“How does it work?”

van der Steen, Stephanie

Publication date:
2014

Citation for published version (APA):
van der Steen, S. (2014). “How does it work?”: A longitudinal microgenetic study on the development of young children’s understanding of scientific concepts [S.l.]: s.n.

“How Does It Work?”

A Longitudinal Microgenetic Study On The Development Of Young Children’s Understanding Of Scientific Concepts

Cover design: Pablo ter Borg
Printed by: Impskamp drukkers

ISBN: 978-90-367-6958-7
ISBN e-versie: 978-90-367-6957-0
NUR-code: 773

The research described in this dissertation was funded by the Curious Minds ('TalentenKracht'—Platform Bètatechniek) program, supported by the Dutch Ministry of Education, Culture and Science, and by the Heymans Institute (University of Groningen).

rijksuniversiteit
groningen

“How Does It Work?”

A Longitudinal Microgenetic Study On The Development Of Young Children’s Understanding Of Scientific Concepts

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 8 mei 2014 om 11.00 uur

door

Stephanie van der Steen

geboren op 18 mei 1986
te Leiderdorp

Promotor

Prof dr. P.L.C. van Geert

Copromotor

Dr. H.W. Steenbeek

Beoordelingscommissie

Prof dr. A.E.M.G. Minnaert
Prof. dr. M.E.J. Raijmakers
Prof. dr. K.W. Fischer

Contents

Chapter 1: Introduction

In this chapter, a broad overview of this dissertation is given. The research questions and motives are stated, and the organization of the chapters in this dissertation is briefly outlined.

What do we need to further promote the sustainable provision of energy? Can we find a cure for life-threatening diseases? Which techniques can help us to put global warming on hold? Is there extraterrestrial life in outer space? Answering these questions requires knowledge of science, technology, engineering and mathematics (STEM fields). Those fields are considered to be of crucial importance to meet societies' most pressing current and future challenges (National Research Council, 2011). Many international organizations have therefore given considerable attention to students' skills needed in STEM and related fields (Langdon, McKittrick, Beede, Khan, & Doms, 2011; Kuenzi, 2008; National Research Council, 2010; OECD, 2004). In addition, students' knowledge of mathematics and science is frequently assessed in large-scale studies, such as the OECD Program for International Student Assessment (PISA; OECD, 2006), and the IEA Trends in International Mathematics and Science Study (TIMMS; Martin, Mullis, & Foy, 2008). These studies' reports have raised the issue that the STEM knowledge of the current student population is insufficient to guarantee future technological advancement, and warn that the number of students choosing a science career is declining (National Research Council, 2011; Roberts, 2002; Van Langen & Dekkers, 2005).

The low number and interest of students in STEM fields is particularly surprising given young children's interest in scientific phenomena and technology. At roughly age 3, children ask their caregivers all sorts of scientific questions, such as: How come the moon changes shape? Why are the dinosaurs extinct? How does a car work, and why do you need gas to drive? Unfortunately, somewhere along their journey to adulthood, the number of these questions decreases and the interest in scientific phenomena declines (Van Geert & Steenbeek, 2007; see also Simonton, 1999 for a general account of the emergence and decline of talent). Physics, mathematics, and chemistry in secondary school seem too abstract and not visibly connected to real life and the challenges society is currently facing. Eventually, a scientific career does not appeal to the majority of college students, and the number of future scientists graduating is low. Are children unable to further develop their STEM skills and interests, despite their early enthusiasm for scientific phenomena?

1.1 Research questions

This dissertation focuses on the longitudinal development of young children's STEM skills in interaction with their material and social environment. Our main research question was: How do children's (3-5 years old) STEM skills develop over the course of 1.5 years in interaction with the social and material context, and are special needs students equally able to acquire these skills? To be more specific, we focused mostly on children's conceptual STEM skills, that is, their understanding of the scientific concepts gravity and air pressure embedded in practical tasks, and how these develop over time in interaction with the tasks and the researcher guiding the child through them, by using an inquiry-based approach. This means that students were actively engaged in the investigation of questions, hypothesizing, gathering evidence, and explain findings (Gibson & Chase, 2002; National Research Council, 2000). The ultimate goal was to provide more information on how children—from both regular schools and special educational facilities—learn in the fields of science and technology. In combination with other studies, this dissertation can eventually help to construct effective science lessons for young children, which can possibly stimulate the STEM knowledge and careers of the future student population. Indeed, there is some evidence that inquiry-based science activities are an effective way of teaching science (Hodson, 1999; Van Schijndel, Singer, Van Der Maas, & Raijmakers, 2010), and have long-term positive effects on students' science achievement and understanding (Gibson & Chase, 2002).

1.2 Broad overview of this dissertation

To examine the development of children in depth, this dissertation adopts a process approach. This entails that we closely look at children's real-time construction of understanding scientific concepts, taking into account the child-context dynamics. To achieve this, we used a microgenetic method to code children's understanding, and their interactions with the task and researcher. In addition, we included children with special needs (i.e., with externalizing and internalizing problems) in this study. Numerous studies have shown that these children score significantly lower on standardized tests (Lane, Barton-Arwood, Nelson, & Wehby, 2008; Reid, Gonzalez, Nordness, Trout, & Epstein, 2004). The question was how these children would develop their understanding of scientific concepts during our tasks, and if their delays would also be present when using a process-oriented and inquiry-oriented approach to their scientific knowledge and skills. Lastly, given the cyclical causal relationship between the short- and long-term timespan of learning (Steenbeek & Van Geert, 2013), we saw an additional necessity to couple several microgenetic codings of the interactions, to get an idea of the mechanisms on the long-term time scale of development.

1.3 Organization of chapters

This dissertation is organized as follows. Chapter 2 is focused on the set up of the longitudinal microgenetic study we conducted to examine the development of young children's understanding of scientific concepts over time. The theoretical and practical foundations of this study are discussed, and we added an extensive description of the participants, materials, data collection, and coding procedures. This chapter is aimed to serve as an overview, which can be used as a reference when reading the next chapters.

In chapter 3, a new theoretical model of children's understanding of scientific concepts is proposed, based on a number of complex dynamic systems properties and skill theory (Fischer, 1980) principles. This model can give guidance to both research and practice in science education. More specifically, it helps to understand how children construct their knowledge in concordance with the (social) context, and highlights the importance of the real-time person-context dynamics. Throughout this chapter, the model is illustrated with an empirical example of the development of a child's understanding during an air pressure task.

Chapter 4 is focused on a cross-sectional comparison of regular and special needs students in terms of their understanding of the scientific concepts gravity and air pressure during one visit. In what way does special needs students' understanding of the scientific concepts differ from their peers in regular schools? In this comparative study, we compared the mean understanding level, number of correct and incorrect answers, as well as the distribution of understanding levels for the two groups. Differences were examined for the whole group as well as for separate age groups.

In chapter 5, we describe a case study in which we explored the couplings between the short- and long-term time scales of development. We focused on three interactions between a 4-year old boy and a researcher while working on an air pressure task. Using microgenetic codings of the complexity of the boy's reasoning and the researcher's questions, we show how fluctuations in the boy's understanding complexity are organized, how the child-researcher interaction dynamics shape this learning process, and how these dynamics change over time.

In chapter 6, we compare the relative importance of general (e.g., standardized test scores, gender, and age) and interaction measures (e.g., number of follow-up questions, off-task behavior) to characterize the development of scientific understanding over the course of 5 visits (comprising 1.5 years). Using a cluster analysis, we first explored how many distinct developmental pathways in understanding we could find, and described their shape. Subsequently, we used a decision tree analysis to investigate which variables (demographic, questionnaire data, test scores, and the microgenetic codings of the interactions) could best predict these distinct developmental pathways.

Lastly, chapter 7 provides a summary of the thesis and a general discussion, covering the practical and theoretical consequences of this study's process approach and its outcomes. We also discuss the performance of special needs students in this study, and how our results can potentially influence (special) educational policy and practice. In addition, we illustrate our ideas on how to improve the current standardized tests used to measure children's academic performance. Lastly, we cover how this study's setting has given us insight into how learning in STEM areas occurs, and how this setting can be translated to the educational practice.

Chapter 2: How 32 Children Worked on Air Squirts and Marble Tracks: Background, Research Design, Participants and Methods

This methodological chapter focuses on the set-up of the longitudinal microgenetic study on the development of young children's understanding of scientific concepts over time (May 2009 – May 2012). This chapter covers the theoretical and practical foundations of this study, as well as the participants, materials, and coding of video data. The purpose of this chapter is to provide an overview of our research aims and methodology. In this way, the chapter can be used as a reference when reading other parts of this dissertation.

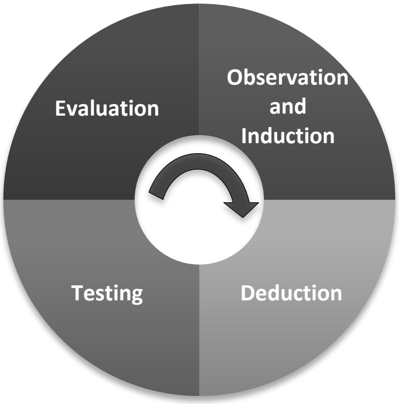
2.1 Scientific concepts and children's understanding of these

Skills in the fields of Science, Technology, Engineering and Mathematics (STEM) can be defined in two broad categories. The first category comprises the conceptual aspects, that is, domain-specific scientific concepts (Zimmerman, 2000). Scientific concepts can be defined as ideas about phenomena in STEM domains, such as chemistry, physics, and biology (Baartman & Gravemeier, 2011; OECD, 2003). In this dissertation, we refer to "understanding of scientific concepts" as the student's current understanding of a particular scientific concept, which has a specific level of complexity. In the last decades, children's understanding of various scientific concepts has been studied, such as gravity (Novak, 2005; Palmer, 2001; Sharp & Sharp, 2007), air pressure (Séré, 1986; She, 2002; Tytler, 1998), electricity (Chiu & Lin, 2005; Shipstone, 1984; Zacharia, 2007), energy (Papadouris, Constantinou, & Kyratsi, 2008; Trumper, 1993), chemistry (Garnett, Garnett, & Hackling, 1995; Taber, 2001), gear wheels (Dixon & Bangert, 2002; Lehrer & Schauble, 1998), the universe (Albanese, Neves, & Vicentini, 1997; Dunlop, 2000), and many more (see for example Rohaan & Van Keulen, 2011).

The second category of STEM skills comprises the domain-general procedural skills needed to acquire the scientific concepts (Zimmerman, 2000). These skills can be roughly attributed to various parts of the empirical cycle (De Groot, 1969; see Figure 1). For *induction*, these skills are observing, asking questions, hypothesizing, and designing experiments; for *deduction*, these are using materials, observing, measuring, predicting, and recording; for *testing*, these skills are (statistical) calculations and interpreting data, and for *evaluation* these are confirming or rejecting evidence, and making inferences (Zimmerman, 2000). Lastly, to succeed as a scientist, several other skills are needed, such as adaptability, communication and social skills, and self-regulatory skills (Bybee, 2010).

This dissertation focuses mostly on the development of children's understanding of scientific concepts. The nature of these conceptual STEM skills is currently under discussion (see Van Geert, 2011a for an overview). Two contrasting theoretical views exist in the scientific literature: a *representationalist* and a *dynamic embedded* view. From a *representationalist* view, scientific concepts are a collection of internally stored symbolic structures representing scientific facts or concepts, which are processed by an individual (Posner, Strike, Hewson, & Gertzog, 1982; Vosniadou & Brewer, 1987). A child's understanding of a particular scientific concept thus consists of a collection of these internal structures, representing scientific facts and ideas, which can be activated and used to coordinate his/her behavior toward the current environment (Haselager, De Groot, & Van Rappard, 2003).[1] Development of understanding scientific concepts over time is seen as a process of conceptual change (Posner et al., 1982).[2] That is, children's initial concepts, based on their interaction with the world, are inaccurate reflections of the scientific reality (a famous example is the pre-operational child's inaccurate understanding of conservation; Piaget, 1947/2001). Through teaching, children can restructure their initial concepts and transform these into more accurate versions over a longer period of time (Vosniadou, 1994; 2007).

Figure 1: Empirical cycle



---

[1] An important notion within the representationalist framework is that of "mental models". These are special kinds of mental representations that constrain the knowledge acquisition process in ways that are similar to the individual's current beliefs, or to specific theories a person holds (Vosniadou, 1994).

[2] Allen and Bickhard (2013) call this "foundationalism": Knowledge is constructed from a representational base.

In contrast, from a *dynamic embedded* view (scientific) concepts are constructed in real-time, and develop over multiple interactions (Greeno, 1989; Thelen and Smith, 1994; Van Gelder, 1998; Zednik, 2011). That is, concepts are no internal structures, but emerge from a current (real-time) process of construction. This process consists of interactions among many components of both the child (e.g., motor skills, sensory systems, and memory), and the context (e.g., the characteristics of the material, the contents and nature of the questions asked by a teacher). Hence, from a *representationalist* framework, a child's answer to a question about a scientific concept reflects his or her internal representation of that concept, while from a *dynamic embedded* view, the representation is in fact the child's answer, which is a locally and temporally emergent structure constructed in a specific context, and not an internal reflection of a concept (Van Geert, 2011a; Van Geert & Steenbeek, 2013).

One implication of the *dynamic embedded* view is that children's concepts are softly assembled and can never be completely context-independent. In fact, the context contributes to the construction of the concept. As a result, concepts vary from occasion to occasion, since the context in which the child constructs them changes. However, concepts can only vary within certain boundaries, given that some of the child's characteristics, such as working memory or motor skills, are roughly stable, or at least slowly changing over time. In other words, concepts are history-dependent, in the sense that they depend on the child's earlier experiences and learning. On the long term, after repeated interactions in several contexts, children's construction potential and usable range of contextual opportunities will change and develop (Van Geert, 2011b).

This context-dependence has a consequence: According to the *dynamic embedded* view, it would be impossible to assess the child's ability to reason about scientific concepts independently and across all contexts (Van Geert, 2002; Van Gelder, 1998). Whereas representationalists are concerned with context-independent assessment of children's scientific performance (which researchers have tried to accomplish with standardized paper-and-pencil tests), dynamic theorists argue that context-independence is a myth, even in such tests. We should therefore not try to find a situation that enables us to extract the "real" context-independent reasoning ability of children, but instead try to evaluate children's skills within contexts that are representative or characteristic for the application of these skills (Van Geert, 2002). Hence, it is a legitimate question to ask what a child can accomplish in an educational (classroom) context, guided by a teacher who asks questions, interprets the child's (re)actions, and provides additional material or social support when needed. This setting formed the basis of the *Curious Minds* research project.

2.2 The Curious Minds: Children's STEM skills and talents

Curious Minds (In Dutch: *TalentenKracht*) is an international research project in which Dutch and Belgian research groups work together to study young children's talent for science and technology (www.talentenkracht.nl). Combining studies from the fields of educational science and pedagogics, as well as developmental and neuropsychology, the aim is to help teachers and parents to recognize and foster these talents. Although it is generally known that young children's reasoning skills are more advanced than assumed in times of Piaget and Vygotsky (e.g., Gelman & Baillargeon, 1983), researchers do not know much about *how* young children's science and technology skills develop on the short term (e.g., during a task) and on the longer time scale of development. Do we give children enough opportunities to develop their skills and talents in STEM areas, and how can we support them?

2.2.1 The nature of talent

Much like the two theoretical views on scientific concepts, two broad views on talent exist in the scientific literature, that is, the *genetic endowment* view (e.g., Gagné, 1985; Gardner, 1993), and the *dynamic emergent* view (e.g., Simonton, 1999; 2001; Van Geert, 2011a). The difference between these two views is not a simple nature/nurture distinction. Both approaches assume that talent is the result of multiple components, such as a relatively high level of performance in a specific domain, high intrinsic motivation and extended effort[3]. However, they differ greatly in what they consider the origin of talent. The *genetic endowment view* emphasizes the existence of a specific innate component (i.e., giftedness) that forms the foundation of a person's high level of performance. Effort and motivation can help a person to thrive and become better skilled in a particular domain, but there can be no talent, that is, no exceptional high performance in a specific domain, without a specific genetic aptitude. Or, to use the famous words of John Dryden (1693/1885, p.60): "Genius must be born and never can be taught". Hence, talent is a roughly static characteristic of only a small number of people that have a specific genetic component, which manifests itself at an early age, and can be further developed by practice.

In contrast, from a *dynamic emergent* view, talent is a property that emerges and changes across the lifespan. Talent is emergenic, meaning that interactions between several physical, physiological, and cognitive properties of the child result in an accumulative advantage (Simonton, 2001; 2005). Moreover, talent may emerge at different points in development for different persons. This is what Simonton (2001) calls the epigenetic component of talent: The underlying personal properties have a different maturation rate, and there are vast individual differences in the configuration of those underlying components. In music for example, two underlying properties of talent might be pitch perception and sense of rhythm (which of course have their underlying neurological components). These two properties do not develop at the same time.[4] Hence, the child with a perfect pitch and a reasonable sense of rhythm will develop his/her musical talent in a different way and rate than a child with a reasonable pitch perception and a perfect sense of rhythm. Moreover, if the child with the perfect pitch perception has frequent ear infections at a young age, it might take a while

---

[3] For a theory that mainly emphasizes deliberate practice, see Ericsson, Krampe & Tesch-Römer, 1993.

[4] Pitch perception, depending on the definition, develops roughly at 3 or 4 years, although tones can be discriminated from early infancy on (Trainor & Unrau, 2012). Rhythmic ability roughly develops between 4 and 7 years of age, also depending on the definition (Pollatou, Karadimou, & Gerodimos, 2005).

for this trait to fully develop, and no early indication of this talent may be detectable.

This example highlights the influence of chance or random factors shaping talent, which brings us to the dynamic aspect of the *dynamic emergent* view. In addition to those random factors (e.g., the ear infections in the example above), the dynamic interaction with the environment also greatly influences the development of talent. When there is an early indication of talent, that is, when young children do relatively well in a specific domain, they are likely to attract support from their environment (e.g., their parents and teachers) to further develop their talent. In addition, the fact that they do so well might stimulate them to put in more time and effort to acquire more knowledge and skills within that domain. This results in an accumulated advantage (or a preferential attachment process—see Yule, 1925 for a first account), making the relative difference with the child's peers bigger over time (Van Geert, 2011a; see also Gladwell, 2009). Hence, children with an above-average quality of innate characteristics (whatever these characteristics may be) may actually reach an exceptional level of performance in a certain area, when they have a high level of intrinsic motivation and receive a high quality of support from the environment. These repeated (iterative) interactions between child and environment characteristics may account for more inter-individual variance in performance than the emergenic epigenetic mechanisms (Simonton, 2005). To summarize, the *dynamic emergent* view of talent entails a process that is emergenic (interaction of several child properties, not just a single genetic component), epigenetic (a different onset for these properties, and inter-individual differences in the property configuration), and dynamic (the role of chance; iterative child-context interactions).

A researcher's theoretical view has important implications for the study and stimulation of talent. Taking the *genetic endowment* view, only those children performing excellent in a specific domain at a young age (which is indicative of a specific genetic component) can further develop their talent with help from a stimulating environment. After all, children without the required genetic component can only benefit to some extent, but will never be capable of true excellent performance. From a *dynamic emergent* view however, all young children would benefit from a stimulating environment, which, in interaction with the child's characteristics, can cause an upward spiral. According to this view, no children should be left out, given that the emergenic and epigenetic processes make it hard to predict when a child actually reaches high levels of performance and thus, when talent becomes observable. Talent, it seems, comes in many forms, and develops in many ways. Besides, even if a child is not necessarily capable of truly outstanding achievements, adequate teaching and stimulation would still be beneficial to assist the child in developing his or her own optimal level of performance.

2.2.2 Curious Minds and its view on talent

Although the *Curious Minds* project never explicitly mentioned the *dynamic emergent* view until 2011, it has been an underlying basis since its start in 2006. In an article about the project's aim and scope, Steenbeek and Uittenbogaard (2009) mention that *Curious Minds* intends to investigate and stimulate children's "natural talents" for STEM areas. To be more specific, those "natural talents" are characteristics that all children have to some extent, and that are considered crucial for the development of advanced STEM skills, such as curiosity, problem-solving, and an intrinsic motivation to learn. In this way, *Curious Minds* also adopts a prospective approach, by studying children at a very young age at which they have yet to develop an exceptionally high level of reasoning or performance, making it possible to study talents as they develop over time. This is in contrast with a retrospective approach, by which researchers try to reconstruct the developmental process that has led to a particular excellent performance. This distinction is hence a matter of forwards versus backwards.[5]

The prospective approach is clearly visible in one of the first studies of the Curious Minds program, in which researchers from the University of Utrecht

---

[5] Of course, both approaches have their advantages and drawbacks. While the retrospective approach does not capture the developmental processes and the dynamic interaction with the environment as talent emerges, adopting the prospective approach means that not all participants necessarily develop an excellent performance in a specific domain.

interviewed young children while working on various scientific tasks (e.g., De Lange, Feijs, & Uittenbogaard, 2007). In a setup similar to classical Piagetian tasks (1947/2001), they asked the children (3-5 years old) to classify objects and to take other people's perspective, but also to conduct simple experiments. The video recordings of the interviews show how children reason about a variety of STEM topics, sometimes in a creative or rather advanced way. This has stimulated other researchers to participate in the *Curious Minds* program and conduct systematic studies on the development of children's STEM skills and talents, both on the short term (e.g., Meindertsma, Van Dijk, Steenbeek, & Van Geert, 2012), as well as on the longer term (this dissertation).

