	.60	.07	.47	.75	38	5	.77	.06	.65	.91	42	2	.85	.05	.76	.96

Notes. N.risk = Number of children/group. N.event = number of children explaining with Mass consistently. Surv = percentage of children who did not answer consistently. *SE* = standard error. LL = lower level 95% confidence interval. UL = upper level 95% confidence interval.

Table

Kaplan-Meier analysis for children's acquisition of Mass theory over T2 and T3

	Consistent																	
	Verbal group						Material group						Free play group					
					95% CI						95% CI						95% CI	
time	n.risk	n.event	surv	SE	LL	UL	n.risk	n.event	surv	SE	LL	UL	n.risk	n.event	surv	SE	LL	UL
4	16	2	.88	.08	.73	1.00	18	1	.94	.05	.84	1.00	12	1	.92	.08	.77	1.00
5	-	-	-	-	-	-	17	1	.89	.07	.76	1.00	-	-	-	-	-	-
6	-	-	-	-	-	-	16	2	.78	.10	.61	1.00	11	1	.83	.11	.65	1.00
	Inconsistent																	
	Verbal group						Material group						Free play group					
					95% CI						95% CI						95% CI	
time	n.risk	n.event	surv	SE	LL	UL	n.risk	n.event	surv	SE	LL	UL	n.risk	n.event	surv	SE	LL	UL
4	29	8	.72	.08	.58	.91	26	2	.92	.05	.83	1.00	34	2	.94	.04	.87	1.00
5	21	4	.59	.09	.43	.80	24	2	.85	.07	.72	1.00	32	2	.88	.06	.78	1.00
6	17	3	.48	.09	.33	.70	22	4	.69	.09	.54	.90	30	1	.85	.06	.74	.98

Notes. N.risk = Number of children/group. N.event = number of children explaining with Mass consistently. Surv = percentage of children who did not answer consistently. *SE* = standard error. LL = lower level 95% confidence interval. UL = upper level 95% confidence interval.

Appendix H: Items of the COM Test.


















Appendix I: Items of the Y-CSM

1. Einführung

Schau mal, ich habe hier 2 Puppen. Diese Puppe heißt Kiki/Bodo.

[Interviewer sucht identische Puppe heraus und sagt:]

Hier ist noch ein Kind, genau wie du und Kiki/Bodo. Er/sie ist ein/e Freund/in von Kiki/Bodo. Ihr/Sein Name ist Kora/Momo

[Interviewer setzt sich beide Fingerpuppen auf die Zeigefinger und Mittelfinger der linken Hand.]

Kiki/Bodo **[Finger mit Kiki/Bodo bewegen]** und Kora/Momo **[Finger mit Kora/Momo bewegen]** gehen in die gleiche Kita und haben den/die gleiche/n Erzieher/in, genau wie du. Sie werden über Dinge sprechen, die sie gerne tun und die sie gut können. Die beiden mögen verschiedene Sachen und können verschiedene Sachen unterschiedlich gut, aber das ist okay, weil sie verschiedene Kinder sind. Es ist ganz in Ordnung für verschiedene Kinder, andere Interessen zu haben und auch in anderen Dingen gut zu sein.

2. Übungsfragen

Kiki/Bodo kann gut malen. **[auf Kiki/Bodo deuten]**

Kora/Momo kann nicht so gut malen. **[auf Kora/Momo deuten]**

Was ist mit dir? Kannst du gut malen oder eher nicht so gut malen?

Zeig mir doch mal, wie gut du schon malen kannst.

Kannst du „Gar nicht gut“, „Nicht so gut“, „Eher gut“ oder „Sehr gut“ malen?

[auf dem Dreieck andeuten]

1	2	3	4

Kora/Momo kann gut Fußball spielen. **[auf Kora/Momo deuten]**

Kiki/Bodo kann nicht so gut Fußball spielen. **[auf Kiki/Bodo deuten]**

Was ist mit dir? Kannst du gut Fußball spielen oder eher nicht so gut Fußball spielen?

Zeig mir doch mal, wie gut du schon Fußball spielen kannst.

Kannst du „Gar nicht gut“, „Nicht so gut“, „Eher gut“ oder „Sehr gut“ Fußball spielen?