In one of the first *Curious Minds* project proposals (Van Geert & Steenbeek, 2007), an extensive definition of talent in STEM areas is discussed, which highlights both the dynamic emergent as well as the prospective nature: "Talent is a child's capacity to (ultimately) reach a high level of performance in a specific domain. Characteristics are: a high learning potential; the ability to elicit high-quality support from the (social) environment; in-depth processing of domain-specific information; creativity; belief in one's own competence; enthusiasm, and a strong intrinsic motivation to learn" (p. 4). Hence, talent for STEM fields is a rather extensive construct, comprising both child- and context-related aspects, as well as conceptual and procedural STEM skills. To design a manageable study based on this definition, this dissertation focuses on the following aspect of STEM talent: children's capacity and potential to reach a high level of performance on STEM tasks. To be more specific, we focus mostly on children's conceptual knowledge, that is, their understanding of scientific concepts embedded in scientific tasks, and how this understanding develops over time in interaction with the tasks, and the researcher administrating these. The main research question is therefore: how does children's (3-5 years old) understanding of scientific concepts develop over the course of 1.5 years in interaction with the context (the scientific tasks, the questions and the adult who monitors the child's explorations and explanations)?[6]

---

[6] The study included 3 years of data collection (10 visits), but due to the extensive process of data coding, only the first 1.5 years (5 visits) are subject to this dissertation.

2.2.3 Curious Minds and the inclusion of special needs students

Given that the *Curious Minds* project stresses the natural talents of children, and adopts a *dynamic emergent view* on talent, our study also includes children with special needs (i.e., with externalizing and internalizing problems) of the same age group[7]. Numerous studies have shown that these children score significantly lower on standardized academic tests (Lane et al., 2008; Reid et al., 2004). This, however, does not imply that these children are less curious, creative, or enthusiastic about physical phenomena. The question was how these children would perform on the scientific tasks, and if they would benefit from guidance provided by the researcher during the tasks.

The fact that we recruited young children in special educational settings had two consequences. First, these children had at least moderate to severe behavioral and/or psychological problems at a young age, which required extra care and prevented them from enrollment in regular educational settings (mild problems are more easily overlooked, and easier to cope with in regular educational facilities). In this population, most problems fall in the category of moderate to severe Attention Deficit/Hyperactivity disorders, or Autism Spectrum disorders. Second, the availability of girls in these special settings was limited. Boys with externalizing problems are more likely to get referred to special education, because their behavior is considered more disruptive than that of girls with externalizing problems. In addition, girls with internalizing problems are more likely to be labeled as “just shy”, and more often stay in regular educational settings.

2.3 A process approach and the importance of microgenetic studies

Earlier studies on children’s understanding of scientific concepts predominantly focused on specific outcomes of learning processes, such as scores on knowledge tests (e.g., before and after an intervention), the number and categories of (mis-) conceptions, as well as the coherence and accuracy of

---

[7] In Dutch: kinderen in het cluster 4 onderwijs.

children's concepts. This has given us important information about children's understanding of scientific concepts and global developmental trends across cohorts (cf. Granott, Fischer, & Parziale, 2002). However, it does not give us information about the process of understanding, that is, it does not show in which way children's scientific concepts develop over time.

In addition, most of those earlier studies are conducted from a *representationalist* perspective. That is, they are focused on the concepts that children have, or on differences between their initial concepts and more accurate versions after (e.g.,) an intervention. Although contextual influences are usually acknowledged, the context has not been treated as a continuously intertwining factor in the development of scientific concepts (see Richardson, Marsh, & Schmidt, 2010 for a discussion about this overlook). While these earlier studies may have been beneficial for revealing differences between groups of students, evaluating interventions, or studying cross-sectional developmental trends, they did not answer the developmental question: *How* does development (or learning) emerge in individual children, in interaction with the context? (as opposed to the question: How does it emerge in terms of aggregate measures, such as age averages).

Microgenetic studies—i.e. studies of (learning) processes that unfold during a short time span—can answer this important developmental question. By means of frequent observations during short time periods, these studies provide important insights into how learning occurs in interaction with the material and social context (Granott & Parziale, 2002). For instance, microgenetic studies have focused on children's changing understanding in interaction with the material context while (e.g.,) operating a robot (Granott, 2002), building miniature bridges (Parziale, 2002), working on a computer program for statistical analysis (Yan & Fischer, 2002), solving balance scale problems (Philips & Tolmie, 2007), working on number conservation tasks (Siegler, 1995), and understanding the concept of living organisms (Opfer & Siegler, 2004). Other microgenetic studies have focused on the interaction with the social environment during learning. For example, a teacher's support aimed at a level that is somewhat higher than that of the student (Granott, 2005) increases students' performance over time. Furthermore, frequent mismatches between the responses of the child and the teacher during a learning interaction, or too many self-iterations of the teacher lead to negative academic outcomes in the long run (Steenbeek, Jansen, & Van Geert, 2012).

Microgenetic studies are not omnipresent though. Given that microgenetic studies put an emphasis on short-term learning interactions, they require detailed coding of data, usually aided by video recordings. This level of detail has its price: both the data collection as well as the data processing are very time-consuming, especially when trying to employ an in-depth analysis capturing both the child's changing understanding, as well as the ongoing interaction with the context. Microgenetic studies are therefore usually solely focused on learning processes during a short time span (e.g., a lesson), and repeated series of microgenetic studies of the same child are usually not coupled to obtain a picture of long-term development. However, given the cyclical causal relationship between the short- and long-term timespan of learning (Steenbeek & Van Geert, 2013), we see a necessity to couple several of these repeated microgenetic processes to get a grip on mechanisms on the long-term time scale of development. This dissertation attempts to make this connection (see chapter 5 and 6). In the next section of this chapter, the setup of our longitudinal process study, including the participants, material and microgenetic coding of video data are discussed.

2.4 Study design and methods

2.4.1 Participants and recruitment

The *Curious Minds* project is specifically targeted at preschoolers and kindergarten students, which roughly comprises the ages 3 to 5. To obtain the most representative sample of this age range, we decided to recruit children from all three age groups (age 3, 4, and 5), including students from regular and special schools. After approval from the ethical committee of the psychology department of the University of Groningen, we started to recruit participants three months before the start of the study, in a predominantly rural region in the north of the Netherlands. We visited one regular primary school and one regular daycare center[8] and handed out information packages to the parents. These information packages included an informed consent form, an intake questionnaire and a return envelope. At the daycare center, four parents returned the forms, yielding three 3-year olds, and one 4-year old soon going to kindergarten. At the primary school, nine parents returned the forms, yielding four 4-year olds, and five 5-year olds. One of these parents asked if her 3-year old daughter and a 3-year old classmate, enrolled in a preschool[9] could participate as well. Furthermore, upon hearing about the project, the parents of three-year old fraternal twins indicated they were interested. After that, we stopped recruiting in regular educational settings, given that we had reached our target of 5 children per age group.

With regard to the special needs students, we started recruiting children of the same age in a special needs primary preparatory school[10], and a special needs daycare center in a predominantly rural region in the east of the Netherlands. Three months before the beginning of the study, the daycare center's head psychologist handed out the information packages to parents during intake interviews. He only selected parents of children with a reasonable vocabulary and social skills, excluding children with severely disrupting behaviors.[11] We then started to recruit in the special needs primary preparatory school. Eleven children were pre-selected by the head of the school and the researcher (adopting the same selection criteria), and received an information package including a letter from the head of the school. All informed consent forms were returned. Since special needs children stay longer in their specialized daycare center before going to kindergarten, this group was on average 4.5 months older than the regular education group. All children, however, still fell within the range of 4- and 5-year

---

[8] Kindergarten is an integral part of Dutch primary schools, comprising the first two grades (4-6 years of age). Before kindergarten, children often go to daycare while their parents are at work. We use the term daycare center to refer to the settings for the 3-year olds, and kindergarten to refer to the settings for the older students.

[9] *Peuterspeelzaal*, a preschool for children (2-4 years old). Children go here for only a few hours per week to get used to a school setting.

[10] The purpose of this school is to educate young children with special needs, and assess their capacities and possibilities for their further school career. After this, children continue their education elsewhere.

[11] By adopting these selection criteria, we knew we would be able to interact with the children without encountering severe communication problems. Note, however, that these children still had considerable behavioral and/or psychological problems.

olds, apart from one boy who just turned 6 years old (74 months). For a list of participants and their characteristics, see Table 1.

In order to make the situation equal for the children, they were visited at their schools. When children switched schools, the director of the new school was contacted by telephone to ask if we could visit the child at his/her new school. Over the course of 3 years, most children switched schools, due to the fact that all 3-year olds transferred to kindergarten, and due to the fact that the special needs primary school had a preparatory purpose, after which children transferred to a variety of schools. Most transfers occurred between the fourth and the fifth visit. The study had only one drop-out after the first visit. A child of a similar age was recruited to participate in the study from the second session onward (see Table 1).

2.4.2 Tasks

Within STEM fields, many scientific concepts exist (see for example Rohaan & Van Keulen, 2011). It was therefore crucial to narrow down and select only a small number. To capture the interaction with the context, we also specifically wanted to study children's understanding while working on scientific tasks. Hence, the concepts needed to be embedded in practical (hands-on) tasks that were somewhat adaptable to the cognitive level of young children, and required some exploration. The tasks needed to be brief (maximum duration between 15 and 20 minutes), appealing to children, suitable for indoor use, transportable by car, and safe. To prevent simple testing effects (Allen, Mahler, & Estes, 1969), we also needed an opportunity to construct tasks of increasing complexity, or to highlight different aspects of the concept in subsequent tasks. Finally, some scientific phenomena are more explicitly present in children's daily life (shadows, density—floating or sinking objects) than others (atoms, gasses). To provide the most accurate picture of children's understanding of scientific concepts, we wanted to include both. After consulting with researchers from the University of Utrecht, who conducted the first *Curious Minds* study (De Lange et al., 2007), we chose two of their tasks as a prototype and basis for the development of the other tasks. For the first sequence we used the *open marble track* as a prototype, comprising Newtonian concepts such as gravity, inertia and acceleration. The latter two concepts imply some sort of movement, which is why we frequently used moving objects in this task sequence. This moving object was usually a ball (varying in size and weight), moving over a surface that got increasingly more complex over time.

The second task sequence was based on the *air squirt* as a prototype task. It comprised the scientific concepts air flow/pressure, and Boyle's law ($P \times V = C$).[12] Since Boyle's law underlies many (pneumatic) pump systems, we frequently used pumps, sometimes in a simple form (squeezing a balloon), and sometimes in a more complex form (a ball or water pump). For the sake of simplicity, we refer to the task sequences as the *air pressure* and *gravity sequence* throughout this dissertation. Of these two concepts, children experience gravity-related tasks more often in daily life, for example in ball games, or while playing with marble tracks.

The tasks (including the prototype tasks, which were slightly adjusted) were constructed in collaboration with an expert in the field of physics and engineering, using materials from toy and hardware stores. For a list of the tasks, including pictures, descriptions, and a table with the increasing complexity of the sequences, see appendix A.

---

[12] The children never explicitly worked with Boyle's law in the form of a formula, but it does describe the underlying forces embedded in most of the air pressure tasks.

Table 1: List of participants, including gender, age, and educational setting

Group	M/F	Age in months	Additional information
Regular education 3-year olds	M	38	Preschool
	F	38	Preschool, sister of 5-year old in regular kindergarten
	F	35	Daycare
	M	40	Daycare
	M	35	Daycare
	M	44	Daycare, fraternal twin
	F	44	Daycare, fraternal twin
Regular education 4-year olds	M	49	Kindergarten
	F	56	Kindergarten
	F	53	Kindergarten
	F	49	Kindergarten
	M	51	Kindergarten
Regular education 5-year olds	M	62	Kindergarten
	M	62	Kindergarten, brother of 3-year old in regular preschool
	F	63	Kindergarten
	M	61	Kindergarten
	M	63	Kindergarten
Special education 3-year olds	M	44	Special daycare
	M	36	Special daycare, brother of 4-year old in special school
	M	43	Special daycare, identical twin
	M	43	Special daycare, identical twin
Special education 4-year olds	M	59	Special kindergarten
	F	57	Special kindergarten, sister of 3-year old in special daycare
	M	55	Special kindergarten
	M	50	Special kindergarten
Special education 5-year (and older)	M	66	Special kindergarten
	M	71	Special kindergarten
	M	71	Special kindergarten
	M	68	Special kindergarten
	F	61	Special kindergarten
	M	74	Special kindergarten
	M	62	Special kindergarten, included from second session

2.4.3 Task administration

The task administration was set up to simulate a natural teaching-learning environment as much as possible. To get a grip on children's thinking about the tasks, it was important to not only look at their actions, but to also get them to verbalize their ideas, so we could get more information on how the children understood the tasks. Moreover, it was important to provide some structure, to prevent children from focusing on only a few aspects of the tasks, while ignoring others. To enable children to show their understanding and explore the task in a natural way while still maintaining an acceptable degree of standardization, the preferred choice was an adaptive protocol. This protocol guaranteed that all children were asked the basic questions reflecting the core building blocks of the task and the incorporated scientific concepts. At the same time, the protocol left enough space for children to take initiative and show their understanding spontaneously, and for the researcher to provide scaffolding when needed.

The following points were decisive for our choice of an adaptive protocol: A standardized protocol with fixed task-related questions might hinder children's own exploration process. Moreover, given that we conducted the study with children aged between 3 and 5, a protocol with standardized questions might either be too hard for the youngest children, or too easy for the eldest. Such a fixed protocol would also not allow the use of a variety of scaffolding techniques, such as encouragement or follow-up questions (unless a fixed number of these scaffolding techniques and their timing were determined in advance). This might cause problems for children who need a little more questioning or encouragement to come to a full understanding of the task. On the other hand, a protocol that is too loose might lead to a lack of structure.

Similar to the empirical cycle (see Figure 1), the protocol started by asking children to describe (a specific aspect of) the material. Subsequently, children were asked to predict what would happen if the task would be manipulated in a specific way. Then the task was manipulated (usually by the child), and the researcher asked the child to describe what he/she just observed. Finally, children were asked to explain their observations, that is, the mechanism that was revealed by manipulating the task. Then the cycle started again, focusing on another aspect of the task. Hence, although the study was mostly focused on children's understanding of scientific concepts embedded in the tasks, several procedural skills were also assessed in the process, such as hypothesizing, using materials, and interpreting.

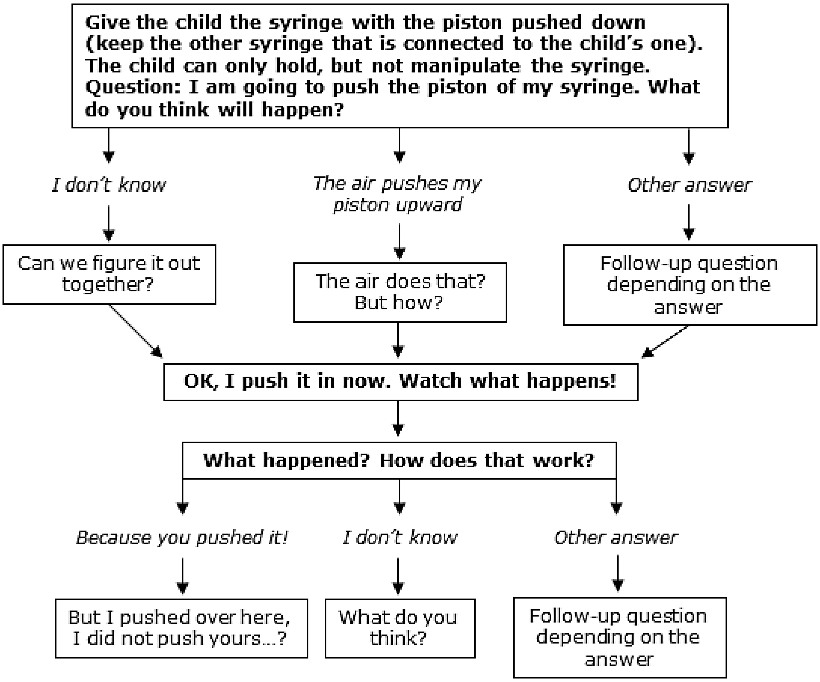
The protocols were written before the round of visits started. Each protocol was written in the same format (see Figure 2; a selection of protocols—in Dutch—can be found in appendix B). The main questions (in bold) were asked to all children, in a predetermined order. These were usually the questions asking the child to describe, predict, or explain task aspects. Anything that happened in between those main questions depended on the child's reaction to the question. To make it easier for the researcher to respond to the child, the protocol contained a few of the child's answer possibilities after each main question, as well as possible reactions to these. However, these served as mere examples, and the researcher was allowed to ask multiple follow-up questions, and use various scaffolding techniques. Besides follow-up questions, the scaffolding consisted of encouraging the child to think about the task and to try out his/her ideas using the material, giving compliments, trying to direct the child's attention, and clarifying/summarizing the child's findings or previous answers.

The researcher was allowed to keep asking follow-up questions until she had an accurate picture of the child's understanding, and was ready to get to the next main question. The child's wrong answers were challenged in the same way, by repeated follow-up questions until the child concluded that his/her line of thinking was incorrect, or until the researcher felt the child would not change his/her mind. Although children's answers were challenged sometimes, the feedback never included statements indicating whether the child was right or wrong. "Don't know"-answers were usually treated by encouraging the child to say what he or she thought, or by emphasizing that the task was no test.

2.4.4 Pilot and start of the study

In February 2009, the setting, and the first two tasks and protocols were tested in a small pilot study with four 4-year old children in a small village in the north of the Netherlands. Two children having Dutch as a second language were included, to test whether the tasks and questions were understandable for children with a smaller vocabulary, as what might be the case for children in special educational settings. Small adjustments to the tasks and protocols were made after this pilot study. For example, materials were placed on the ground, instead of on the table, so children could reach for them, and the protocol was slightly shortened to make sure children were optimally concentrated.

Figure 2: Excerpt of a task protocol (translated in English)



The longitudinal study started in May 2009, and ended in May 2012. Children were visited 10 times over the course of these 3 academic years: In October/November; January/February, and May/June. Because of the time-consuming nature of the data collection and processing, only 5 visits are subject to this dissertation, that is, the videos from May 2009 until October 2010. The data of the remaining 5 visits is yet to be analyzed.

2.4.5 Skill theory: a suitable method for coding understanding in video data

To obtain a measurement of children's understanding during the task administration, we developed a coding system. Given that we specifically view understanding as an ongoing process distributed across child and context, our goal was to measure this process in real time while the children worked on the tasks, incorporating both child and context characteristics. In addition, the coding should yield trajectories of changes in understanding during the tasks (i.e., on the short term), that could eventually be coupled to understand longer-term process characteristics. An additional difficulty was that our measurement of understanding should be comparable for both the gravity and air pressure tasks. For example, the highest understanding level on a gravity task should be comparable to the highest level of the air pressure task.

We chose skill theory as the basis of our coding system. This cognitive developmental theory focuses on the complexity and variability of children's skills— a variety of actions and thinking abilities—and the way these are constructed in specific domains (Fischer, 1980; Fischer & Bidell, 2006). The reason for choosing this theory was threefold. First, the theory assumes that skills are mastered in a specific context, and hold both person-related as well as context-related characteristics (Parziale & Fischer, 1998). An example of a skill is a child's ability to understand how air pressure works while manipulating a task (the context). This understanding is reconstructed when the student works on a similar task in another environment, for example with different materials or without the help of the researcher. People vary constantly between their highest and lowest possible complexity levels in a certain domain. The highest levels within this bandwidth are only reachable when the environment provides sufficient support (Fischer & Bidell, 2006; see also Yan & Fischer, 2002). Skills, such as children's understanding of a particular scientific concept, are thus highly influenced by the possibilities and constraints of the situation in which the skill is used. This view is highly similar to the *dynamic embedded view* of representations, which claims that concepts are no internal structures, but emerge from a current (real-time) process of construction.

Second, skill theory includes a hierarchical scale to measure the complexity of skills over a longer period of time, but also on a short period of time. The scale consists of 10 levels, grouped into 3 tiers. The first tier consists of sensorimotor skills: connections of perceptions to actions or utterances. The second tier consists of representational skills, these are understandings that go beyond current simple perception-action couplings, but are still based on them. The third tier consists of abstractions, which are general nonconcrete rules that also apply to other situations (Schwartz & Fischer, 2004). Within each tier, three levels can be distinguished, each one more complex than the previous one. The first one can be characterized as a single set, (i.e. a single action, a single representation, or a single abstraction). The second level is a relation between two of these sets, which is referred to as a mapping. The third level is a system of sets, which is a relation between two mappings, in which each mapping consists of a relation between single sets. After this level, a new tier starts, which is divided in single sets, mappings and systems as well (Fischer & Bidell, 2006).[13] For an indication of how we operationalized these levels in the current study, see below.