[auf dem Dreieck andeuten]

1	2	3	4

3. Fragen

Kiki/ Bodo und Kora/Momo werden jetzt darüber sprechen, wie viel sie über das Bauen mit Bauklötzen wissen und was sie gut können. Die beiden sind unterschiedlich gut in verschiedenen Bereichen und sie wissen unterschiedliche Dinge. Wie gerade eben, möchte ich, dass du mir sagst, wie das bei dir ist. Wenn du etwas nicht verstehst, kannst du mir das einfach sagen.

<table>
<tr><td>SK1_a</td><td>Kiki/Bodo [auf Kiki/Bodo deuten] weiß schon ganz viel über das Bauen mit Bauklötzen. Kora/Momo [auf Kora/Momo deuten] weiß noch nicht so viel über das Bauen mit Bauklötzen. Wie ist das bei dir? Weißt du schon viel über das Bauen mit Bauklötzen oder weißt du noch nicht so viel über das Bauen mit Bauklötzen?
<table><tr><td>Eher nicht so viel</td><td>Viel</td></tr></table></td></tr>
<tr><td>SK1_b</td><td>Zeig mir mal, wie viel du schon über das Bauen mit Bauklötzen weißt.
„Ganz wenig“, „eher nicht so viel“, „eher viel“, „sehr viel“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]
<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>SK2_a</td><td>Wenn die Erzieherin in der Kita oder ihre/seine Eltern zuhause etwas über das Bauen mit Bauklötzen erzählen, versteht Kora/Momo [auf Kora/Momo deuten] das schnell. Kiki/Bodo [auf Kiki/Bodo deuten] versteht das nicht so schnell. Wie ist das bei dir? Wenn deine Erzieherin/dein Erzieher in der Kita oder deine Eltern zuhause etwas über das Bauen mit Bauklötzen erzählen, verstehst du das dann schnell oder eher nicht so schnell?
<table><tr><td>Eher nicht so schnell</td><td>Schnell</td></tr></table></td></tr>
</table>

<table>
<tr><td>SK2_b</td><td>Zeig mir mal, wie schnell du das verstehst.
„Gar nicht schnell“, „nicht so schnell“, „eher schnell“, „sehr schnell“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]
<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>SK3_a</td><td>Fällt es dir leicht, etwas über das Bauen mit Bauklötzen zu lernen oder eher nicht so leicht?
<table><tr><td>Eher nicht so leicht</td><td>Leicht</td></tr></table></td></tr>
<tr><td>SK3_b</td><td>Zeig mir mal, wie leicht es dir fällt, etwas über das Bauen mit Bauklötzen zu lernen.
„Gar nicht leicht“, „nicht so leicht“, „eher leicht“, „sehr leicht“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]
<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>SK4_a</td><td>Bist du gut darin, etwas über das Bauen mit Bauklötzen zu lernen oder bist du eher nicht so gut, etwas über das Bauen mit Bauklötzen zu lernen?
<table><tr><td>Eher nicht so gut</td><td>Gut</td></tr></table></td></tr>
<tr><td>SK4_b</td><td>Zeig mir mal, wie gut du darin bist, etwas über das Bauen mit Bauklötzen zu lernen.
„Gar nicht gut“, „nicht so gut“, „eher gut“, „sehr gut“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]
<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>SK5_a</td><td>Ist es für dich einfach, etwas über das Bauen mit Bauklötzen zu verstehen oder eher nicht so einfach?
<table><tr><td>Eher nicht so einfach</td><td>Einfach</td></tr></table></td></tr>
</table>

<table>
<tr><td>SK5_b</td><td>Zeig mir mal, wie einfach es für dich ist, etwas über das Bauen mit Bauklötzen zu verstehen.
„Gar nicht einfach“, „nicht so einfach“, „eher einfach“, „sehr einfach“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>SK6_a</td><td>Kennst du dich mit dem Bauen mit Bauklötzen gut aus oder eher nicht so gut?<table><tr><td>Eher nicht so gut</td><td>Gut</td></tr></table></td></tr>
<tr><td>SK6_b</td><td>Zeig mir mal, wie gut du dich mit dem Bauen mit Bauklötzen auskennst.
„Gar nicht gut“, „nicht so gut“, „eher gut“, „sehr gut“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>FR1_a</td><td>Kiki/Bodo [auf Kiki/Bodo deuten] macht es Spaß mit ihren/seinen Eltern oder der Erzieherin/dem Erzieher über das Bauen mit Bauklötzen zu sprechen, Kora/Momo [auf Kora/Momo deuten] macht es eher keinen Spaß. Wie ist das bei dir? Macht es dir eher Spaß oder eher keinen Spaß mit deinen Eltern zuhause oder mit deiner Erzieherin/deinem Erzieher über das Bauen mit Bauklötzen zu sprechen?<table><tr><td>Eher keinen Spaß</td><td>Spaß</td></tr></table></td></tr>
<tr><td>FR1_b</td><td>Zeig mir mal, wie viel Spaß es dir macht, über das Bauen mit Bauklötzen zu sprechen.
„Gar keinen Spaß“, „eher keinen Spaß“, „ein bisschen Spaß“, „viel Spaß“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>FR2_a</td><td>Würdest du gern mehr über das Bauen mit Bauklötzen lernen oder eher nicht so gern?<table><tr><td>Eher nicht so gern</td><td>Gern</td></tr></table></td></tr>
</table>