The least complex skills (the first level of the first tier) are single sensorimotor sets, comprising a single action, or a (nonverbal) understanding of a single observable aspect of a problem. This skill roughly emerges after 3 or 4 months (Fischer & Bidell, 2006). The skills highest in complexity are abstract systems, that people may develop from their mid-twenties on, when they are capable of

---

[13] The original formulation of skill theory (1980) also included a tier with reflexes (encompassing 3 levels within the tier). In later versions of the theory, a level called "single principles" is proposed that some highly skilled people may develop in a certain domain. We did not include these levels in our coding system.

comprehending encompassing abstractions in a specific field, similar in level to a postgraduate student's knowledge of his/her particular field of study. The crux, however, is that skill theory can not only describe and explain the development of skills on the long term, but also describe the microgenesis of problem solving (Fischer, 1980; Fischer & Bidell, 2006; Granott & Parziale, 2002; Schwartz & Fischer, 2004; Yan & Fischer, 2002). When facing a new task, even skilled adults may show skill levels that are mostly sensorimotor at the beginning, building up to more elaborate levels. During tasks, people do not go through the skill cycles in a linear fashion. Instead, they may repeatedly build up skill levels and regress before they obtain their highest possible level (Yan & Fischer, 2002).

A third reason to apply skill theory to this longitudinal microgenetic study is that the scale focuses on the hierarchical complexity of skills rather than their content. Because of this content-independent nature, skill theory enables researchers to compare skills (including understandings) across multiple time points, contexts, persons, and age ranges (Parziale & Fischer, 1998). Skill theory is therefore especially suitable to compare individual pathways across tasks (Fischer, Rose & Rose, 2007).

2.4.6 The construction of a coding system

The coding system to obtain measures of children's understanding of the scientific concepts during the tasks consisted of 4 phases. We started by assigning a time stamp to the beginning and end of each utterance of both the child and the researcher, using the program Mediacoder (Bos & Steenbeek, 2006). In the second phase, we changed the codings that marked the beginning of an utterance into categories. The researcher's utterances were coded as descriptive, predictive, and explanatory questions; encouragement; follow-up questions; compliments; clarifications; procedural remarks; directing the student's focus, and off-task utterances. The student's utterances were classified into descriptive, predictive, and explanatory answers/remarks; initiatives; content-related questions, and off-task utterances. In the third phase, we combined children's coherent descriptive, predictive, and explanatory answers into meaningful units, to be able to assess their level of understanding in the next phase of coding. The unit ended when the next utterance of the student fell into another category, or when the researcher interrupted the student, for example by asking another question. Hence, the units exclusively contained a series of descriptive, or predictive, or explanatory answers. One exception was made: If the researcher interrupted by simply encouraging the student to tell more about the same topic, the unit would not end.

In the final phase of the coding system, the complexity of the answers within a unit was determined using a scale based on skill theory. The complexity levels of the units ranged from single sensorimotor actions (Level 1) to single abstractions (Level 7). At Level 1 (sensorimotor actions), the child mentioned single characteristics of the task, such as “This tube is long”. At Level 2 (sensorimotor mappings), two elements of the task were coupled, such as “I can push this (piston) into here (the tube of the syringe)”. At Level 3 (sensorimotor systems), simple causal mechanisms were stated, such as “If I push this (piston) in, the other one goes upward”. At Level 4 (single representations), two causal mechanisms were coupled, or an “invisible” causal mechanism was mentioned, such as “When I push this (piston) in, air causes the other one to move upward”. At Level 5 (representational mappings), mechanisms were explained or predicted in terms of two causal relationships including an additional step, e.g., “The piston pushes the air down, which goes through the tube to the other syringe, which piston then gets pushed out by the air”. At Level 6 (representational systems), the system under question (e.g., the mechanism behind the task) was described in terms of all relevant elements and couplings between these. Finally, at Level 7 (single abstractions), the child’s answer should contain an abstraction, that is, the answer should contain an accurate immaterial concept (such as gravity, friction, inertia) that can be used in general, and thus goes beyond the task material. When a child simply answered “yes” or “no” to a close-ended question, the answer was simply rated as correct or incorrect. More extensive incorrect, irrelevant, and “don’t know”-answers were rated as incorrect, and were not assigned a level of complexity. The child’s other utterances, such as requests and off-task utterances, and the researcher's utterances were also not rated using the complexity scale.[14]

We explicitly want to emphasize that we used skill theory as a *basis* that we tailored to our needs. Throughout this dissertation, the coding is based on skill theory, but does not encapsulate all aspects of the original theory. For example, for the sensorimotor tier, we coded answers that were (at least partly) verbalized and not just purely nonverbal actions. In this way, our coding system resembles the new applications of the skill theory scale (e.g., Rappolt-Schlichtmann, Tenenbaum, Koepke, & Fischer, 2007) more than the older ones (e.g., Fischer, 1980; Schwartz & Fischer, 2004). To give an example: the child's verbalization of an observation (e.g., "you can push the piston into the syringe") is in our coding system treated as a manipulation involving two objects, and thus coded as a sensorimotor mapping. The fact that the child verbalizes this relationship (and not just manipulates the material) does, according to our coding system, not mean that this reflects a higher level.

Another difference is that our coding system is explicitly focused on correct task-related utterances. Incorrect answers or remarks are only labeled as 'false', even though these could technically be assessed in terms of their complexity. In our study, however, there was usually no complex reasoning behind the false answers, and we felt that we could get an accurate picture of children's understanding by focusing on their correct answers or remarks.

Lastly, we also made some coding rules that were more or less in accordance with the theory, but above all made the coding easier and the inter-rater reliability higher. For example, correct predictions, even if they were simple, were always coded as level 4 (single representations) or higher, because in order to predict an event that has not happened yet, one has to go beyond the task material and reason about a hypothetical outcome. In addition, correct explanations, even if they were simple, were always coded as a level 3

---

[14] For one of the studies in this dissertation, we also coded the researcher's descriptive, predictive, explanatory, and follow-up questions using a complexity scale based on skill theory.

(sensorimotor system) or higher, because a correct explanation needed to include at least 3 elements: cause, effect, and a relationship between these two.

2.4.7 Interrater reliability

In order to make sure that the codings of different raters were reliable, a standardized codebook was used. For each round of coding (categories, units, and understanding levels), 10 students went through a training by coding 3 video fragments of 15 minutes. The codings of the third fragment were compared to the codings of the researcher who constructed the codebook and percentages of agreement were calculated. On average, these were: categories: 83% (range 80-93; $p < .01$), units: 87% (range 80-100; $p < .01$), and level of understanding: 84% (range 78-92; $p < .01$).[15] In total, 160 videos were coded for this dissertation.

2.4.8 Questionnaires

Although the main focus of the study was on the video data codings, the parents of the children were also asked to fill out questionnaires after each visit (10 times in total). Questions focused on home environment characteristics that may influence STEM skills, such as: 1) parents' perception of the child's problem solving skills, curiosity, and exploratory behavior; 2) the child's play behavior at home, the use of educational toys, cooperative play with parents and sports, and 3) parental stimulation in the form of household chores, stimulation of early arithmetic skills such as counting and recognizing numbers, and stimulation of playing with construction toys. The first questionnaire also contained questions about demographics, such as the child's age, gender and diagnosis (if applicable), family composition, nationality and the parents' educational level.

Depending on the parents' preference, the questionnaires were either sent by e-mail, or given to the children to pass to their parents after the visit. The questionnaire was filled out by the same parent each time, who was not informed about the child's performance on the tasks. If the questionnaire was not returned

---

[15] *P-values* are calculated using Monte Carlo permutation tests, see below for an explanation.

within two weeks, the parents received two reminders via e-mail. Given that the questionnaires were not the main focus of the study, we stopped reminding the parents after the second e-mail. On average, 24.2 questionnaires (78%) were returned after each visit.

The purpose of these questionnaires for this study was twofold. First, they were used to get a general idea of the occurrence of major life events that could affect children’s performance on the tasks. Before each visit, we made sure to look at the parents’ answers to the final question of the previous questionnaire, which was: “Have any major events occurred in your child’s life during the last 3 months? Major events include e.g., moving to another house or town, the death of a family member or pet, a long-term illness in your child’s family, changing schools or classrooms, getting a new classroom teacher, etcetera”. If the children would perform considerably worse during the tasks compared to the previous visits, and recently went through a major life experience, we could possibly link this performance decline to the life event. During the study, some major life events indeed occurred in the children’s lives, but they did not seem to have a profound negative affect on their work on the hands-on tasks.

Second, the questionnaires were used in chapter 6 of this dissertation, to see if children’s development over time—more specifically the developmental trajectory of the cluster they were assigned to by means of a cluster analysis—could be predicted by “home environment variables”, which were derived from the questionnaires of visit 1 through 5. Variables included e.g., children’s language, emotional, physical and motor development rated by the parents as below average/average/above average; children’s preference for playing with educational toys as rated by their parents; the average number of educational toys used during cooperative parent/child play as reported by the parents, etcetera. For a comprehensive list of these “home environment variables” and their predictive value, see chapter 6.

2.4.9 Standardized learning achievement (Cito) scores

Like the questionnaires, the children’s standardized learning achievement test (Cito) scores were not the main focus of the study, but provided important information on children’s school performance. Cito tests are standardized assessments of learning achievement, administered 2 times a year to keep track of children’s progress on the subjects math and (Dutch) language.[16] Given that children in kindergarten have limited spelling and number skills, the early math and language tests administered in kindergarten mostly focus on mathematical and language *reasoning* (Wijnstra, Ouwens, & Béguin, 2003). This means that they address the ability to phrase words, understand questions (Cito, 2009), classify objects, and to measure and observe differences and similarities (Koerhuis, 2011).

After asking the parents for permission, we collected the early math and language test scores, provided by the (remedial) teachers of the children around the time of the third visit. On both tests, children could get a score from A (25% highest-scoring students) to E (10% lowest-scoring students). We obtained both test scores, because we considered them equally important for the performance on the tasks. The math test measures early analytical skills, whereas the language test gives information about the child’s ability to understand questions. The test scores of 4 of the special needs students were missing, because the ongoing standardized assessment is not yet obligatory for special schools.[17] The test scores were used in chapter 4, to examine whether there was a difference between the test scores of the regular and special students, and in chapter 6 to determine their predictive value for the three cluster’s developmental trajectories over the course of 5 visits.

---

[16] In Dutch: leerlingvolgsysteem. Since august 2012, schools are obliged to test children with standardized tests and keep track of their progress. Schools can choose from several variants of these standardized tests, but the Cito learning achievement tests are used by most schools of the children in this study.

[17] This will be obligatory from August 2014 on.

2.4.10 Data analysis

Depending on the research question(s) of the chapter, we used a variety of graphical techniques to display patterns in the data, such as (normalized) Loess curves and frequency distributions. In addition, we used Monte Carlo permutation tests to determine the statistical significance of differences throughout most chapters, and a hierarchical cluster analysis and decision tree analysis for chapter 6. More information about the data analysis can be found in the next chapters. Here we would like to cover one technique that is widely used in this dissertation, namely Monte Carlo permutation tests.

Monte Carlo permutation tests can be used when the assumptions underlying conventional statistical techniques cannot be met, for example when one has a small or skewed data set (Todman & Dugard, 2001). This highly flexible method can be used to answer a variety of questions about developmental processes. In a simple “construct-your-own-test” kind of way, it can compare test statistics that cannot always be used in conventional statistical techniques, for example distributional characteristics, slopes of graphs, overall trends in the data, but also the more common proportions and group averages (Van Geert, Steenbeek, & Kunnen, 2012).

A Monte Carlo test (also known as random sampling or permutation technique) determines the chance that a test statistic is accidental, that is, caused by chance alone. This chance can be determined by drawing a large number of “accidental” samples from the original data, by means of either resampling or random permutation (shuffling) of the empirical data. The difference between random permutation and resampling is that the first technique draws random cases from the original distribution without replacing them, whereas the second technique does replace the original cases, considering the sample as an infinite pool to draw from. After repeatedly shuffling or resampling the data (1000, 5000, or even 10.000 times), the number of instances that the empirically observed test statistic occurred in these random samples is counted. Dividing this count by the total number of drawn samples results in a *p*-value, comprising the chance that the test statistic can be found in these random samples. If the *p*-value is low, the chances are low that our observed statistic is based on chance, or in other words, only based on the properties of our sample (see Van Geert, Steenbeek, & Kunnen, 2012 for a tutorial on using Monte Carlo tests).

2.5 Summary of this chapter

This dissertation focuses on the longitudinal development of young children's STEM skills in interaction with their material and social environment. Throughout this dissertation, we focus mostly on children's conceptual STEM skills, that is, their understanding of the scientific concepts gravity and air pressure embedded in practical tasks, and how these develop over time in interaction with the tasks and the researcher guiding the child through them. The nature of these conceptual STEM skills is currently under discussion. We have mentioned the *representationalist* and the *dynamic embedded* view, and how the latter view underlies the current study. We proceeded by describing that this study is part of the Curious Minds program, which concentrates on young children's natural talents for science and technology. Much like the two theoretical views on scientific concepts, two broad views on talent exist in the scientific literature, the *genetic endowment* view, and the *dynamic emergent* view. The current study has adopted the latter view, by taking a process-oriented, prospective approach to the study of young children's skills in the domain of science and technology. This entails that we 1) focus on young children's understanding of scientific concepts as these develop both on the short-term during tasks, as well as on the long-term; 2) take a microgenetic approach by coding children's (verbalized) understanding of scientific concepts in real time; 3) take the person-context dynamics into account by not only coding children's understanding, but also the researcher's utterances and linking these to one another; 4) include a special needs student population, to see if their delays would also be present when using a process-oriented and inquiry-oriented approach to their scientific knowledge and skills, and 5) couple several short-term microgenetic codings of the interactions to provide a picture of the longer-term development of understanding scientific concepts.

Chapter 3: Using the Dynamics of a Person-Context System to Describe Children's Understanding of Air Pressure[18]

This chapter explains how children's understanding can be studied from a dynamic systems complexity approach and skill theory perspective, and illustrates this with an example of understanding an air pressure task. Using dynamic systems principles, we can take the dynamics of children's understanding into account, without reducing its complexity or the role of the environment. We argue that understanding is a continuous person-environment loop, which emerges through iteration (every understanding is based on the previous one and embedded in the current context). Using skill theory, a framework for cognitive complexity, we can describe understanding in terms of complexness, ranging from basic perception-action connections to abstractions, and detect microgenetical variability in understanding. While developing, children repeatedly (re)construct their understandings. The long-term development of understanding therefore constitutes of an aggregation of multiple short-term interactions in different contexts, which also govern the iterative sequence of short-term interactions. The proposed framework enables us to closely monitor interactions between child and environment to determine how understanding is formed. Given that understanding is a process of intertwining person-context dynamics, it is important for parents and educators to be aware of the ways in which they interact with their children or pupils.

---

[18] This chapter is published as: Van Der Steen, S., Steenbeek, H., Van Geert, P. (2012). Using the Dynamics of a Person-Context System to Describe Children's Understanding of Air Pressure. In H. Kloos, B. J. Morris, & J. L. Amaral (Eds.), Current Topics in Children's Learning and Cognition (pp. 21-44).

Understanding refers to “the ability to understand”, which means “to comprehend, to apprehend the meaning or import of, or to grasp the idea of [something]” (Oxford English Dictionary, 1989). Understanding is a key concept within all fields of study concerning learning and development, such as cognitive psychology, pedagogy, educational sciences, and developmental psychology. Within these fields of study, understanding has been studied for different domains, such as scientific reasoning (e.g., Grotzer, 2004; Inhelder & Piaget, 1958/2001; Rappolt-Schlichtmann, Tenenbaum, Koepke, & Fischer, 2007), social development (e.g., Blijd-Hogeweys, 2008), mathematics (e.g., Dehaene, 1997; Gilmore & Bryant, 2008), and many more. In the field of education, children’s understanding is especially important, as understanding involves deep knowledge of concepts, and the active manipulation of this knowledge in the form of explaining, predicting, applying, and generalizing (Perkins & Blythe, 1994).

A model of understanding can give guidance to both researchers and educators dealing with children’s understanding and the development of their understanding. In this chapter, we will present such a model, based on dynamic Systems and skill theory principles. The model is illustrated throughout this chapter with examples of children’s understanding of scientific concepts, or more specifically, children’s understanding of air flow and air pressure during a syringe task, which is described below. The syringes task is designed to let children explore how air flows through a system, and to introduce them to the relationship between pressure and volume, as well as the way in which pressure can exert forces on objects (see also De Berg, 1995). Although there are some basic questions the researcher asks every child during the administration of the task, most of the interaction between the boy and the researcher emerges in real-time, i.e. during the task itself.

Between three and seven years of age, important changes in children’s conceptual understanding of scientific concepts take place (Van Geert & Steenbeek, 2008), in addition to changes in curiosity and exploration tendencies (Simonton, 1999), which are probably related to important changes in children’s lives. That is, they go through a major transition when they enter first grade, and start learning to read, write, and to do arithmetic (Carrière, 2009). During this age period children’s learning behavior gets shape, attitudes toward school are formed, and first interactions with peers and teachers in a school setting emerge, which are the building blocks of academic performance at a later age.

Moreover, this is also the age at which important cognitive developmental transitions take place. From the work of Piaget (1947/2001) we know that children between three and seven years old are in the pre-operational stage of development, which is characterized by the forming of concepts, and the use of symbols to think about the world, but also by centrism, i.e., focusing on a single aspect instead of more aspects while children reason or solve problems. More recently, research using skill theory, which is inspired by Piaget’s theory, illustrated that the highest skill (understanding) level that children first reach between 3 and 7 years of age develops from single representations (understandings that go beyond specific actions on objects) to representational systems (linking several of these representations that define the object or concept at hand—see also section 3) (Fischer & Bidell, 2006). However, this research also showed that children vary enormously in their skills across context, tasks, and within short periods of time. This variation is due to the fact that context dynamically contributes to the deployment of skills in the form of a real-time activity. That is, thinking or understanding takes place in the form of action. How does the process of understanding occur in action, taking into account the real-time interactions that constitute this process in a teaching environment, and taking into account the vast amount of intra-individual variability?

Based on our ongoing longitudinal research project, we will illustrate how short term “building blocks” of understanding give rise to various long-term patterns of understanding. In order to fully understand these short-term building blocks, we have selected one particular problem domain for this chapter, namely air flow and air pressure, because it provides a domain that is both limited and rich enough to study. Zooming in on these short-term interactive processes gives us important information to understand the development and transformations of understanding on the long term (Steenbeek, 2006; Thelen & Smith, 1994).

During the ongoing longitudinal research project, a researcher repeatedly visits 32 young children (3 to 6-years old) as part of an ongoing longitudinal study on children’s understanding of scientific concepts, such as the flow of air and air pressure. During one visit, the researcher presents each child with two empty medical syringes without a needle, which are joined together by a small transparent tube. One of the syringes’ pistons is pulled out. “What do you think will happen if I push this [piston] in?” is one of the questions the researcher asks. This question triggers a variety of answers from the children. Some children think nothing happens, others say the tube will pop out, whereas others even think the material will explode. Some children say they don’t know, while others predict that the piston of the other syringe comes out, which is the right answer in this case. After the researcher demonstrates what happens, researcher and child discuss about possible explanations for this phenomenon. Again, multiple answers are given. Some children simply say they don’t know. A few mention batteries or electricity as a causal explanation, whereas others say that water flows through the syringes and causes the piston to move upwards. Some children emphasize the tube that connects the syringes, and others understand that air flows through the tube and syringes.

What accounts for the differences in young children’s understanding of scientific concepts, and what is the role of the environment, i.e., the teacher in supporting and promoting this understanding? To answer this question, a model of children’s scientific understanding should take the complexity and dynamic nature of this into account, as well as the complex interactions with the environment on which the understanding of children is often based (Fischer & Bidell, 2006). This chapter aims at explaining how children’s understanding of scientific concepts can be studied using a model based on properties derived from dynamic systems theory (e.g., Van Geert, 1994) and skill theory (Fischer, 1980; Fischer & Bidell, 2006).

3.1 Dynamic Systems and understanding

A dynamic systems complexity approach describes how one condition changes into another, and how different time scales are interrelated (Van Geert, 1994; Van Geert, 1998; Van Geert & Steenbeek, 2005a, 2008; see also the theory of embedded-embodied cognition of Thelen & Smith, 1994). Research in the dynamic systems paradigm investigates real-time processes and captures development as it unfolds through multiple interactions between a child and the environment (Van Geert & Fischer, 2009). Such development can be viewed as a self-organizing process, since the state of the system organizes from the multiple interactions among the elements (e.g., the child and environment). Over time, the system's state may emerge toward certain stable states, or attractors (e.g., Thelen & Smith, 1994). Dynamic systems theory has so far proven to be a valuable framework for studying human development, including reflexes (Smith & Thelen, 2003), parent-child interactions (Fogel & Garvey, 2007), language development (Van Dijk & Van Geert, 2007), scaffolding in teaching-learning situations (Van Geert & Steenbeek, 2005b), dyadic play interactions (Steenbeek, 2006), identity development (Lichtwarck-Aschoff, Van Geert, Bosma, & Kunnen, 2008), and cognitive development (Fischer, 1980; Fischer & Bidell, 2006). The approach makes use of methods to investigate time-serial processes, and test dynamical relations between these processes (Cheshire, Muldoon, Francis, Lewis, & Ball, 2007; Lichtwarck-Aschoff, et al., 2008; Steenbeek & Van Geert, 2005; Van Geert & Steenbeek, 2005a; 2007). For example, Van Geert and Steenbeek (2005; 2007) present mathematical models to predict patterns and variations in combinations of variables over time. Other authors used time series to describe relationships between variables (Van Dijk & Van Geert, 2007) or state space grids (Hollenstein, 2007) to investigate interactions between dyads; as opposed to probabilistic approaches which rely on deviations from the mean and group differences.

Applying a dynamic approach to the study of understanding scientific concepts means that several properties of this approach have to be taken into account. Below, four properties (intertwining person-context dynamics, iterativeness, interconnected time scales, and microgenetical variability)[19] and examples of their application to the study of understanding (of e.g., scientific

---

[19] Actually, the dynamic systems approach has many more properties or "tools" (Howe & Lewis, 2005) to study development. However, we highlighted these four specific properties to illustrate how this approach sheds new light on the study of understanding scientific concepts.

concepts) will be discussed. In section 5, the properties will be illustrated in light of an empirical example, in combination with skill theory's framework to measure the complexity level of understanding (Fischer & Rose, 1999).

3.1.1 Intertwining person-context dynamics

Vygotsky (1934/1986) already pointed out that children develop understanding in close cooperation with their teachers and the material. His concept of the zone of proximal development is a dynamically changing concept, in which teacher and child co-construct the child's development. This means that the child's skills and understanding are constructed by a series of actions guided by the educator, instructions and tool-use, which are then internalized and personalized (cf. Van Geert, 1998; Van Geert & Steenbeek, 2005a).

From a dynamic systems perspective, understanding is seen as a process of intertwining person-context dynamics (Thelen & Smith, 1994), meaning that the social (e.g., the science teacher) and material environment (e.g., materials used in science class) play an active part in the process and cannot be viewed separately, or merely as an outside-based influence. In fact, these elements are intertwined across time, in a continuous person-environment loop: at any moment in time, one component (e.g., the child) affects the other (e.g., the teacher) and the other affects the first, thus creating the conditions under which both components will operate during the next moment in time (Steenbeek, 2006). For example, interactions between a child, a researcher, and the syringes-task will organize toward certain distributed patterns of understanding at that moment (in real time), which eventually evolve toward stable attractors on a longer time scale (Halley & Winkler, 2008; Thelen, 1989). Hence, understanding is an active process of what the child constructs in interaction with (not just within) a specific environment, in which each individual contribution is virtually meaningless if not viewed in light of the interaction (Van Geert & Fischer, 2009). Merged together, person and context become what Fogel and Garvey (2007) call a "cooperative unit", in which both components not only contribute to the process of development, but are highly intertwined and form a unique process together.

Representationalists, such as Fodor (1981) hold the idea that understanding takes the form of internal structures (representations) within the child's mind. A child's scientific understanding thus consists of a collection of these internal structures which represent scientific facts and concepts, which are activated and used to coordinate our behavior toward the current environment (Haselager, De Groot, & Van Rappard, 2003). In this case, a concept or representing model of the air pressure task would be represented in the child's mind, and this representation would guide the child's behavior as he or she is working on the actual air pressure task.

Terms such as “concept” or “representation” are actually more or less undefined, and derive their meaning from a particular theoretical framework. From a representationalist (or information-processing) view, these words refer to internal entities responsible for our thinking or actions toward the environment. From a dynamic view, however, these words refer to processes, perception and action structures that emerge within a specific environment (Van Geert & Fischer, 2009). Perceiving, acting and thinking are conscious processes that take a particular shape in the stream of consciousness of the participants, such as a child and the researcher (Van Gelder, 1995; 1998). This shape is governed by the participants' actions on the objects, such as the syringes, or on physical representations of the syringes, such as prints or drawings, within their current context, and should not be identified with a retrieval of internally stored representations (Van Geert, 2011a).

We can construct much of this stream of consciousness by carefully watching the ongoing interaction between child and environment in terms of the intertwining of various forms of verbal and non-verbal behavior, such as eye and head movements, gestures, pointing, verbal descriptions, manipulations of the materials, etcetera. The child's current understanding of the concept at issue (for instance, the flow of air through two syringes connected by a tube), is the child's continuously changing state of mind, or stream of consciousness, as he picks up and reacts to whatever goes on in the current dynamic interaction. Thus, despite the fact that the process of constructing an understanding is a distributed process, involving the intertwining of person and context, understanding can still be specified as an individual and "internal" process corresponding with the individual child's ongoing state of mind, but only as a changing state that unfolds in this active process (Van Geert, 2011a). Hence, representations are structures that emerge during a specific interaction in a specific environment, and are not internal symbolic structures which guide behavior.

3.1.2 Iterativeness

Within the process that results from an intertwining between person and context, understanding emerges through iteration, that is, every step in understanding is based on the previous one and embedded in the current context. More precisely, iterativeness (sometimes referred to as recursiveness) involves a series of computational operations, in which the input of the next operation is the output of the previous one. For instance, if a child determines that an empty syringe contains air, he can build on this knowledge by trying out what happens if he joins two of these syringes together by using a tube. Understanding changes through repeated interactions, instead of being the retrieval of a complete representation that is already there in memory. During a teaching interaction, each previous action of the child has an influence on the subsequent (re-)action. In other words, the existing understanding is the basis for the emergence of the next understanding as it develops in the interaction.

In its simplest possible form, a dynamic systems model specifies the change in a variable (L) over time (t) as a function of the current level of the variable: $L\ t+1 = f\,(Lt)$. The function f refers here to the change in 'understanding', but can specify any sort of influence or mechanism of change (Steenbeek, 2006). Understanding does not consist of particular moments within the interaction (e.g., when the child answers), but is in fact the whole iterative process itself, and every interaction unit is a component of this holistic understanding process during a particular problem solving event. Even though understanding consists of the whole iterative process, the child's answers are a reflection of the child's ongoing state of mind within that process and reveal his or her understanding at that very moment in time.

As Howe and Lewis (2005) point out, the iterative nature of the process of understanding can also explain some of the differences between children. When children’s understanding depends on interactions, and each interaction is based on the previous one, small differences between children’s initial states of understanding can grow bigger over several interactions. This is particularly so if the process takes the form of a positive feedback loop amplifying idiosyncratic properties of the answers, i.e. properties that are typical of a particular child. For example, if the child focuses on only one syringe and the researcher’s follow-up questions center on that syringe as well, the difference between this child and another child who focuses on both syringes grows bigger. However, if the process takes the form of a negative feedback loop reducing the idiosyncrasies, small differences in initial states will most likely remain small over the course of the problem-solving process. This would be the case if the researcher switches the focus of her follow-up questions to the other syringe, thereby scaffolding the child towards a more complete picture of the task. The difference between this child and the child who initially focused on two syringes then becomes smaller.

3.1.3 Time scales

The property of interconnected time scales entails that the dynamics of long-term development of understanding are intrinsically related to the dynamics of short-term processes of understanding (Thelen & Smith, 1994; Lewis, 2000). That is, in order to get a grip on long-term changes in understanding of children, it is worthwhile to focus on the short-term (micro-genetic) process, and examine properties of that process, such as variability (Granott, Fischer, & Parziale, 2002; Steenbeek, 2006).

Iterativeness occurs on the short term as well as on the long term, meaning that on the short term (e.g., during one interaction between child and teacher in science class), each step in understanding is based on the previous step in understanding, while on the long term each interaction builds on the preceding interaction (e.g., the interaction during last week’s science class). In this way, the same mechanisms are sculpting the development of understanding over a shorter and longer period. Thelen and Corbetta (2002) indicate that the general principles underlying behavioral change work at multiple time scales. The short- and long-term scales interact, in that repeated (iterative) processes on the short term time scale influence processes on the long-term time scale (Lewis, 2000). In addition, the emergence of large-scale patterns also influences what happens on the short-term time scale, by shaping the structure and function of the interaction on the short term (Lewis & Granic, 2000; Smith & Thelen, 2003; Steenbeek, 2006; Van Geert & Steenbeek, 2005a). The underlying idea is that all levels of the developing system interact with each other in a self-organizing way, and consist of nested processes that unfold over many time scales, from milliseconds to years (Lewis, 2000; Thelen & Smith, 1994).

3.1.4 Microgenetical variability

As a result of the iterative organization of the components and the intertwining between child and context that mark the process of children's understanding, we can observe microgenetical variability. This means that the complexity of children's understanding fluctuates within very short periods of time, e.g., during one task. While studying the processes of developmental change, it is crucial to take many observations (adopting a microgenetic research method) to detect the subtle changes that constitute understanding and its development (Kuhn, 1995; Siegler & Crowly, 1991). Researchers note that, driven by bi-directional interactions with the environment, the complexity of children's understanding can increase during a task, but also temporally decrease, for example when the task difficulty increases, when the teacher's support decreases, or when children encounter something unexpected while working on a task. Understanding can change gradually or abruptly in a stage-like pattern in a short timeframe, even during a single task (Siegler & Crowly, 1991; Yan & Fischer, 2007).

Researchers have suggested that this variation is an important factor in development, since an increase in variability may be related to the ability to reach higher levels of skill (Howe & Lewis, 2005; Thelen, 1989), or, more generally, to a transition to another pattern of behavior (i.e., attractor) (e.g., Thelen & Smith, 1994; Van Geert, 1994). The variability on the short-term (e.g., during the syringes-task or during a science lesson) can therefore yield important information about how the developmental pathways of understanding will be shaped on the long term.

In order to capture the complexity of understanding and variations in complexity over a short and longer time periods, we can use skill theory's framework of cognitive development (Fischer, 1980; Fischer & Bidell, 2006). This framework can be used on both the long- and short-term time scale and is compatible with a dynamic systems approach. Even more so, skill theory could be considered as a specific dynamic system's theory applied to human skill development, since it assumes skills are built in an iterative and hierarchical way, i.e. each skill level builds on the previously obtained skill level. Moreover, skills are highly context-dependent and fluctuate over time, that is, they depend on the constraints and affordances of the context in which they are mastered (Fischer & Bidell, 2006).

3.2 Skill theory and understanding

Skill theory focuses on the complexity and variability of children's skills, which consist of actions and thinking abilities, and the way these are constructed (Fischer, 1980; Fischer & Bidell, 2006). Since skills are thinking structures mastered in a specific context, such as a science class, they hold both person-related as well as context-related characteristics (Parziale & Fischer, 1998). An example of a skill is a child's ability to understand how air pressure works while manipulating the syringes-task. This understanding is reformulated when the student works on a similar task in another environment (e.g., with different materials or without the help of the researcher). Skills are thus highly influenced by the possibilities and constraints of the situation in which the skill is used.

Skill theory explains both long- and short-term development of skills by measuring these on the same hierarchical complexity scale. This complexity scale consists of 10 levels, grouped into 3 tiers, which are sensorimotor, representational or abstract by nature. The scale can be applied to different cognitive (Fischer & Granott, 1995; Schwartz & Fischer, 2005), social (Fischer & Bidell, 2006) and language domains (Fischer & Corrigan, 1981), as it focuses on hierarchical complexity rather than content. This makes skill theory especially suitable to describe differences between children, as well as differences between skills in different domains for the same child (Parziale & Fischer, 1998).

A child's understanding within a domain, as an emergent process in real-time, can be viewed along two dimensions: the first being the dimension of content (the subject), the second of complexity (the complicatedness). In order to evaluate children's understanding (of, for example, air pressure), we need a fair ruler to determine how elaborate their understanding is, and to evaluate whether they need extra help in some areas. One of the most powerful characteristics of skill theory (Fischer, 1980) is that it extracts complexity from content, resulting in a content-independent ruler of understanding. Because of the content-independent nature of the way skill theory approaches understanding (or other skills), it enables researchers to compare understanding across multiple time points, contexts, persons, and for different age ranges.

According to Fischer and colleagues (Fischer, 1980; Fischer and Bidell, 2006), development in a particular domain goes through 10 levels of skills hierarchically grouped into three tiers that develop between 3 months and adulthood. The first tier consists of sensorimotor skills: simple connections of perceptions to actions or utterances. An example is a statement that two syringes are attached to a tube. Sensorimotor skills form the basis of the skills in the two subsequent tiers, i.e. they are the building blocks of the higher levels. The second tier constitutes of representational skills, these are understandings that go beyond current simple perception-action couplings, but are still based on them. Hence, the term representation refers to the coordination of several sensorimotor skills at the same time, not to an internal symbolic structure (Fischer, 1980). Within the context of the air pressure task for example, the child can predict what will happen if the piston is pushed in without literally touching or manipulating the syringe. Nonetheless, what he or she predicts depends on the material context, and on the sensorimotor skills that he or she mastered before. The third tier consists of abstractions, which are general nonconcrete rules that also apply in other situations (Schwartz & Fischer, 2005). This would be an explanation about the relationship between pressure and volume inside a syringe.

Within each tier, three levels can be distinguished[20], each one more complex than the previous one. The first one can be characterized as a single set, meaning a single action (or a single representation, or a single abstraction). The second level is a relation between two of these sets, which is referred to as a mapping. The third level is a system of sets, which is a relation between two mappings, in which each mapping consists of a relation between single sets. After this level, a new tier starts, which is divided in single sets, mappings and systems as well (Fischer & Bidell, 2006). For the emergence of each level, evidence of discontinuities and differences between levels has been demonstrated using analysis methods based on Rasch scaling (Schwartz & Fischer, 2005).

Fischer and colleagues (Fischer, 1980; Fischer & Bidell, 2006; Granott & Parziale, 2002; Schwartz & Fischer, 2005; Yan & Fischer, 2002) showed that skill theory can not only describe and explain the development of skills on the long term, but also describe the micro-genesis of problem solving. When facing a new task or problem within a domain, even high-skilled adults go through the same cycles of development. That is, at the beginning they show skill levels that are mostly sensorimotor, which build up to more elaborate levels during the course of the task. During a task (and also during the long-term development of skills), people do not go through the skill cycles in a linear fashion. Instead, they repeatedly build up skill levels and show collapse before they obtain their highest possible level, something Yan and Fischer (2002) call “scalloping”. During a task, people vary constantly within a bandwidth between their highest and lowest possible complexity levels, which is also known as the developmental range. The highest levels within the bandwidth are only reachable when the environment provides sufficient support (Fischer & Bidell, 2006; see also Yan & Fischer, 2002).

---

[20] After the 3 levels of the abstraction tier, a higher complexity level emerges, also known as ‘single principles’, which is the 10th level of the scale. Additionally, people function on the few highest levels usually in early adulthood, but only for their domains of expertise. For most other domains, people function on a lower complexity level.

Skill theory also accounts for inter-individual differences in understanding and is therefore especially suitable for describing individual developmental pathways (Fischer, Rose & Rose, 2007). Yan and Fischer (2002) showed that adults' performance on a computer task can move through a variety of pathways, each one showing nonlinear fluctuations. Of all participants, novices showed the most frequent and rapid fluctuations in performance. Experts however fluctuated less frequent in their performance, meaning that variations followed on each other in a slower fashion.

In sum, a model of understanding needs some kind of ruler to determine the complexity of understanding levels children show. Skill theory (Fischer, 1980; Fischer & Bidell, 2006) provides a content-independent ruler for understanding, which can be applied to different time scales of development, and takes both the role of context, as well as inter- and intra-individual variability into account.

3.3 A model of understanding

Using the four properties from the dynamic systems paradigm and Skill theory's ruler, we can construct a model of understanding to guide research and practice in education, but also in other areas that require the evaluation of cognitive growth. The general model of understanding here is that it is an active process, distributed across people involved, and that it is dynamic, i.e., it continuously changes, and self-organizes through iteration. It is important to keep in mind that, even though the four properties describe distinct mechanisms, they all work at the same time while the process of understanding unfolds. Below, we will present the model and briefly highlight its components, after which we discuss these in more detail by using an empirical example.

As Figure 3 shows, children construct levels of understanding during short-term interactions with the environment, such as during a task they are working on together with an adult. Both child and adult are characterized by specific distal factors (e.g., years of schooling) that influence their behaviour. However, those distal factors are not what we focus on, since the figure can be characterized as an action model, that is, it focuses on understandings which are constructed during an interaction by means of a process that is distributed across the child, the adult, and the material context with which they interact or which they manipulate. This means that during an interaction, there is a bidirectional influence between the child's answers and the adult's questions within the material context. This is illustrated in the big square (part A) of Figure 3.

Moreover, the process is iterative, meaning that it changes through repeated interactions, instead of being the retrieval of a complete representation that is already there in memory. During a teaching interaction, each previous action of the child has an influence on the subsequent (re-) action. This is illustrated by the big arrows between adult and child (part B of Figure 3) and the small arrows on the side of the boxes indicating the child and adult.

Each task-related utterance has two dimensions: a specific content and a complexity level. During interactions, we can observe the complexity level of understanding, as it comes forward in the child's distinct utterances, which are often reactions to what the adult is saying, or are part of the ongoing discussion between an adult and a child. This complexity level, measured by skill theory (Fischer, 1980), will vary between different children, and will fluctuate over time within the same child. This is illustrated by part C in Figure 3.

Lastly, the long-term development of children's understanding unfolds through several of these short-term interactions. As an example, Figure 3 displays the sessions with 3-month intervals we used in our study of young children's understanding of scientific concepts. The link between short- and long-term development is indicated in part D of Figure 3.

3.4 An empirical example and illustration of the model

In the next sections, we illustrate the model and the four properties by using an example (see Table 2) derived from our empirical study focusing on the long-term development of understanding air pressure (and other scientific concepts, such as gravity) in three to seven year old children. Table 2 is an excerpt of a transcribed session in which a boy (4 years, 6 months) and a researcher explore the syringes task mentioned in the introduction. The transcript starts right after the point in which the researcher and the boy explored the exterior of the syringes. That is, they compared them in size and examined the numbers written on the outside.

Figure 3: A conceptual (action) model of understanding based on principles derived from dynamic systems theory and skill theory

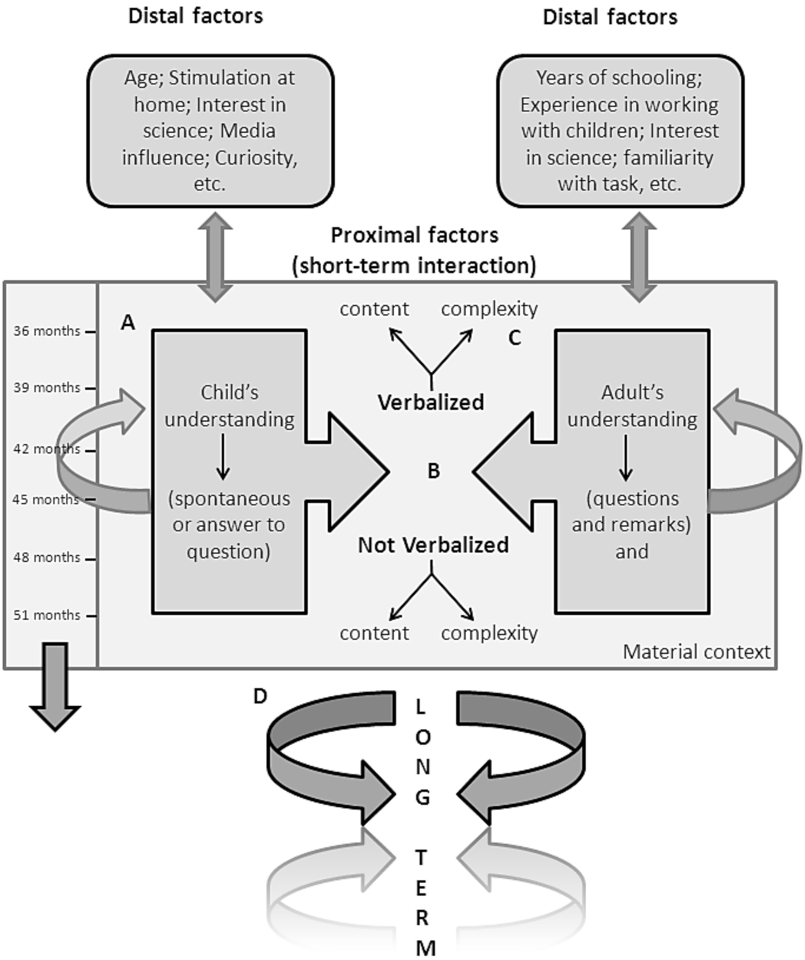


Table 2: Excerpt of a session from our longitudinal project in which a boy (4 years, 6 months) explores the syringes task together with a researcher