<table>
<tr><td>FR2_b</td><td>Zeig mir mal, wie gern du noch mehr über das Bauen mit Bauklötzen lernen möchtest.
„Gar nicht gern“, „eher nicht so gern“, „ein bisschen gern“, „sehr gern“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>FR3_a</td><td>Wenn die Erzieherin den Kindern etwas über das Bauen mit Bauklötzen erklärt, macht es Kora/Momo [auf Kora/Momo deuten] Spaß etwas über das Bauen mit Bauklötzen zu erfahren. Kiki/Bodo [auf Kiki/Bodo deuten] macht es eher keinen Spaß. Wie ist das bei dir? Macht es dir eher Spaß etwas über das Bauen mit Bauklötzen zu erfahren, oder eher keinen Spaß?<table><tr><td>Eher keinen Spaß</td><td>Spaß</td></tr></table></td></tr>
<tr><td>FR3_b</td><td>Zeig mir mal, wie viel Spaß es dir macht, über das Bauen mit Bauklötzen zu erfahren.
„Gar keinen Spaß“, „eher keinen Spaß“, „ein bisschen Spaß“, „viel Spaß“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>FR4_a</td><td>Würdest du gern von deiner Erzieherin/deinem Erzieher oder deinen Eltern zuhause etwas über das Bauen mit Bauklötzen erzählt bekommen oder eher nicht so gern?<table><tr><td>Eher nicht so gern</td><td>Gern</td></tr></table></td></tr>
<tr><td>FR4_b</td><td>Zeig mir mal, wie gern du etwas über das Bauen mit Bauklötzen erzählt bekommst.
„Gar nicht gern“, „eher nicht so gern“, „ein bisschen gern“, „sehr gern“ [auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]<table><tr><td>1</td><td>2</td><td>3</td><td>4</td></tr></table></td></tr>
<tr><td>FR5_a</td><td>Macht es dir Spaß, wenn du etwas Neues über das Bauen mit Bauklötzen erfahren kannst, oder eher keinen Spaß?<table><tr><td>Eher keinen Spaß</td><td>Spaß</td></tr></table></td></tr>
</table>

FR5_b	Zeig mir mal, wie viel Spaß es dir macht, etwas Neues über das Bauen mit Bauklötzen zu erfahren. „Gar keinen Spaß“, „eher keinen Spaß“, „ein bisschen Spaß“, „viel Spaß“ **[auf dem Dreieck andeuten. Die Zahlen nie vorlesen!]**

1	2	3	4

Appendix J: Curriculum Vitae

ANKE M. **WEBER**

Gerberstraße 5, 76829 Landau, Germany • weber-a@uni.landau.de • +49 152 57220636

CAREER

Research associate
University of Koblenz-Landau | Landau, Germany
Institute for Children and Youth Education, Primary School Education

Since 2019-02
Project DiLeSaM (The future of STEM learning: Digital learning environments for science and mathematics), funded by the Telekom-Stiftung. Development of scaffolding and learning materials, investigation of primary school teacher students' expectancy (e.g., self-concept) and values (e.g., motivation) for science teaching

Since 2018-10
Conception, planning and realization of science seminars for primary school teacher students; supervising master theses

2016-10/2018-09
Project BauSpiel – Playing with building blocks, funded by the German Research Association. Development of test instruments, planning and conducting of experiments on children's motivation, self-concept and intuitive science theories; implementing different forms of play with and without scaffolds; data analyses; supervising the student researchers