Person	**Content:** *verbal (gestures, manipulations, gaze directions between brackets)*	**Complexity**	**Nr**
Researcher	*(Attaches the two syringes by a small transparent tube, gives one syringe to the boy)* "I attached the tube to these. What do you think will happen if I push mine in?"		1
Boy	*(Looks at his own syringe)* "I don't know"	No level	2
Researcher	"But what do you think?"		3
Boy	*(Looks from the researcher to his syringe)* “Uhm..."	No level	4
Researcher	*(Pauses)* "You said they are the same. I pulled this piston out *(Touches the piston)*, and pushed the other piston in *(Points down to the other piston)*. Then I attached the tube. What do you think will happen if I push this one in?” *(Gestures as if she is pushing down)*		5
Boy	"Then this one will go up like this." *(Holds his syringe in one hand, while his other hand pushes the end of the piston on the table, then he moves his hands up)*	Single representation (prediction)	6
Researcher	*(Points to this syringe the boy holds)* "Is that one going up?"		7
Boy	"Yes, and then that one is going down" *(Points at the piston of the syringe the researcher is holding)*	Single representation	8
Researcher	"Really? Why does that happen?"		9
Boy	"Because we attached the tube." *(Follows the tube with his finger to the tip of his syringe)*	Sensorimotor system	10
Researcher	"I see… If we would take away the tube, it wouldn't work?"		11
Boy	*(Shakes his head)* "No".		12
Researcher	*(Pushes her piston in, pauses)* "Were you right?"		13
Boy	*(Watches his own syringe as the piston pulls out)* "Yes"		14
Researcher	"Can you do it as well?" *(Holds her syringe up)*		15
Boy	*(Looks at both syringes, pushes the piston of his syringe in)*		16
Researcher	“How is this possible? You're pushing it over there *(Points at the piston of the boy’s syringe)* and then this one goes backwards!”		17
Boy	*(Pushes piston in and pulls it out)* "I don't know"	No level	18
Researcher	"OK, but it has something to do with the tube, you said. What do you think is inside the syringes and tube?"		19
Boy	*(Pauses for a long time, looks around)* "I don't know"	No level	20
Researcher	"I think there's no water in it" *(Shakes her syringe)*		21
Boy	"No" *(Starts shaking the syringe)*		22
Researcher	"But then, what is in it? And how is it possible that we can move one by pushing the other?"		23
Boy	"Because this is attached *(Touches the end of the*	Sensorimo-	24

	tube) and then it can move out" *(Pulls the piston out)*	tor system	
Researcher	"You know what; we can also attach a longer tube! *(Gets a longer tube)* "What do you think will happen then?"		25
Boy	*(Gets the end of the tube and attaches it to his syringe)* "I think still the same."	Single representation (prediction)	26
Researcher	"Even with a longer tube?"		27
Boy	"Yes" *(Looks at his syringe)*		28
Researcher	*(Pushes her piston in, it works)* "So now it works as well"		29
Boy	*(Pauses, pushes the piston of his syringe in, then pulls it out)*		30
Researcher	"So it has to do with the tube or something like that..."		31
Boy	"Yes, because the tube is attached to this one *(Looks at syringe while he pushes the piston back in)*, and it is attached to here *(Points at the point where syringe and tube are connected)*, and then goes *(Makes a gesture for pushing the piston in)* this *(Points at the tip of the syringe)*, it goes like this" *(Follows the tube from the tip until he is halfway)*	Sensorimotor system/ single representation	32
Researcher	"I see…what do you mean when you say 'this'?"		33
Boy	*(Keeps on following the tube with his finger, can't reach for the last bit, so follows it in the air)* "The tube, it goes like this"	Sensorimotor system/ single representation	34
Researcher	*(Follows the last bit of the tube with her finger)* "Yes, but what is going through the tube?"		35
Boy	"That... *(Pauses and looks at the tip of his syringe)* "The sigh is going through the tube *(Gestures for pushing the piston in)* "And then it goes, like this, and this, and this" *(Follows the tube until halfway)*	Single representation/representational mapping	36
Researcher	"The sigh is going through the tube and flows to mine?"		37
Boy	"Yes" *(Pulls the piston of his syringe out)*		38

3.4.1 Example of person-context dynamics – social construction

An important part of these context dynamics is the social part of the context, meaning the people around the child. Thus, the development of the child's understanding occurs in interaction with the social environment (e.g., the teacher), and it is this interaction that drives the process of understanding, enabling the student to receive adaptive assistance and make progress step by step (Hirsch-Pasek, Golinkoff, Berk, & Singer, 2009; Van Geert & Steenbeek, 2005a). In our example (see Table 2), the child constructs his answers together with the researcher. The researcher's questions are guided by, and on their turn guide, the child's answers. An illustration of this can be seen in fragments 2 to 6 of Table 2. After the boy answers he does not know what happens with the syringe he is holding if the researcher pushes the piston of the other one in, the researcher asks him "What do you think?" In this way, she is trying to get the boy to make predictions, encouraging him to hypothesize. In response, the boy looks around and does not answer the question. The researcher, in turn, helps him getting started by summarizing what he said before and by a verbal repetition of her actions with the task material. After having heard the adult's repetition of her actions, the boy starts to construct an answer on a higher complexity level than before. In terms of skill theory, this answer can be classified as a single representation, as he makes a prediction that goes beyond simpler perception-action couplings (skill levels, when applicable, are indicated the right column of Table 2).

Two things are important here. First, the researcher is responding to the boy in this way, because he did not know the answer. Had the boy given the answer, she may had pushed the piston in, or asked him to elaborate on his answer. Because the boy does not know the answer, she needs an approach to determine whether he really has no idea, and if so, how she can help him to make a prediction based on what he knows about the syringes. In order to do this, she tries out two different approaches. First, she asks him what he thinks, which can be a starting point for further elaboration on his side. When the boy does not reply, she decides to help him to get started by giving some information about what they have done and seen before. The boy now hypothesizes what happens if the piston of one of the syringes is pushed in. The answer to the question "What do you think will happen?" (see fragment 1 of Table 2) is therefore the product of the interaction between the boy and researcher. In her reactions to the boy's "I don't know" the researcher is trying to guide his understanding. In turn, after hearing the researcher's summary, the boy constructs his understanding. What happens with regard to the boy's understanding during the interaction with the researcher is not mere retrieval of earlier gathered knowledge, or a reaction to a trigger (whether it be the syringe itself or the questions), but a (re)construction of knowledge through a constellation of interactions with researcher and material. If we look at understanding while it occurs in real time, we can only study the person-context aggregation that results from this interactive process and cannot distinguish the unique contribution of the individual components (Van Geert & Fischer, 2009). Even though one can describe what the child does in answer to a specific action or expression of the adult; it is not possible to distinguish the adult's or child's contribution to the (variance in) understanding during the task.

Parallels can be drawn with other teacher-student interactions, such as in scaffolding during instructions in arithmetic lessons. In their model of scaffolding, Van Geert and Steenbeek (2005) model the process of scaffolding during an arithmetic class taking a dynamic systems complexity approach. Scaffolding is an interactive process in which the student makes progress using the help of a teacher, which scaffold-level should be adapted to the student's level in order to have the right effect. One of the most interesting properties of this dynamic model is that it accounts for transactions between teacher and student, and that it portrays a dynamic, real-time combination of both the student's performance level and the scaffold-level of the teacher. One of the parameters in the model is the *optimal scaffolding distance*, a bandwidth which differs among individuals and contexts, within which help stimulates learning. Within that bandwidth, the optimal scaffolding distance is the distance between the pupil's level and the level of help or scaffolding for which the learning effect is maximal. Just like in our model of understanding, the actions of student and teacher form a unique process built of bi-directional relationships (Fogel & Garvey, 2007).

3.4.2 Example of person-context dynamics—the material context

In addition to the social context, the material context (such as the syringes) also plays an important role in the process of understanding. The syringes should not be conceived of as fixed or monolithic things, but are instead part of the emerging dynamics. Even an unmovable material object is dynamic in terms of its effect on the child, in the sense that the child continuously changes his angle of vision towards the object and thus sees different parts of the object. The dynamic and intertwining nature of the material context is even more strongly illustrated by the syringes task, in which the child or the adult manipulate the syringe, and are thus changing the nature of the object in line with their activities.

In the example (Table 2), the syringes and tube are frequently touched by the boy and the researcher to emphasize or guide their verbal expressions (see fragments 5, 6, and 10). The best illustration of this, however, can be found in fragments 32 to 36. In this fragment, the boy uses the material extensively, after which a higher level of complexity emerges: he transitions from a sensorimotor systems level to a single representation/representational mappings level. Note how the boy substitutes words for gestures and pointing in fragments 32 and 34, following the process of what happens with his hands. Parallels can be drawn with fragment 5, in which the researcher is talking the boy through what happened before. In fragments 32 and 34, however, the boy uses the material instead of the researcher's words to construct his understanding. Before fragment 32, he predicted that one piston comes out when you push the piston of the other syringe in. However, so far, he was not able to explain why. Now, using his hands to examine the syringe, he is able to represent the process, and concludes that "it" is going through the tube. Eventually, guided by the researcher's question "But what is going through the tube?" which seems to suggest that he is on the right track, he is able to replace the word "this" in his explanation for "sigh".

3.4.3 Example of understanding as an iterative process

In Figure 3, the iterative character of the understanding dynamics between student and researcher is shown in that each previous action of the student has an influence on the subsequent (re-)action of the researcher, and vice versa. Over time, each session has an influence on the subsequent session of this student-researcher pair, which implies that the influences between the child and environment are bidirectional, meaning that not only the action of the researcher influences the next (re)action of the student, but also that the previous interaction influences the next interaction. Iterativeness is thus the form in which the cyclical or reciprocal character of causality occurs.

In our example (Table 2), the iterative nature of the process is not only illustrated by how the researcher and child react to what has been said previously throughout the whole transcript, but also by how the child's understanding develops during the interaction. With regard to the prediction he makes in the first half of the interaction, the child goes from "I don't know" (fragments 2 and 4; no skill level) to "This one goes up like this" (fragment 6; single representation). This change in understanding is constructed in reaction to what the researcher said right before in fragment 5. With regard to the explanation of the boy why this happens, his understanding goes from "Because this [the tube] is attached" (fragment 24; sensorimotor system), to "Something goes like this [through the tube]" (fragment 32; sensorimotor system/single representation), to "The sigh is going through the tube" (fragment 36; single representation/representational mapping)." The statement that the tube is attached, which the researcher repeats and emphasizes in fragments 19 and 31, leads to the conclusion that there must be something flowing inside the tube. Since there is no water in the tube fragments 21 and 22), or anything else visible for that matter, it must be "sigh" (fragment 36).

This step-wise refining of the boy's understanding, in which each previous step is the beginning of the next step, illustrates the iterative nature of the process nicely. Not only does iterativeness occur on the conversation level (what the child says depends on what the researcher said previously and vice versa), it also occurs on the complexity level of understanding (each understanding of the child depends on the previous understanding). Finally, the iterative nature of the process can also be seen over sessions, meaning that previous sessions influence subsequent sessions.

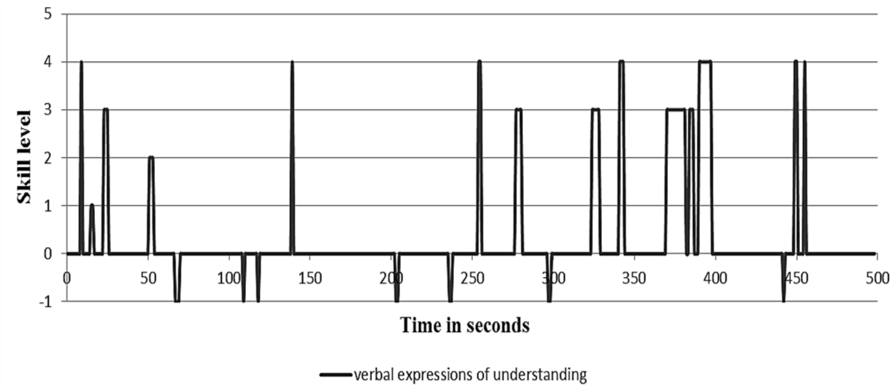
3.4.4 *Example of micro-genetic variation*

In our example (see Table 2), microgenetical variability is seen in the child's understanding of how the material works. First, in fragment 10 the boy names a single cause for what happens: "Because we attached the tube". This is an answer on a sensorimotor system level; he gives a single, observable causal explanation for the phenomenon, not taking the volume of the syringes or the air into account (see also the third column of Table 2). Over the course of the interaction, he briefly regresses to "I don't know" (fragments 18 and 20; no skill level), and restores his previously gained skill level again in fragment 24: "Because this [the tube] is attached". From there, he further constructs his understanding, and eventually reaches a higher level in fragment 36: "The sigh is going through the tube", for which he needs a representation of the role of air in the system.

In Figure 4 a time-serial illustration of the fluctuations in the boy's answer levels during the air pressure task is depicted. The graph shows how the understanding of the boy fluctuates over time. While skill theory's level 4 (single representation) is mostly observed during the interaction, the boy also regularly shows understandings at level 3 (sensorimotor system). Even though his understanding seems to increase in complexity over time (on average the boy reaches level 4 more often in the second half of the interaction), his understanding often regresses to level 3 and to incorrect/irrelevant understandings. Hence, understanding is not a fixed entity, but varies over time, even within a single task.

The short-term intra-individual variability influences the variations in development we can see on the long term (Fischer & Bidell, 2006; Van Geert & Fischer, 2009). If microgenetical variability is associated with reaching higher-level skills (Howe & Lewis, 2005; Thelen, 1989), long-term trajectories of understanding may differ between children showing more periods of variability versus children showing little periods of variability within short-term interactions. This also makes sense in combination with the property *Iterativeness*, as a short-term interaction showing a broad range of skill levels makes it more likely that skill levels subsequently move toward a higher level (cf. a phase transition), compared to a previous interaction showing a narrow range of skill levels. After all, the interaction with a broad range of skill levels yields more possibilities for the next interaction than an interaction with a narrow range. In conclusion, as Howe and Lewis (2005) mention, understanding gets form over various instances and in turn, drives long-term developmental change. This connection between the short- and long-term scale of development brings us to the next property, that of interconnected timescales.

Figure 4: Time-serial illustration of the variability in the boy's understanding during the air pressure task.



Note. Complexity levels are measured using a coding system based on skill theory. For this boy, levels on the y-axis range from 1 (single sensorimotor set) to 4 (single representation). A score of -1 represents an incorrect or irrelevant answer.

3.4.5 Example of interconnected timescales

Three months later, the researcher returns with the syringes and the tube. The researcher starts by asking "Do you remember what we had to do with this?" In response, the boy immediately grasps the material and attaches the tube to the syringes. Then he replies: "Yes, when you push this one in, the air will go over here". He doesn't need more time to think about the process in a stepwise fashion: That it works like this because the tube is attached, that there must be something going through that tube, etcetera. Based on the previous interaction, he now knows that air is going through the tube and makes the pistons move. Note, however, that this is not a mere retrieval from memory. The boy first attaches the syringes to the tube, and answers afterwards. Moreover, the question of the researcher is phrased in a way that encourages him to think about what they did before. Even though the researcher's role is not as prominent as it was in the previous interaction, the social context still plays a role in the construction of understanding. However, three months earlier, the understanding was clearly a co-construction between child and researcher. Now the child can directly introduce this understanding to the interaction, triggered by the researcher's question and the material, but without further interference.

3.5 Discussion

From a theoretical point of view, we discussed a number of dynamic properties in combination with skill theory's ruler of cognitive development. We argued that using these properties and ruler give both educators and researchers important means to get a grip on how children's understanding of scientific concepts builds up over time. More specifically, it helps to understand how children organize their knowledge in concordance with the context, i.e. the teacher, and highlights the importance of being aware of teachers' accounts in conversations with children, for example during a science lesson.

There are many different types of knowledge generation processes, one of which is the socially situated process between adult, child and task that we are discussing here. When a child is assessed or diagnosed, a different process of knowledge generation occurs. In these instances, the child is asked to construct knowledge without the help of an adult, but usually in interaction with a particular symbolic substrate, such as a piece of paper to draw on, or the structure of language that the child is using to describe knowledge. It is however wrong to think that only the latter process (in which the child works without help) is a reflection of the child's "real" knowledge. In fact, both the co-constructed as well as the individually constructed knowledge reflect the child's "real" understanding. Variations in complexity levels within one type of knowledge generation, but also between different types of knowledge generation, illustrate the intrinsic variation of understanding as such.

The model we proposed helps in re-conceptualizing the process of understanding in individual children, and the underlying mechanisms of change in their understanding. The latter is especially important, since “Developmental psychologists are not simply interested in the stable states achieved by individuals along their lifespan, but also about the mechanisms of change that lead from one state to the next.” (Howe & Lewis, 2005, p.248). The advantage of a dynamic systems approach to the study of understanding is that it makes the development of understanding more transparent and no longer limited to an invisible process inside the individual learner (Van Geert & Fischer, 2009). Instead, it enables us to closely monitor interactions between child and environment to determine how the outcome (a form of understanding at some point) is constructed in real time.

In an applied sense, it is of great importance for parents, (science) teachers, and other practitioners to have knowledge about how children grasp varied concepts and how their understanding develops over time. By having this knowledge, they will be able to challenge children in their current level of understanding in order to promote children’s optimal developmental trajectories with regard to cognitive understanding, and by doing so, promote children’s optimal development in a broader sense. Departing from the idea of understanding as a process of change in which the child and the (social and material) context intertwine, the ways and complexity levels at which educators interact with their pupils have an important influence on the development of understanding. With regard to iterativeness, it is important for educators to acknowledge that how understanding changes at one moment in time depends on the understanding at a previous time point. That is, from a dynamic systems perspective, there are no internal operations on representations of knowledge that cause intellectual growth. Understanding organizes on the spot, and gets internalized over time through multiple interactions with the environment. Regarding microgenetical variability, it is important for educators to understand that the highest complexity level on which children operate (e.g., when they learn about scientific concepts) can change rapidly during short-term interactions, not only when the environment or the amount of support visibly changes. Finally, a better understanding of the temporal stream of understanding will help educators to become aware of their own role in the long-term learning process, and may help them to change their actions when necessary or wanted. Students who are engaged in (scientific discovery) learning need adequate support to construct their knowledge (Alfieri, Brooks, Aldrich, & Tenenbaum, 2010).We claim that teachers' awareness of their own role is an important indicator for the quality of their support, which is a crucial factor in improving children's learning (McKinsey, 2007).

We need to work further on completing the empirical picture of possible trajectories of understanding that can emerge in individual children and investigate how these are related to processes on the short-term time scale. This will help us to differentiate components that build up to children's successful and unsuccessful learning trajectories with regard to scientific understanding. This knowledge will also help science educators to teach children to successfully master scientific concepts, as children's understanding of scientific concepts is not always accurate (Grotzer, 2004). When children have more expertise in science, feel confident about this, and enjoy science lessons, this may eventually boost the current number of young people pursuing a scientific academic career. In order to maintain economic growth, people with a scientific education who can ensure continuous technical capability of the highest standards in all fields of expertise are very much needed.

An important next step in the study of the development of children's understanding of scientific concepts as a dynamic system is to try to map individual learning trajectories and build a dynamic simulation model, based on a general theory of action or agent behavior on interacting time scales, and a general theory of mechanisms of change (Steenbeek, 2006; Van Geert, 1994; Van Geert & Steenbeek, 2008). With the help of such a simulation model, the important role of the (science) educator in the emergence of understanding can be unravelled. As a result, such a simulation model will have an important educational value, by making the dynamic principles that play a crucial role in the development of understanding accessible for a broader public of educators. Based on the short-term interaction patterns we see emerge, and the implications this has for the long term, we can eventually construct adaptive teaching programs, lessons and materials for science education, which are better adapted to children's current levels of understanding and how this understanding develops in interaction.

An example of an adaptive educational and assessment (computer) program is Math Garden (Gierasimczuk, Van Der Maas, & Raijmakers, 2012; Van Der Maas, Klinkenberg, & Straatemeier, 2010), an educational computer game with a wide range of sums children that can play at school or at home. Children's responses (the short-term child-computer interactions) are frequently analyzed and reported to their teachers by means of error analyses, individual growth curves, and comparisons between the particular child and his classmates (or the broader population of peers). The program itself uses the child's data by varying the complexity of the sums adaptively, depending on the percentage of right answers, but also on the child's reaction time. Moreover, using the responses and reaction times of all individual children, the items of Math Garden are arranged (and get frequently re-arranged) in terms of complexity. This program shows how multiple short-term interactions provide information about the individual's long-term development and how this information can inform educational practice. These kinds of adaptive teaching and assessment programs translate dynamic principles into concrete materials that help children to develop their understanding in an optimal way.

In conclusion, as Vygotsky (1934/1986) already noted: "To devise successful methods of instructing the schoolchild in systematic knowledge, it is necessary to understand the development of scientific concepts in the child's mind. No less important than this practical aspect of the problem is its theoretical significance for psychological science." (p. 146). We think that by studying the development of children's understanding of scientific concepts using a model based on properties derived from dynamic systems theory and skill theory an important contribution to both this applied and scientific goal is made.

Chapter 4: A Comparison between Young Students with and without Special Needs on their Understanding of Scientific Concepts[21]

This research examines whether young special needs students with emotional/behavioral difficulties (age 3-5, n = 14) reach lower understanding levels than regular students (age 3-5, n = 17) while working on two scientific tasks under a condition of scaffolding (e.g., follow-up questions depending on students' levels of understanding). Understanding was measured microgenetically, per utterance, using a scale related to skill theory. Monte Carlo analyses showed that special needs students gave more wrong and (lowest) level 1 (single sensorimotor set) answers than regular students, and fewer answers on (higher) level 3 (sensorimotor system). However, no difference was found in their mean understanding level, and mean number of answers. Both groups also had a comparable number of answers on the highest levels (level 4 and 5; single representation and representational mapping). These results do not point to substantial differences in scientific understanding between special needs and regular students, as earlier studies using standardized tests have pointed out, and highlight the important role of scaffolding students' understanding. Standardized tests do not seem to indicate the bandwidth of possible scores students show, or give an indication of their optimal scores, whereas a gap exists between student's task performance under conditions of individual performance and performance under a condition of support.

---

[21] This chapter is published as: Van Der Steen, S., Steenbeek, H., Wielinski, J., & Van Geert, P. (2012). A Comparison between Young Students with and without Special Needs on Their Understanding of Scientific Concepts. *Education Research International, 2012*. doi: 10.1155/2012/260403

Numerous studies have shown that students with special needs do not reach the level of academic performance of regular students, since their behavioral or emotional problems interfere with their ability to use their cognitive skills at an optimal level (Scruggs & Mastropieri, 1986; Epstein, Kinder, & Bursuck, 1989; Trout, Nordness, Pierce, & Epstein, 2003). The focus of these studies is primarily on academic achievement, measured with summative assessment methods or standardized tests. However, do we obtain a valid picture of the capabilities, skills and talents of students if we measure these with standardized tests, mostly referring to specific domains such as arithmetic and spelling? Instead, research should also focus on other domains, measures, and conditions of performance in order to identify skills and capabilities that would otherwise be missed. This research aims to contribute to this matter by examining 31 regular and special needs students' understanding of scientific concepts by using a microgenetic design, and an alternative method of measuring understanding. The students (age 3-5) explored two scientific tasks under a condition of optimal scaffolding, meaning that they were encouraged and assisted by an adult while working on the tasks. The aim of this study is to examine whether differences between special needs and regular students will be revealed in the process of building their understanding of scientific concepts, under the guidance of an experienced adult who provides adaptive scaffolding.

4.1 Children's understanding of scientific concepts

Children's understanding of scientific concepts develops from a very young age on (Siegler & Alibali, 2005). Recently, researchers have argued the importance of studying the development of young children's understanding of scientific concepts. Young children's cognitive skills in the domain of science are the foundations of later literacy in this area, and assist children in developing their reasoning about complex relationships (National Research Council, 2005). The degree of understanding scientific concepts reflects the level of scientific thinking skills children can use while working on a problem solving task. Scientific thinking skills can be defined as the skills needed for describing a problem-solving situation, for forming hypotheses, testing hypotheses, and explaining as well as evaluating outcomes (Koslowski, Okagaki, Lorenz, & Umbach, 1989; Kuhn & Franklin, 2006; Wilkening & Sodian, 2005; Zimmerman, 2000; 2007). In the last decades, children's understanding of various scientific concepts has been studied. These studies predominantly focused on specific outcomes of individual learning processes, such as pre- and post-test scores on questionnaires (see chapter 2 and 3 of this dissertation). In order to study students' understanding of scientific concepts, it is important to look at their achievements under a condition of individual performance, but also – even more importantly – under a condition in which they are supported (Zimmerman, 2007).

The concept of scaffolding (Wood, Wood, & Middleton, 1978) comprises the temporary support of a child's learning process by an adult or more capable peer. The support is only temporary, since it is gradually reduced when the child reaches higher levels of competence, and is capable of independent problem-solving (Pressley, Hogan, Wharton-McDonald, & Mistretta, 1996). Scaffolding unfolds dynamically (Van Geert & Steenbeek, 2005b) in that it describes how a particular level of knowledge or skill in a student changes as a result of the scaffolding process, but also how the scaffolding shifts as a result of the change in the student's performance. Teacher and student are engaged in a mutual process, in which the level of the student influences the level of the scaffold (which should be ahead of the first), while the level of the scaffold influences the level of the student. Given this definition of scaffolding as a dynamic mechanism of coupled teaching-learning processes, optimal scaffolding implies a student's optimal understanding as well as optimal teaching at the same time.

Researchers have pointed out the existence of a gap between children's task performance under conditions of individual performance (also referred to as the functional level), and performance under a condition of support (known as the optimal level, see (Fischer & Bidell, 2006). This dichotomy dates back to the work of Vygotsky (1934/1986). The general idea behind this dichotomy is that children do not show a single competence level, but instead vary across a range of possible levels. With help and guidance under a condition of scaffolding, students show an increase in understanding (or an increase in certain capacities), compared to a condition in which they work without receiving support (Fischer & Bidell, 2006). In educational testing, unfortunately, emphasis is put on the functional level, meaning what a student can do alone (an exception are dynamic testing methods, in which repeated testing is alternated with specific forms of feedback). The problem with these standardized methods of individual testing is twofold. First, it does not give us an idea of the student's learning potential, meaning the levels the student can reach with support, which will soon be mastered individually. Second, student's difficulties that interfere with scoring optimally on these tests, such as problems with focusing attention, or understanding the wording of questions, remain unnoticed. Hence, the scores of students with special needs might not only reflect their understanding of a particular concept, but also to a great extent the problems they encounter in an individual testing situation. Under a condition of scaffolding, a teacher (or researcher) can not only attend to the student's needs in a testing situation, but also observe the capabilities of the student when receiving adequate support.

In this study, students were presented with two scientific tasks, while a researcher provided a variety of scaffolding techniques depending on the student's needs. This condition of optimal scaffolding differs from a dynamic testing (or assessment) method, which aims to measure students' learning potential in a particular domain by testing repeatedly and giving feedback after each test (Lidz, 1991; Sternberg & Grigorenko, 2002). Even though dynamic testing methods are used to unravel the process of learning, they are generally standardized, meaning that the questions, the moments of feedback and the types of feedback are defined beforehand. In our condition of optimal scaffolding, we tried to create a naturalistic context somewhat similar to science classes in primary schools. That is, adult and student were constantly talking and working on the task; there were no long-lasting monologues, and they did not take turns in manipulating the task. Moreover, feedback was not given at fixed intervals, but continuously during the interaction, mostly in the form of follow-up questions adapted to the student's answer, such as "Can you explain that?" or "How do you think we should figure that out?"

4.2 Special needs students

The Organization for Economic Co-operation and Development (OECD) defines students with special educational needs as those students who require “additional public and/or private resources to support their education” (OECD, 2005). Since this definition is quite broad, the OECD has defined three cross-national subcategories in which special needs students can be divided: students with disabilities (e.g., sensory, motor or neurological disabilities), students with difficulties (e.g., emotional and/or behavioral difficulties that have a negative effect on learning) and students with disadvantages (e.g., disadvantages due to socio-economic or linguistic factors). Depending on the country and the student’s condition, students with special needs receive extra resources within regular educational facilities, or are placed in special classrooms or schools. In the current research project, we visited special needs students with emotional and/or behavioral difficulties who were enrolled in special educational facilities. Most of these students were officially diagnosed with ADHD or mild forms of autism spectrum disorders (ASD), such as pervasive developmental disorder- not otherwise specified (PDD-NOS). A literature search showed that special needs students with difficulties usually perform below the level of regular students (Mooney, Epstein, Reid, & Nelson, 2003; Reid, Gonzalez, Nordness, Trout, & Epstein, 2004) on academic achievement tests that are usually standardized. This leads to the question whether a condition of optimal scaffolding would yield the same results.

In general, children diagnosed with ADHD show inattention (e.g., difficulty staying focused, often distracted and unorganized), hyperactivity (e.g., motoric restlessness, excessive talking) and impulsivity (e.g., cannot wait for his/her turn, doing before thinking) (American Psychological Association, 2000), which seem to impair their ability to learn (Humphries, 2007). Luo and Li (2003) found that the memory capacity (including short-term and working memory) of children with ADHD was impaired compared to that of typically developing children. Moreover, studies examining the processing level of children and adults with ADHD indicated that they have deficits in higher-level processing (Kalff et al., 2003) and that they use different brain areas to encode complex or low-salient stimuli (Hale, Bookheimer, McGough, Phillips, & McCracken, 2007).

Children diagnosed with ASD are impaired in initiating and sustaining appropriate social interactions (e.g., maintaining relationships, limited social or emotional reciprocity) and communication (e.g., stereotyped use of language, impaired Theory of Mind). In addition, they often show limited and repetitive behavioral patterns (American Psychological Association, 2000). Barnes et al. (2008) stated that ASD students are not able to learn as easily as regular students, since they do not make deliberate use of their (social) environment, even though their implicit learning processes seem to be intact. Studies on higher-level processing of children with ASD showed that they exhibit difficulties when higher-level language processing (the use of meaning and context of a word) is needed to encode information (Noens & Van Berckelaer-Onnes, 2005).

Many special needs students with difficulties (in our sample as well as in the broader population) have a combined diagnosis, such as Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS) with hyperactivity symptoms, or ADHD with symptoms of Oppositional Deviant Disorder (ODD). While there are differences with regard to the specific difficulties that students with different diagnoses encounter in learning situations, they do resemble each other in that special needs students with difficulties generally display significant academic delays across all placements (including all forms of special education and general education; for a meta-analysis, see Reid et al., 2004), which do not seem to improve over time.

4.3 Measuring children’s understanding of scientific concepts

In this study, the levels of understanding were operationalized by using a scale related to the 10 levels of skill theory, developed by Fischer (1980). Skill theory focuses on the complexity and variability of children’s skills, which consist of actions, verbalizations, and thinking abilities, and the way these are constructed (Fischer, 1980; Fischer & Bidell, 2006). One of the most powerful characteristics of skill theory is that it extracts complexity from content, resulting in a content-independent measure of understanding. Because of this content-independent nature, skill theory enables researchers to compare understandings across multiple time points, contexts, persons, and age ranges (Fischer & Bidell, 2006; Fischer & Corrigan, 1981; Fischer & Granott, 1995).

According to Fischer (1980), development in a particular domain goes through 10 levels of skills, hierarchically grouped into three tiers that develop between 3 months and adulthood. The first tier consists of sensorimotor skills: simple connections of perceptions to actions or utterances. For example, the child states that two syringes are attached to one another by a tube. Any statements or actions going beyond the observation of elements, or observable mechanisms, fall in the second and third tier. The second tier constitutes of representational skills, understandings that go beyond current simple perception-action couplings, but are still based on them. That is, the term representation refers to the coordination of several sensorimotor skills at the same time. Within the context of the two connected syringes for example, the child can predict what happens if one of the pistons is pushed in, without literally touching or manipulating the syringe. Nonetheless, what he or she predicts depends not only on the context, but also on the sensorimotor skills mastered before. The third tier consists of abstractions, general rules that also apply to other situations. This would be an explanation about the relationship between pressure and volume inside a syringe (Schwartz & Fischer, 2004). Earlier (basic) skills form the basis of the more advanced skills across all tiers, i.e. they are the building blocks of the higher levels.

Within each tier, sensorimotor, representational or abstract, three levels can be distinguished, each one more complex than the previous one. The first one can be characterized as a single set, (e.g., a single representation, or a single abstraction). The second level is a relation between two of these sets, which is referred to as a mapping. The third level is a system of sets, which is a relation between two mappings, in which each mapping consists of a relation between single sets. After this level, a new tier starts, which is divided in single sets, mappings and systems as well (Fischer & Bidell, 2006).

Fischer and colleagues (Fischer & Bidell, 2006; Granott & Parziale, 2002; Schwartz & Fischer, 2004; Yan & Fischer, 2002) showed that skill theory can not only describe and explain the development of skills on the long term, but also describe the microgenesis of problem solving. When facing a new task or problem, even highly skilled adults go through the same cycles of skills. At the beginning they show skill levels that are mostly sensorimotor, which later build up to more elaborate levels. During a task, people do not go through the skill cycles in an orderly linear fashion. Instead, they repeatedly build up skill levels and regress before they obtain their highest possible level (Yan & Fischer, 2002). This variation between their highest and lowest possible complexity levels is also known as the developmental range. The highest levels within this range (reflecting the student's optimal level) are only reachable when the environment provides sufficient support (Fischer & Bidell, 2006; Yan & Fischer, 2002).

Given that students constantly vary within their developmental range (and given that we used a condition in which scaffolding was provided), it is important to measure understanding repeatedly during a task, and capture the full range of skills students master in this context. Measuring students' understanding in a microgenetical way enables us to closely examine variations in students' understanding which reflect their thinking processes, and prevents us from losing that information if we were measuring understanding at one point in time (Siegler, 2006). We therefore decided to register the skill theory levels of all task-related utterances. By not only looking at students' mean understanding level, but also at the distribution of their understanding levels, a more complete picture of their understanding can be revealed.

4.4 Research questions and hypotheses of this study

This chapter addresses the following questions: First, on average, do the special needs students reach a lower (skill theory) level of understanding than the regular students during the two scientific tasks while they are scaffolded by an adult? Second, if we look at the data from a more microgenetic point of view, does the proportion of the answer levels of special needs students differ from that of the regular students during the scientific tasks?

To see whether the special needs students would benefit from our scaffolding approach, we decided to take a falsification-approach. If the scaffolding would *not* have a positive effect, we would, based on previous literature, expect to find that special needs students' difficulties would impair them in crucial aspects relevant for the tasks, such as staying focused, and being able to process complex information. In line with this, we would expect that (a_1) their mean level of understanding would be lower than that of the regular students, and that (a_2) they would have a lower mean number of correct task-related utterances (answers to questions), but (a_3) a higher mean number of incorrect task-related utterances (wrong answers to questions, i.e., mistakes). This leads to the hypothesis that (b_1) special needs students would have a higher proportion of Level 1 (single sensorimotor set) and Level 2 (sensorimotor mapping) correct answers, which are the lowest skill theory levels. In contrast, regular students were expected (b_2) to answer more questions correctly on the three higher levels: Level 3 (sensorimotor system), Level 4 (single representation) and Level 5 (representational mapping)[22]. However, if special needs students would benefit from the scaffolding condition, we should be able to reject all hypotheses mentioned above, and find no substantial differences between the two groups.

4.5 Method

4.5.1 Participants

The participants consisted of 14 Dutch special needs students with emotional/behavioral difficulties (12 male, 2 female) enrolled in special educational facilities, and 17 Dutch regular students (10 male, 7 female) enrolled in regular educational facilities. Each group consisted of three cohorts recruited at the start of the study: 3-year olds (M_{age} = 40 months, SD = 3.74), 4-year olds (M_{age}

---

[22] We did not include levels higher than 5 into our hypotheses, because the ages associated with the emergence of these levels are above the age range of the students included in our study (see Fischer & Bidell, 2006 for the ages of emergence).

= 54 months, SD = 4.09), and 5-year olds (M_{age} = 65 months, SD = 4.52). Although technically the 3-year old students should be classified as preschoolers, we refer to them as students for the sake of simplicity. The two oldest special needs cohorts (n = 10) attended kindergarten at a special needs primary school, and the youngest special needs cohort (n = 4) attended a special needs day-care center. The two oldest regular cohorts (n = 10) attended kindergarten at a normal primary school, and the youngest regular cohort (n = 7) attended a regular daycare center. Recruitment took place at two schools and daycare centers in the Netherlands. Within these schools and centers, students' parents were asked if their children could participate in a study on scientific reasoning. All students whose parents responded positively were included in the study.

The special needs students included in this study had emotional and/or behavioral difficulties that have a negative impact on their learning. They were officially diagnosed by psychological institutes or pedagogic professionals, most of them with ADHD (about 70% of the special needs students), or a form of ASD (30% of the special needs students). In the Netherlands, an official diagnosis is required to be able to enroll in a special school or educational facility. Given the severity of their problems and their developmental delays, these students were unable to follow the educational program offered at regular schools. The educational program in their special schools takes a slower pace, and focuses more on the students' behavior and basic skills and knowledge. The lower percentage of female special needs students (21.4%) is comparable to that of other mixed-gender studies on special needs students with difficulties. Within the 13 mixed-gender studies included in their meta-analysis, Reid et al. (2004) found percentages of females ranging from 9.3% to 63%, with an average percentage of 22.6%.

4.5.2 Procedure

During each visit, the students explored two scientific tasks individually, guided by a researcher, who was extensively trained into working with an adaptive protocol (see below). The first task involved the scientific concepts air pressure and Boyle’s law, demonstrated by a task in which two syringes were attached to each other through a tube. When the piston of one syringe was pushed in, air travelled through the tube to the other syringe, which piston got pushed out as a consequence. During this task, syringes of different volumes were used. The second task during this visit was about the scientific concepts gravitation, inertia and acceleration, which were demonstrated with a ball-run. Balls of different texture and weight were released at one end of the run, and slid down a path with different colors in order to determine which ball would come the farthest. The concepts of air pressure and gravity/inertia/acceleration were chosen because they provided a domain that was both limited and rich enough to study students’ understanding of scientific concepts. Moreover, given their young age, the students had probably never encountered tasks like this, which meant that a continuous interaction with some form of scaffolding could be established.

To create a condition of optimal scaffolding, but also reach an acceptable level of standardization, an adaptive protocol was constructed. This guaranteed that all students were asked the basic questions that reflected the core building blocks of the scientific concepts incorporated in the task. At the same time, the protocol left enough space for students to show their understanding spontaneously, and for the researcher to provide scaffolding when needed, without prompting the student with answers. This was done by asking follow-up questions related to the student’s earlier answers, encouraging the student to elaborate on an answer, or asking for short explanations.

For each task, the researcher showed the student the material and asked the student for its purpose and functioning at the very beginning. Afterwards—regardless whether the student answered the previous questions right, wrong, or at all—the student was encouraged to explore the material by him/herself. Subsequently, the researcher asked questions about the task’s functioning, as well as the underlying mechanisms, such as “Why does the piston of the other syringe get pushed out when you push the piston of this syringe?” The researcher gave the student time to answer, asked follow-up questions (related to the level of understanding as shown by the student) and encouraged him/her to think about the task and try out his/her ideas using the material. Even though students’ answers were challenged sometimes, the feedback never included statements indicating whether the student was right or wrong. When the student could not give an explanation, the researcher proceeded with another question or subject. Each task took approximately 15 minutes. All interactions were recorded on video.

4.5.3 Coding of verbal understanding

In order to determine students' levels of understanding throughout the tasks, their verbal utterances were coded in four steps using the computer program MediaCoder (Bos & Steenbeek, 2007). The videos were coded in great detail, which enabled us to assign a range of understanding levels during a task. The first step in the coding procedure was the determination of the exact points in time when episodes of utterances started and ended. The second step involved the classification of all utterances of the student into several categories: descriptive, predictive, and explanatory answers/utterances; requests; content-related questions, and other utterances. After this initial classification, meaningful units of the student's coherent utterances were formed in the third step of the coding procedure (units of analysis). This meant that the student's utterances about a single topic were combined. The unit of analysis ended when the next utterance of the student fell into another category, or when the researcher interrupted the student (e.g., by asking another question). However, if the researcher simply encouraged the student to tell more about the same topic, the unit of analysis would not end.

Lastly, the level of understanding per unit was determined by rating each unit on a ten level scale, which follows the model of skill theory (Fischer, 1980). These were the levels ranging from single sensorimotor sets (Level 1) to representational mappings (Level 5). At Level 1, students stated single characteristics of the task, such as "This ball is fast". At Level 2 (sensorimotor mapping), single characteristics were linked and comparisons between task elements were made, such as "This ball rolls faster than the other one". At Level 3 (sensorimotor system) students described aspects of the tasks in terms of causal observational relationships, such as “If I push the piston of this syringe, then the piston of the other one moves”. At Level 4 (single representation), students were able to predict non-observable characteristics and relations by saying e.g., “I think this ball will come further than the other”, or “Air causes the piston of the syringe to move”. Lastly, at Level 5 (representational mapping), students could explain and predict in terms of two causal relationships including an additional step, e.g., “The piston pushes the air, which travels through the tube to the other piston, which then gets pushed out by the air.” Next to these five levels, an answer could also be classified as a “mistake” when it was simply wrong, irrelevant, or when the student indicated that he or she did not know the answer to a question.

Videos were coded by two independent raters using a standardized coding book. For each round of coding (categories, units, and understanding levels), raters went through a training of coding three 15-minute video fragments and compared their codings with those of an expert-rater —the researcher who constructed the codebook. Initial differences between the raters and the expert-rater were solved through discussion. The codings of the third fragment were compared to the codings of the expert-rater and a percentage of agreement was calculated. The percentages of agreement on the third fragment were: categories: 93% ($p < .01$), units: 94% ($p < .01$), and level of understanding: 92% ($p < .01$). The advantage of reporting simple percentages is that these are intuitively clear measures of agreement. Nevertheless, percentages provide no indication to what extent they depend on chance, which is why a *p*-value (within brackets) was added (Van Geert & Van Dijk, 2003). The *p*-values were calculated using a Monte Carlo procedure; for a description of this statistical procedure see section 2.4.

4.5.4 Data analysis

After coding special needs and regular students’ answers during both tasks, the frequencies for each level of understanding were determined. The mean level of understanding, the number of mistakes and answers, as well as the proportion of answers on each level were compared. For these comparisons, we used Monte Carlo permutation tests (Todman & Dugard, 2001), which have great explanatory value in the case of small or skewed samples and result in reliable p-values, since they do not assume any underlying distribution, or a minimum sample size (Van Geert, Steenbeek, & Kunnen, 2012). Given our small sample size and skewed distribution of data, an ANOVA design (with accompanying assumptions) would decrease statistical power (Baguley, 2012). The Monte Carlo procedure estimates the probability that a certain difference between two groups is caused by chance alone. This is done by drawing a number of random samples from the original data (for this study 5000 random samples were drawn for each test), and determining how often the observed, or a bigger difference occurs in these random samples (positive cases). This number of positive cases is divided by the number of random samples in order to produce a *p*-value for the tested difference, comprising the probability that the observed difference occurs in the distribution of 5000 random samples of the data. If the probability that this occurs is small, we can conclude that the observed difference is not merely caused by chance and thus that it is a legitimate difference.