Student Researcher
2011-11/2012-07
University of Koblenz-Landau | Landau, Germany
Institute for Developmental and Educational Psychology
Data processing and analyses; archive work; advising B.Sc. Psychology students; supporting the teaching staff

EDUCATION

Doctoral studies
Since 2016-10
University of Koblenz-Landau | Landau, Germany
As a psychologist in educational sciences

M.Sc. Psychology
2013-10/2016-09
University of Koblenz-Landau | Landau, Germany
Graduation grade: 1.2 (excellent)
Title of master thesis: "The development of the academic self-concept: Stability, the influence of gender and parents' appraisals", grade: 1.0 (excellent)

B.Sc. Psychology
2010-10/2014-02
University of Koblenz-Landau | Landau, Germany
Graduation grade: 1.4 (very good)
Title of bachelor thesis: "The influence of achievement and mothers' achievement perceptions on children's academic self-concept and the stability of the academic self-concept over time", grade: 1.0 (excellent)

Abitur
2010-06
Leer, Germany
Graduation grade: 1.8 (good)

TEACHING EXPERIENCE

Lecturer Since 2018-10	**University of Koblenz-Landau** | Landau, Germany Seven seminars on scientific projects for M.Ed. students: Topics include programming and magnetism; basic teaching techniques such as scaffolding, diagnosis, and feedback
Thesis advisor Since 2018-10	**University of Koblenz-Landau** | Landau, Germany 15 master theses of M.Ed. students: Topics include effects of scaffolding on motivation in science learning and children’s science self-concepts, and impact of play on children’s learning and self-concepts

FUNDING & AWARDS

2020-07	**Teaching award** Hochschuldidaktische Arbeitsstelle of the University of Koblenz-Landau, Germany
2020	**German Research Foundation** Assistance in study designing and writing a research proposal for the PFKiNaT project (Learning support by educational specialists in early scientific and technical education during construction play), 283 940 €
2017/2018	**University of Koblenz-Landau** | Landau, Germany Mentee in the mentoring programme ment² for PhD students
2017	**ERASMUS+** Course on scientific English, Malta, 900 €

KNOWLEDGE & SKILLS

Languages: fluent German (native language), English; basic Spanish, French

Technical skills: office suites; statistics software R, SPSS, MPlus

Statistical procedures: cross-sectional and longitudinal data; multilevel growth curve modelling, structural equation models, survival analyses, Cox regressions, path analyses, latent class analyses, cross-lagged models, Markov chains, item response modelling, latent state-trait modelling

Landau, 2021-01-04

Anke Weber

Anke M. Weber, M.Sc.

Appendix K: Publication List

Publications

Peer-reviewed journals

Weber, A. M., Reuter, T., & Leuchter, M. (2020). The impact of a construction play on 5- to 6-year-old children's reasoning about stability. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2020.01737 (Open Access)

Weber, A. M., & Leuchter, M. (2020). Measuring preschool children's knowledge of the principle of static equilibrium in the context of building blocks: Validation of a test instrument. *British Journal of Educational Psychology, 90*, 50–74. https://doi.org/10.1111/bjep.12304

Weber, A. M., & Trojan, J. (2018). The restorative value of the urban environment: A systematic review of the literature. *Environmental Health Insights*, *12*, 1–13. https://doi.org/10.1177/1178630218812805 (Open Access)

Journal articles submitted or in preparation

Reuter, T., Weber, A. M., Flottmann, J., & Leuchter, M. (2020). Ability of 6- to 7-year-old children to choose the control of variables strategy. *Manuscript submitted for publication*.

Sasse, H., Reuter, T., Weber, A. M., & Leuchter, M. (2020). The influence of adaptive scaffolding during guided discovery on science learning performance of primary school children. *Manuscript in preparation*.

Weber, A. M., Barkela, V., Stiel-Dämmer, S., & Leuchter, M. (2020). Informatische Problemlösekompetenz und logisches Schlussfolgern als Voraussetzungen für emotionale Kosten bei Grundschullehramtsstudierenden [Computer science problem solving and logical thinking as prerequisites for primary school student teachers' emotional costs]. *Manuscript submitted for publication*.

Weber, A. M., Bastian, M., Barkela, V., Mühling, A., & Leuchter, M. (2020). Effects of motivation and self-concept on student teachers' informatic problem solving. *Manuscript in preparation*.

Weber, A. M., & Leuchter, M. (2020). Change in 5- to 6-year old children's science knowledge about stabilities and self-concepts promoted by construction play. *Manuscript in preparation*.

Peer-reviewed book chapters

Weber, A. M., & Leuchter, M. (2018). Frühe naturwissenschaftliche Bildung und Förderung [Early science education and support]. In T. Schmidt & W. Smidt (Eds.), *Handbuch empirische Forschung in der Pädagogik der frühen Kindheit* [*Handbook of empirical research on education in early childhood*] (pp. 333–347). Münster: Waxmann.

Conference contributions (selection)

Dämmer, S., Weber, A. M., & Leuchter, M. (2019, March). Vergleiche als lernförderliche Maßnahme im Lehramtsstudium [Comparisons as a measure promoting learning in teacher training]. Paper presented at the annual meeting of the Gesellschaft für Didaktik des Sachunterrichts, Lüneburg, DE.

Weber, A. M., & Leuchter, M. (2017, September). Erfassung und spielerische Förderung des Statikverständnisses von Vorschulkindern [Measurement and playful support of preschool children's statics knowledge]. In F. Lang, & J. Kant, Entwicklung und Förderung des Wissenschaftsverständnisses vom Kindergarten bis zum Studium [Supporting scientific understanding from kindergarten to the university level]. Paper presented at the biannual conference of the Specialist Group for Educational and Developmental Psychology of the German Psychological Association, Münster.

Weber, A. M., & Leuchter, M. (2017, September). Balancing blocks: Young children's understanding of statics. Paper presented at the biannual meeting of the European Association for Research on Learning and Instruction, Tampere, FI.

Weber, A. M., & Leuchter, M. (2017, August). Measuring preschoolers' statics understanding. Paper presented at the biannual meeting of the Junior Researchers of the European Association for Research on Learning and Instruction, Tampere, FI.

Weber, A. M., & Leuchter, M. (2018, August). Guided-play, children's motivation, and statics knowledge. In T. Reuter, *Guided play in early STEM education – adults' process competencies and children's outcomes*. Paper presented at the biannual meeting of the SIG5 conference of the European Association for Research on Learing and Instruction, Berlin, DE.

Weber, A. M., & Leuchter, M. (2019, August). The effect of construction play on 5- to 6-year-old children's knowledge in the balance domain and their self-concepts. In T. Reuter, *New perspectives on playful learning in early STEM education*. Paper presented at the biannual meeting of the European Association for Research on Learning and Instruction, Aachen, DE.

Zucker, V., Reuter, T., Weber, A. M., Dämmer, S., & Leuchter, M. (2019, February). Subjektive und objektive Evaluation eines videobasierten Seminars für Grundschullehramtsstudierende [Subjective and objective evaluation of a video-based seminar for primary school teacher training students]. In P. Hörter, & M. Holodynski, *Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden in den MINT-Fächern der Primarstufe: Ergebnisse von videobasierten Interventionsstudien in der ersten Phase der Lehrerbildung* [*Professional teaching perception of student teachers in STEM at the primary school level: Results of video-based intervention studies in the first phase of teacher education*]. Paper presented at the annual meeting of the Society for Empirical Educational Research, Cologne, DE.

Appendix L: Declaration of originality

I hereby declare that I have written the present dissertation independently, without assistance from external parties and without the use of other resources than those indicated. This dissertation has not been submitted previously for grading at this university or any other academic institution.

Authorship and publication status

The articles presented in the present dissertation were submitted for publication or have been published in international peer-reviewed journals. In collaboration with the co-authors, the articles were primarily written by Anke Maria Weber, doctoral candidate at the Institute for Children and Youth Education, University of Koblenz-Landau. The authors and publication statuses of the articles are entitled in the following.

Article 1	Weber, A. M., & Leuchter, M. (2020). Measuring preschool children's knowledge of the principle of static equilibrium in the context of building blocks: Validation of a test instrument. *The British Journal of Educational Psychology*, *90*, 50–74. https://doi.org/10.1111/bjep.12304
Article 2	Weber, A. M., Reuter, T., & Leuchter, M. (2020). The impact of a construction play on 5- to 6-year-old children's reasoning about balance relationships. *Frontiers in Psychology*, *11*, 1737. https://doi.org/10.3389/fpsyg.2020.01737
Article 3	Weber, A. M., & Leuchter, M. (2021). *Construction play promotes change in 5- to 6-year old children's science knowledge about stabilities and science self-concept*. Manuscript submitted for publication.