Since we compared a number of differences between conditions and variables, we have decided to discuss only the interesting differences, which we defined as all differences for which the *p*-value was equal to or smaller than .1 (which would support the hypotheses, and literature on academic differences between regular and special needs students), and all differences that were contrary to our expectations (i.e., those results that would make us reject the hypotheses that the two groups differ, which would possibly indicate the positive effect of scaffolding). The effect sizes of these differences (*d*) were calculated by dividing the difference in means by the standard deviation of the youngest age group (in case of within-group differences), or the standard deviation of the regular students (in case of between-group differences). These standard deviations were chosen because they were usually the biggest, and hence yielded the most conservative measure of the effect size.

4.6 Results

4.6.1 Mean levels of understanding

Before testing our hypotheses, we first looked at the within-group differences in mean understanding level to see if similar patterns would evolve within each group. The results of the analysis are displayed in Table 3 and Figure 5. For the regular students, a significant difference in mean level of understanding was found between the 4-year olds and the 5-year olds, and between the 3- and 5-year olds ($p < .01$ for both differences, $d = 1.81$ and $d = 2.24$ respectively). For the special needs students, a very similar pattern emerged: The 3-year olds and the 4-year olds differed significantly in their mean level of understanding from the 5-year olds ($p < .05$; $d = .97$ and $d = 1.33$ respectively).

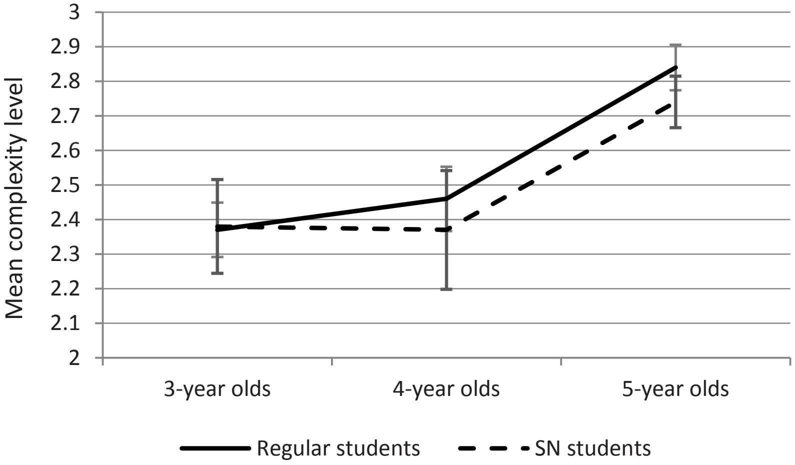
4.6.1.1 Hypothesis a1: lower mean level of understanding for special needs students

Table 3 also shows the overall mean understanding level of the regular and special needs students. Contrary to the hypothesis (a_1), the regular group reached only a slightly higher mean level of understanding ($M = 2.54$, $SD = .27$) compared to the special needs group ($M = 2.50$, $SD = .32$). This difference was not statistically significant ($p = .36$). When looking at the differences in means for each age group, the results were similar. Even though the special needs students had lower mean understanding levels in the two oldest age groups, and a comparable level of understanding in the youngest age group (see Figure 5), the differences with the regular students were too small to be statistically significant. We can therefore reject hypothesis a_1, and conclude that there are no significant differences in mean level of understanding, both in the group as a whole and across all age groups.

Table 3: Mean, standard deviation, minimum and maximum level of understanding per group of students (regular and special needs) and cohort.

Group	Age	N	Mean	SD	Min	Max
Regular	**All**	17	2.54	0.27	0	5
	3	7	2.37	0.21	0	4
	4	5	2.46	0.21	0	4
	5	5	2.84	0.15	0	5
Special needs	**All**	14	2.50	0.32	0	5
	3	4	2.38	0.27	0	4
	4	5	2.37	0.38	0	5
	5	5	2.74	0.17	0	5

Figure 5: Mean understanding level (Y-axis) displayed by age (X-axis) for each group. Error bars refer to the standard error of the means.



4.6.2 Mean number of correct answers and mean number of mistakes

Subsequently, the mean numbers of answers and mistakes were analyzed (see Table 4 and Figure 6). Again, the within-group differences were explored first to see if we could detect similar patterns in the two groups. Within the regular group, the mean number of answers first decreased with age and then slightly increased, albeit not statistically significant. However, there were some significant differences regarding the mean number of mistakes for the regular group, that is, the difference between the 3- and 4-year olds (*p* = .05, *d* = .77), and the difference between the 3- and 5-year olds (*p* < .05, *d* = .91). The special needs group showed a non-significant decrease in the mean number of answers between the 3- and the 4-year olds, and a significant increase between the 4- and the 5-year olds (*p* < .05, *d* = 1.26). Their mean number of mistakes, however, differed only slightly, and none of the differences between the age groups were statistically significant.

4.6.2.1 Hypothesis a_2 and a_3: special needs students have a lower mean number of correct answers, and a higher mean number of mistakes

The mean number of answers did not differ significantly (*p* = .42) between the two groups, which was in contrast with the hypothesis (a_2) that the mean number of answers would be lower in the special needs group. The mean number of mistakes, however, was significantly higher for the special needs students (*p* < .01, *d* = .91), which supported hypothesis a_3. This was also found when we corrected for the number of answers, i.e. when we compared the mistakes proportional to the total number of answers, which yielded a higher proportion (0.46) for the special needs students compared to the proportion (0.32) for the regular students (*p* < .01, *d* = 1.45).

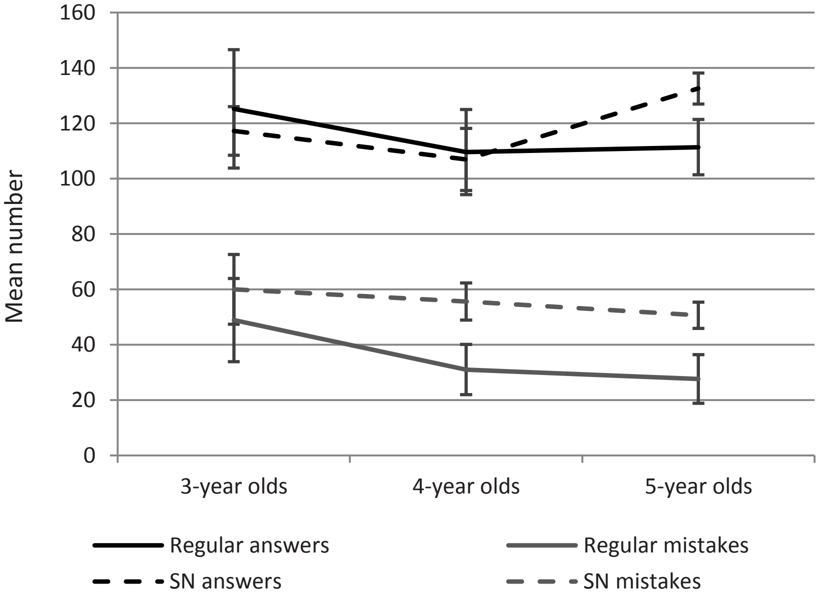
When looking at the different age groups, the 3-year old regular students did not differ significantly from the 3-year old special needs students in terms of their mean number of answers, but also not in their mean number of mistakes. However, the ratio wrong/total number of answers of the 3-year old special needs students (0.5) was significantly higher than that of the 3-year old regular students (0.39), *p* < .05, *d* = 1.19. The mean number of answers of the 4-year old regular students also did not differ from that of the special needs students. That said, their mean number of mistakes was significantly higher (*p* = .01, *d* = 2.09). This was also the case when the ratio wrong/total number of answers was compared: The ratio of the 4-year old special needs students was significantly higher (0.52) than that of the regular students (0.29), *p* < .01, *d* = 3.47. Lastly, the 5-year old regular and special needs students differed significantly with respect to both their mean number of answers and their mean number of mistakes (*p* = .05, *d* = .95 and *p* < .01, *d* = 1.83 respectively). Note that the 5-year old special needs students answered more questions than the regular students (*M* = 132.6, *SD* = 19.55 vs. *M* = 111.4, *SD* = 22.35), contrary to hypothesis a_2. Nevertheless, they also made more mistakes (*M* = 50.6, *SD* = 10.46 vs. *M* = 27.6, *SD* = 12.58), and the ratio wrong/total number of answers was higher for the special needs students than for the regular students (0.38 and 0.24 respectively, *p* < .01, *d* = 1.95), which was in line with what was expected (a_3).

To summarize, we found no evidence for the hypothesis that special needs students have a lower mean number of correct answers across all age groups, so we can reject hypothesis a_2. On the other hand, we did find evidence for the hypothesis that special needs students have a higher mean number of mistakes, and cannot reject hypothesis a_3.

Table 4: Means and standard deviations of the number of answers and mistakes per group of students (regular and special needs), per cohort.

Group	Age	N	Mean answers	SD	Mean mistakes	SD
Regular	**All**	17	116.59	40.95	37.35	19.44
	3	7	125.29	56.59	48.86	23.31
	4	5	109.60	34.33	31.00	11.79
	5	5	111.40	22.35	27.60	12.58
Special needs	**All**	14	119.07	24.11	55.07	16.30
	3	4	117.25	30.08	60.00	25.13
	4	5	107.00	20.35	55.60	15.08
	5	5	132.60	19.55	50.60	10.64

Figure 6: Mean numbers (Y-axis) of answers and mistakes by age (X-axis) for each group. Error bars refer to the standard error of the means.



4.6.3 The proportion of the (skill theory) answer levels

In order to answer whether the distribution of the answer levels of special needs students differed from that of the regular students, the number of answers were counted for each level and divided by the total number of answers within each (age) group. To test the differences between the groups, the mean proportions were used (see Table 5).

4.6.3.1 Hypothesis b_1: special needs students have a higher proportion of correct answers on Level 1 and 2

When we compared the regular students with the special needs students across all age groups (see the left upper graph of Figure 7), special needs students had a significantly higher proportion of Level 1 answers ($p < .01$, $d = 2.0$), as was hypothesized. However, the regular group had more answers on Level 2 ($p = .05$, $d = .55$), which was in contrast with hypothesis b_1. When looking at the 3-year olds, a similar difference between the groups emerged for Level 1 ($p < .05$, $d =$

1.06). The 4-year old special needs students also had a higher proportion of Level 1 answers compared to their regular peers (*p* < .01, *d* = 4.4), and given the large effect size, this seems to be a considerable difference. The 4-year old regular students had a higher mean proportion of level 2 answers than the special needs students (*p* = .05, *d* = 1.06), which was in contrast with hypothesis b_1. For the 5-year old students, the difference in the proportion of Level 1 answers between the special needs students and the regular students was significant (*p* < .01, *d* = 3.3). In sum, special needs students had indeed a higher proportion of correct Level 1 answers across all age groups, which was in line with hypothesis b_1. For Level 2 answers, however, the overall group of regular students had a significantly higher proportion, as well as the 4-year olds. For the 3- and 5-year olds, no significant difference in the proportion of Level 2 answers was found. Hence, the results for the proportion of Level 2 answers are not in line with hypothesis b_1.

4.6.3.2 Hypothesis b_2: Regular students have a higher proportion of correct answers on level 3, 4, and 5

In the overall group, the regular students had a higher proportion of Level 3 answers (*p* = .06, *d* = .49), which supported hypothesis b_2. On Level 4, however, the special needs students outperformed the regular students, which was unexpected (*p* = .1, *d* = .49). No significant difference between the groups was found for Level 5 (*p* = .31). When looking at the separate age groups, the 3-year olds showed a similar difference between regular and special needs students on Level 3 (*p* < 0.05, *d* = .86). For this age group, the difference on Level 4 was also noteworthy, since the 3-year old special needs students had a higher proportion of answers on this level than the regular students (*p* = .07, *d* = 1.04). For the 4- and 5-year olds, the differences between the groups on Level 3, 4 and 5 were too small to be statistically significant. To conclude, the only evidence in line with hypothesis b_2 was found for the proportion of Level 3 answers in the overall group and for the 3-year olds. All other differences were not in line with hypothesis b_2.

Figure 7 shows the proportion of answer levels, both for the groups as a whole and for the separate age groups. Despite some small differences (mostly on Level 1 and 2), the shape of the graphs of the two groups is strikingly similar, with peaks at Level 2 and 4, low values at Level 1 and 5, and a dip at Level 3. In the graph of the 3-year olds (right upper graph), the dip at Level 3 is clearly lower for the special needs students than for the regular students, whereas the rest of the proportions seem to be similar. The graphs for the 4- and 5- year old students (lower two graphs) look even more similar. The difference in the proportion of Level 3 answers is smaller for these age groups, and the proportions of answers on Level 4 and 5 seem to be equal.

Table 5: Proportions of correct answers per level of understanding (the number of correct answers for each level divided by the total number of correct answers of each (age) group).

Group	Age	N	1	2	3	4	5
Regular	**All**	17	0.04	0.60	0.14	0.21	0.005
	3	7	0.05	0.67	0.13	0.17	0.00
	4	5	0.04	0.65	0.12	0.19	0.00
	5	5	0.02	0.47	0.17	0.32	0.01
Special needs	**All**	14	0.13	0.51	0.09	0.26	0.007
	3	4	0.11	0.64	0.02	0.23	0.00
	4	5	0.18	0.51	0.09	0.21	0.01
	5	5	0.09	0.42	0.15	0.33	0.01

4.7 Discussion

The aim of this research was to examine whether differences between 3- to 5-year old special needs and regular students would emerge in the process of building their understanding of scientific concepts while working on two scientific tasks: one about air pressure and Boyle’s law, and one about gravity, inertia and acceleration, under a condition of optimal scaffolding in a natural setting.

4.7.1 Overview of our findings

With regard to the mean level of understanding, the hypotheses that special needs students' mean level of understanding would be lower (a_1), and that they would have a lower mean number of answers (a_2) must be rejected. The hypothesis that special needs students would make more mistakes (a_3) was the only hypothesis that was mostly supported by our data. That is, the overall special needs group made more mistakes than the regular group. This was also the case when the 4- and 5-year old special needs and regular students were compared. For the 3-year olds, no difference was found when absolute measures were compared; however, the ratio wrong/total answers was significantly higher for the 3-year old special needs students.

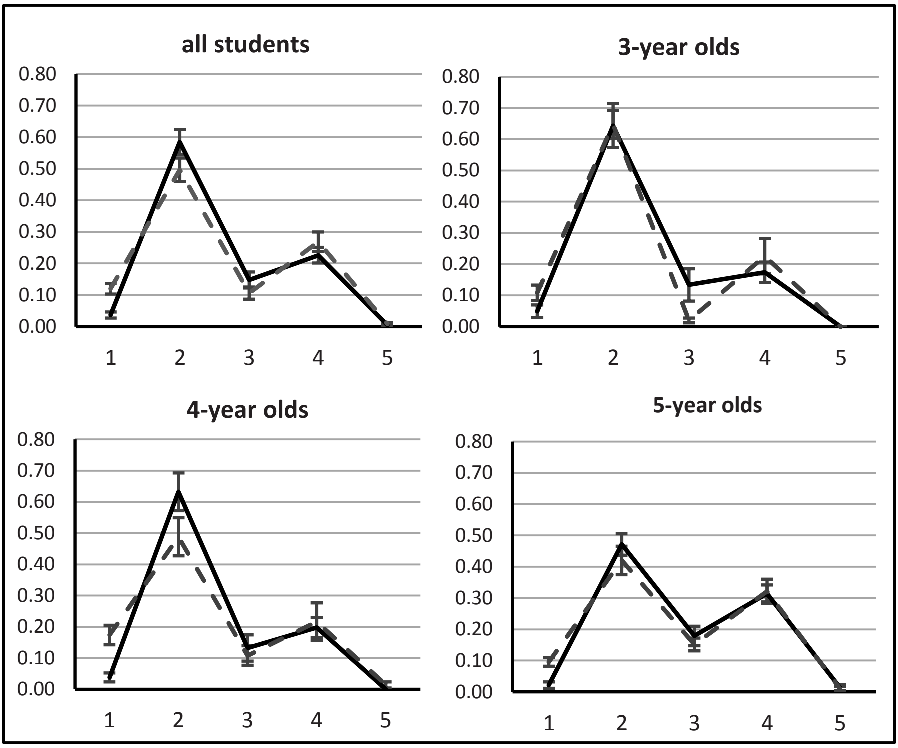
In line with hypothesis b_1, special needs students had a higher proportion of Level 1 (single sensorimotor set) answers compared to the regular group. Contrary to this hypothesis, however, the regular students outperformed the special needs students on Level 2 (sensorimotor mapping) in the overall group and most age groups. In addition, the regular students had indeed a higher proportion of Level 3 (sensorimotor system) answers (hypothesis b_2), but this was mostly caused by the difference between the 3-year old special needs and regular students. On Level 4 and 5 (single representation and representational mapping), the groups scored roughly equal; which was not in line with hypothesis b_2. In general, most findings were in contrast with the hypotheses and previous research.

4.7.2 The positive effects of optimal scaffolding conditions

In the last years, studies showed that students with special needs are not learning the required basic academic skills, and perform below the level of regular students across several domains. Most of these studies focused on math and reading skills (Epstein et al., 1989; Scruggs & Mastropieri, 1986; Trout et al., 2003), measured with standardized tests (Reid et al., 2004), although some have focused on scientific thinking (Mooney et al., 2003). The outcomes of these studies are in contrast with the performance of special needs students under our optimal scaffolding condition. In fact, our results are even in contrast with the standardized test scores of the special needs students included in this study, on which they performed below the regular students. Most Dutch schools take part in a national assessment program (Cito) and regularly evaluate their students' progress on several subjects, such as math and language skills. We collected the regular and special needs students' test scores on their first Cito language and math tests, administered in kindergarten. On both tests, students could get a score from 1 (E, lowest score) to 5 (A, highest score). We obtained data for 28 of our students; the data of three special needs students were not available, because they had not yet been tested. Taking the mean score of these two tests, our regular students had a score of 4.4 on average, whereas the special needs students had a score of 3.68. Using a Monte Carlo test, we found this difference to be statistically significant ($p < .05$), with an effect size (d) of .67. This means that at this time, the regular students performed two-third of a standard deviation better on these two academic tests compared to the special needs students in our sample.

The question arises whether the skills and performances examined with standardized tests are similar to those in this research. Standardized tests do not indicate the bandwidth of possible scores children show, or give an indication of their optimal scores, whereas researchers have pointed out the existence of a gap between children's task performance under conditions of individual performance and performance under a condition of support (Fischer & Bidell, 2006). In other words, the context in which one assesses students' capabilities influences the results to a great extent. This context can be a difference in terms of measurement setting or presentation of tasks (standardized versus scaffolding), but also in terms of the type and phrasing of questions. In a study of Ayoub and colleagues (2006) maltreated children (42 months old) were not able to re-tell stories involving nice interactions as accurately as non-maltreated children. However, both groups showed roughly the same scores when asked to re-tell stories involving mean interactions. The authors conclude that maltreated children are not cognitively impaired in the traditional sense, but instead have learned to focus more on negative aspects, which can be an adaptive response to threat.

Figure 7: Proportion (Y-axis) of answer levels 1-5 (X-axis) for all students, and for the 3 age groups.



Note. Regular students are displayed by the solid line, special needs students by the dashed line. Error bars refer to the standard error of the proportions.

The current research shows that special needs students with behavioral difficulties perform on the same level as regular students on tasks requiring scientific thinking and reasoning, if they are guided by an adult who uses appropriate scaffolding techniques to respond to the student's emotional and cognitive needs. On the other hand, standardized tests in math and language seem to be too demanding. Cooper, Baum and Neu (2004) indicated that standardized test scores are not always appropriate to measure problem-solving skills of special needs students. In their study on problem-solving, which included experiential science materials, a mentoring component, and assessment of students' scientific products instead of their test scores, the problem-solving skills of special needs students were comparable to those of regular students. This study also seems to indicate that special needs students' scientific problem solving skills (and their understanding, which reflects the level of these skills) are more advanced in conditions in which they receive adaptive support from the environment. Their individual performance, in the literature mostly measured by standardized tests (and in the case of our sample by math and language tests), might not accurately reflect the special needs students' full potential.

4.7.3 Standardized tests vs conditions of scaffolding: what do they measure?

For many special needs students, the validity of (standardized) tests depends on the accessibility of test items and tasks. As an example, a dyslexic student's score on a standardized math test might not only reflect the student's math skills, but also the ability to read the test items and instructions (Almond et al., 2010). Hence, standardized tests do not only measure the constructs they claim, and students' test scores might reflect some construct-irrelevant noise. The students included in our study were not print-disabled, but had other difficulties, and formal testing situations might be unable to meet their individual needs. These needs might well be met in a scaffolding condition, in which the researcher continuously draws the student's attention, changes the wording of questions if necessary, and uses follow-up questions to get a complete picture of the student's understanding, or challenges an earlier given answer. Moreover, the hands-on tasks used in this study enabled the students to try out their ideas, and, if necessary, change their explanations of the mechanisms at work.

Scaffolding does not mean that students get so much help that they simply surpass their own level of performance, nor does it mean that students are prompted with answers. Instead, scaffolding sets a context in which students can access the upper section of their range of possible scores. Although scaffolding is seldom used in summative assessment methods, Almond and colleagues (2010) note that scaffolding provides students with supports that help them to answer questions at their individual level, which allows us to better measure students' knowledge and skills. Under a condition of scaffolding, teachers can see what students do know about a particular item, instead of simply marking their answer as wrong or incomplete. This study shows that when children are in a situation in which scaffolding is applied frequently, differences between special needs and regular children almost disappear. We therefore advise teachers in special educational settings to use a wide range of adaptive scaffolding techniques (follow-up questions, encouragement, instructions, and feedback) during their lessons. In doing so, teachers can pay particular attention to the mistakes special needs students make (which they made more in this study compared to the regular students), and encourage them to elaborate on the correct parts of their thinking. By carefully watching students' responses in the classroom, the difficulties of special needs students can be detected and further addressed by using scaffolding techniques. For example, the 3-year old special needs students in this study had difficulties in expressing causal relationships, that is, they had significantly less answers on Level 3 (sensorimotor system). These young students might benefit from more scaffolding directed towards this type of reasoning.

New initiatives show that scaffolding conditions are not as far from formal testing situations as one would imagine. Research suggests that applying universal design principles can improve testing of special needs students with difficulties, by providing alternative forms of instructions (e.g., not only text, but also graphs or pictures, or videos), alternative forms of expression (e.g., not only writing down answers, but also drawing or using graphic organizers), and alternative forms of engagement (e.g., choosing a topic for a test on reading comprehension) (Almond et al., 2010; Dolan & Hall, 2001).

4.7.4 Suggestions for future research

The number of special needs students is growing (U.S. Public Health Service, 1999) and therefore it becomes more and more important to assess not only their disabilities, but also their capabilities both in the academic context and beyond. Identifying their strengths and providing help to make use of these strengths could support students in developing a more positive self-concept and self-efficacy, which they often lack due to failure experiences in the academic context (Cooper et al., 2004). Future research could investigate what characteristics of students' environment (materials, tasks, and interactions with adults or peers) support the development of their (scientific thinking) skills, in order to advise teachers, parents and therapists regarding the optimal adjustment of academic contexts to students' individual needs. In addition, the microgenetic approach we used (coding per utterance), yielded a continuous measurement of students' understanding, and showed that understanding shifts regularly between levels over time (see also Granott & Parziale, 2002). Measuring understanding using aggregated data of single tests might prevent us from detecting these variations in students' understanding and could possibly lead to inaccurate measures. Further research should both investigate the benefits of scaffolding for special needs students in more detail, as well as the variations in their academic achievements over time. The results of these studies can then be used to optimize standardized tests, so that special needs students can make optimal use of these situations.

Chapter 5: A Process Approach to Children's Understanding of Scientific Concepts: A Longitudinal Case Study[23]

In order to optimally study changes in the complexity of understanding, microgenetic measures are needed, and a coupling of these to longer-term measures. We focus on the interaction dynamics between a 4-year old boy and a researcher while they work on tasks about air pressure in three subsequent sessions. The complexity of the utterances of the researcher (questions) and the boy (answers) was measured using a skill theory-based scale. Over the course of the three sessions, an increase in the boy's number of right answers occurred, and the frequencies of the complexity levels shifted. With regard to the interaction dynamics, the boy initiated significantly more simultaneous in- and decreases in complexity level over time, whereas the researcher initiated less. At the same time, the boy showed an increase in his mean understanding level. Therefore, on the longer term, learning may be related to taking more responsibility for generating lines of thought.

---

[23] This chapter is published as: Van Der Steen, S., Steenbeek, H., Van Dijk, M., & Van Geert, P. (2014). A Process Approach to Children's Understanding of Scientific Concepts: A Longitudinal Case Study. *Learning and Individual Differences, 30, 8- 91*. doi: 10.1016/j.lindif.2013.12.004

As developmental psychologists studying educational settings, we are interested in how children learn during a task, how the person-context dynamics shape this learning process, and how understanding develops over time. While studies taking measures over longer time periods (over the course of months) reveal general developmental trends of learning, they provide little insight into the short-term mechanisms of change (e.g., during a lesson). In contrast, microgenetic studies— studies of processes that unfold during a short time span—provide important insights into how actual change in learning occurs, and how the link between teaching and learning is formed (Granott & Parziale, 2002; Siegler, 2006). Given the cyclical causal relationship between the short- and long-term timespan of learning, we see an additional necessity to couple these microgenetic processes to mechanisms on the long-term time scale of development. That is, one should describe and explain how short-term learning events influence long-term development and vice versa (Granott, 2002; Steenbeek & Van Geert, 2013).

This chapter focuses on three interactions between a 4-year old boy and a researcher while working on scientific tasks about air pressure. Using time-serial microgenetic data of the boy's reasoning, we explore fluctuations in his understanding, and examine how the child-researcher dynamics shape this learning process, as well as how these dynamics change over time during two subsequent visits. We will use tools inspired by the (dynamic systems) complexity approach (Van Geert, 2008; Van Geert & Steenbeek, 2005a), and dynamic skill theory (Fischer & Bidell, 2006). First, however, we define the concept of scientific understanding from a macro- and microdevelopmental perspective.

5.1 Defining scientific understanding

Multiple studies on scientific learning show that students develop various concepts about scientific phenomena during their (early) school years (Linn & Eylon, 2006; Zimmerman, 2005). These scientific concepts can be defined as ideas about phenomena in the domains of chemistry, physics, and biology (Baartman & Gravemeier, 2011; OECD, 2003). Children use these concepts in combination with inquiry skills (tool use, analogical reasoning, manipulation of variables) to reason scientifically (Zimmerman, 2005). From a macro-developmental perspective, children's understanding of various scientific concepts has been studied, such as gravity (Novak, 2005; Palmer, 2001; Sharp & Sharp, 2007), air pressure (Séré, 1986; She, 2002; Tytler, 1998), electricity (Chiu & Lin, 2005; Shipstone, 1984; Zacharia, 2007), chemistry (Garnett, Garnett, & Hackling, 1995; Taber, 2001), gear wheels (Dixon & Bangert, 2002; Lehrer & Schauble, 1998), and the universe (Albanese, Neves, & Vicentini, 1997; Dunlop, 2000). These studies have given an idea of global developmental trends across cohorts by focusing on specific outcomes of the learning process, such as scores on knowledge tests (e.g., before versus after an intervention), as well as the number, categories and accuracy of children's concepts. Microgenetic studies, on the other hand, have investigated the developmental trajectories of scientific concepts in detail, mostly over a short period of time, such as during a task or science lesson. In particular, these studies have examined the short-term path (changes in conceptual understanding), rate of change, breadth (whether acquired skills generalize to other tasks), source (what contextual factors influence learning progress), and intra-individual variability in strategies, actions, or thinking (Siegler, 2006).

Despite the progress microgenetic studies have made in unraveling the characteristics of learning and development (see for example Goldin-Meadow, Alibali, & Church, 1993; Granott, Fischer, & Parziale, 2002; Kuhn, 2002), more processes of change and mechanisms facilitating change in learning situations have yet to be identified (Flynn & Siegler, 2007). Researchers studying complex systems can offer a rich set of tools to analyze microdevelopmental patterns and link these to general developmental trends. The properties associated with complex systems, such as the soft-assembly of multiple components, and the recursive nature of development, may help to interpret and explain patterns found in microgenetic studies (Thelen & Corbetta, 2002). Of particular importance is the connection of several microgenetically coded learning interactions to provide a picture of learning over a longer term. Focusing on two dynamic properties (intra-individual variability and person-context dynamics), this chapter shows how learning interactions can be microgenetically analyzed to examine how a boy's understanding is constructed during one science task, and how this relates to his learning over the course of two subsequent tasks.

5.2 Using dynamic skill theory to take microgenetic measures of understanding

In many microgenetic studies, researchers choose to code and analyze video-data, to prevent disrupting the unfolding process as much as possible. Skill theory (Fischer, 1980; Fischer & Bidell, 2006) includes a scale that provides a useful tool for coding such data. Skill theory focuses on the complexity and variability of children's skills, which consist of actions and thinking abilities, embodied in verbal and non-verbal behavior. Used in a microgenetical way, the scale enables researchers to extract the complexity (of e.g., utterances) from content, which makes it possible to compare understanding across multiple time points, contexts, and persons (Parziale & Fischer, 1998). Learning is defined as building collections of skills, which are hierarchically ordered in 10 levels grouped into three tiers. The first tier consists of sensorimotor skills: simple connections of perceptions to actions or utterances. The second tier consists of representational skills; these are understandings that go beyond current perception-action couplings. The third and final tier consists of abstractions, which are general nonconcrete rules that also apply to other situations (Schwartz & Fischer, 2004). Within each tier, three levels can be distinguished: single sets, mappings (a relation between two single sets), and systems (a relation between two mappings).

Although skills are hierarchically ordered, learning does not entail a linear progression through the levels. Instead, it is driven by many microdevelopmental steps forward and backward (Van Geert & Fischer, 2009). Even during a single task, people vary constantly within a bandwidth between their highest and lowest possible complexity levels, also known as the developmental range. The highest levels of this range are only reachable when the environment provides sufficient support (Fischer & Bidell, 2006; Yan & Fischer, 2002). Skill theory thus accounts not only for intra-individual variability in learning, which has been of growing interest in developmental psychology (e.g., Thelen & Smith, 1994; Van Geert & Van Dijk, 2002; Van Orden, Holden, & Turvey, 2003), but also for the dynamics between person and environment (skills emerge in specific contexts, and differ depending on the support offered), which have been emphasized by many (Fogel & Garvey, 2007; Thelen & Smith, 1994; Van Geert & Fischer, 2009). These two properties will be illustrated below.

5.3 Structured intra-individual variability

Intra-individual variability is crucial to understand developmental phenomena (Siegler, 1994), given that development is by definition a real-time iterative process within individuals (Van Orden et al., 2003). Information about fluctuations in people's actions or thinking can thus help to describe and understand cognitive change (Siegler, 2007). From a dynamic point of view, variability is seen as a system-specific property (Steenbeek, Jansen, & Van Geert, 2012; Van Geert & Steenbeek, 2005a), meaning that the complexity of children's understanding fluctuates, even within short periods of time. Researchers studying microdevelopment found that people particularly show an increase in variability (in e.g., actions or strategies) before transitioning to a more advanced strategy (Bassano & Van Geert, 2007; Van Dijk & Van Geert, 2007), or a higher level of understanding during a task (Jansen & Van Der Maas, 2001; Yan & Fischer, 2002). Such an increase in variability is needed to explore new strategies, and ultimately, to anchor a more advanced strategy for a longer period of time (Shrager & Siegler, 1998; Siegler, 1996; 2007; cf. Simonton, 2011). The structure of intra-individual variability can be analyzed not only statistically (see Van Orden et al., 2003; Kello et al., 2010), but also functionally by describing which levels are observed and how these relate to the ongoing interaction with the context. That is, one can investigate how fluctuations in the complexity of children's understanding relate to complexity fluctuations of the interaction partner, or in other words, focus on the child-context dynamics during a learning process.

5.4 Child-context dynamics

Most studies do not specifically address the continuous intertwining of person and context (Richardson, Marsh, & Schmidt, 2010), but instead view the environment as “system input” (p. 5), that is, an independent variable that influences the person, or interacts with certain characteristics of the person. Viewed dynamically, however, behavior is a “dynamic, self-organized consequence of the physical laws and informational constraints that are mutually structured across mind, body, and environment” (Richardson et al., 2010, p.8). The child's understanding of a concept, is the child's continuously changing cognitive state, as he or she reacts to the current dynamic interaction (Van Geert, 2011b).

Since understanding is a self-organizing process assembled of three interactive components (boy, researcher, and task), certain patterns in the interplay of the complexity of questions and answers might emerge. For example, fluctuations (i.e., intra-individual variability) in understanding may be influenced by not only the ongoing interaction with the context, but also the other way around (see Chapter 3). That is, increasing complexity of the researcher’s questions about the task may be related to increasing complexity of the boy’s answers. In addition, one would also expect the researcher to adjust the complexity of her questions to the complexity of the boy’s previous answers (see the literature on scaffolding, e.g., Van Geert & Steenbeek, 2005b). Over time this process might change. When the boy and researcher are more adapted to one another, and when the boy has a (partial) understanding of the procedure and concepts asked during a task, he might take more initiative in directing the conversation. As a metaphor, one could picture a dance. The researcher can only lead if the boy follows, and vice versa. A switch in this lead might indicate that the boy has at least a partial understanding of the task, and that he feels confident to demonstrate this. It is, however, important to keep in mind that there is always a mutual coupling between dance partners. That is, there is no simple notion of unidirectional causality, since the coordinated movements emerge as a result of joint activity.

5.5 A case study—Research questions and hypotheses

This case study is focused on a typically developing 4-year old boy, who worked together with a researcher on a task about air pressure during three visits. Skill theory was used to code the cognitive complexity of the boy's answers and the researcher's questions. The central research question was: How can we characterize the interaction dynamics—the boy's and the researcher's fluctuations in complexity levels—during one session, and how does this change over the course of three sessions? To answer this question, we first adopted a systematic exploratory approach to examine the fluctuations in the boy's understanding levels during one session, and explored similarities and differences in the two subsequent sessions. Second, we specifically focused on the child-researcher dynamics during the three sessions. Our first hypothesis was that the fluctuations in complexity levels of the boy's answers and the researcher's questions would be related during session 1. To be more specific, we expected a covariation within a temporal range, in which changes in the researcher's complexity levels would be followed by similar changes in the boy's understanding levels. Over the course of the next two sessions, we hypothesized that the interplay between the boy and researcher would shift from oscillatory movements mostly initiated by the researcher to a situation in which these were also initiated by the boy (hypothesis 2).

5.6 Method

5.6.1 Participant information

For this study, a typically developing boy (4 years and 8 months old) was chosen as a case. He attended kindergarten at a primary school in the north of the Netherlands, and his scores on early arithmetic and language tests (measured in the Cito national ongoing assessment program) fell within the range of the 25% highest-scoring 4-year olds.

5.6.2 Material

The boy worked on a hands-on air pressure task, while the researcher asked about the functioning of the task, and provided adaptive scaffolding. During the first visit (session 1), the task involved two syringes of the same volume attached by a tube. When the piston of one syringe was pushed in, air traveled through the tube and pressed the piston of the other syringe out (see Tytler, 1998 for a similar task). At the end of the task, a longer tube was connected to the syringes, and differences in the functioning of the task were explored. The two subsequent tasks involved connecting syringes of different volumes (session 2 – administered 3 months after session 1), and using syringes to lift a miniature version of an elevator (session 3 – administered 3 months after session 2). The tasks of sessions 2 and 3 required extra manipulations or more elaborate thinking to explain their functioning.

5.6.3 Procedure

For each task, the researcher showed the material, and asked the boy for its purpose and functioning. After this, the boy was encouraged to explore the material while the researcher asked about the task's functioning and underlying mechanisms. These questions depended on what emerged from the interaction. To create an optimal learning situation, the researcher asked follow-up questions related to the boy's level of understanding, and encouraged him to elaborate on his answers. However, the researcher was not allowed to prompt the boy with answers. Each session took approximately 10 minutes and was recorded on video.

5.6.4 Coding of verbal expressions

In order to determine the boy's level of understanding throughout the task, the verbal expressions were coded in four steps using the computer program MediaCoder (Bos & Steenbeek, 2006). First, we started with the determination of the exact points in time when utterances of both the boy and researcher started and ended. The second step involved the classification of these verbal utterances into categories. As a third step, meaningful units of the student's coherent expressions were formed (units of analysis). In the fourth and final step, the complexity of the boy's answers within a unit, and the complexity of the researcher's questions were determined using a scale based on skill theory.

In order to make sure that the codings were reliable, a standardized codebook was used (see appendix C for a description). For each round of coding, three raters went through a training of coding 3 video fragments of 15 minutes and compared their codings with those of an expert-rater (who constructed the codebook and training). The codings of the third fragment were compared to the codings of the expert-rater and a percentage of agreement was calculated. On average, these were: categories: 87% (range 81-93; $p < .01$), units: 93% (range 89-96; $p < .01$), level of understanding: 90% (range 83-95; $p < .01$), and complexity of the researcher's questions: 84% (83-86%; $p < .01$).[24]

5.6.5 Data analysis

For our exploratory analysis of the fluctuations in the boy's understanding, we plotted a time series of the (skill theory) complexity levels measured in the boy's utterances during session 1. Using a Monte Carlo permutation test (Todman & Dugard, 2001), we compared the fluctuations in two sections of the interaction by taking the mean absolute difference between each complexity level and the next. To analyze how the boy's complexity levels were organized, we calculated the frequencies and used Monte Carlo tests to see whether these changed significantly over the three sessions.

The first hypothesis (fluctuations in the boy's and researcher's complexity levels are related during session 1) was answered by plotting a Loess smoothing of the two time series of complexity levels during session 1. To investigate the interaction, the smoothed graphs were normalized and compared. Hypothesis 2 (the interplay between the boy and researcher shifts from oscillatory movements mostly initiated by the researcher to a situation in which these are also initiated

---

[24] Percentages are intuitively clear measures of agreement, but provide no indication to what extent they depend on chance, which is why a p-value was added (cf. Van Geert & Van Dijk, 2003) using a Monte Carlo procedure.

by the boy) was answered by plotting the smoothed normalized graphs of the next visits. Using Monte Carlo permutation tests, the numbers of simultaneous in- and decreases in complexity levels during the three sessions were compared; as well as differences in initiations (who started the in- or decrease before the other followed). In addition, we repeatedly calculated the covariance while shifting the researcher's graph alongside the graph of the boy, to see how many time steps we had to shift the graph in order to get the highest covariance (overlap). For more information about the statistical procedures we refer to appendix D.

5.7 Results

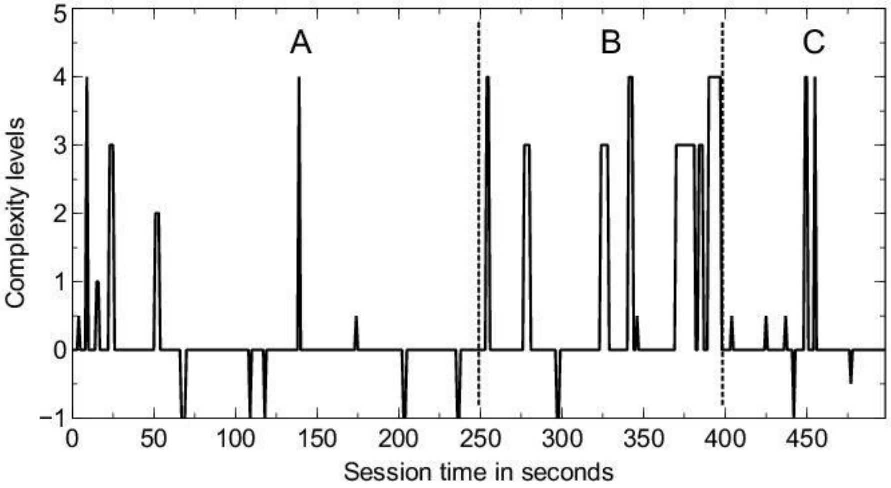
5.7.1 Microgenetical variability—exploratory analyses

During session 1, the boy more often responded with false and correct yes/no answers to close ended-questions compared to the other levels. In addition, answers on level 3 and 4 were more often observed, whereas he almost never answered on level 1 and 2 (see Table 6). Most frequencies, however, were not significantly higher or lower than expected based on the total number of answers in that session, apart from the low frequencies of level 1 and 2 answers. From a visual inspection of Figure 8 it seems that the complexity of the utterances first decreased (section A, the first half of the interaction). During the first part of the second half of the interaction (section B), higher complexity levels occurred. Toward the end of the interaction (section C) the complexity levels decreased again, although at the end of the interaction two higher complexity levels can be observed.

Although a visual inspection of Figure 8 seems to indicate that more fluctuations are present during the second half of the interaction (sections B and C), a Monte Carlo analysis revealed no significant difference (n = 30; p = .34) between section A and the other two sections. This result could be influenced by the researcher's questions. For example, after an answer of the boy, the researcher could ask about another task-related topic on a lower complexity level. In that case, the difference in complexity between the boy's current and previous answers might reflect a difference in accurate reactions to the questions asked. Nonetheless, when only taking into account answers about the same topic (answers to questions asked on the same complexity level), no difference between section A and the other two sections was found (*n* = 11; *p* = .72).

A next step was to explore how the boy's fluctuations in understanding were organized over a longer period. Table 6 shows the frequencies of the complexity levels during the three sessions. Using a Monte Carlo procedure, we tested whether the frequencies of the levels changed over time. The total number of right answers increased (*p* < .1) from session 1 (23) to session 2 (37). This was the same for the number of answers on level 2 (from 1 to 13, *p* < .01), which significantly decreased again in session 3 (4 answers, *p* < .05). The third session yielded a higher number of level 1 answers, as opposed to session 1 (1 vs. 7 answers, *p* < .05), but not as opposed to session 2. Lastly, there was an increase in level 3 answers during session 3 (from 2 to 9 answers, *p* < .1).

Figure 8. Time-serial Illustration of the complexity levels measured in the boy's answers during session 1.



Note. Utterances classified as incorrect are depicted as -1, and right answers to close-ended questions are marked as 0.5.

Summarizing these exploratory analyses focusing on the boy's variability in complexity levels, we described how his understanding fluctuated during session 1, showing no difference in variability between the first half and second half of the interaction. In addition, we focused on how the frequencies of complexity levels changed over time, showing that the boy's level 2 answers increased during the second session, while his level 3 answers decreased. In session 3, this was exactly the other way around. Given this information on how the fluctuations were organized during the sessions, the question may be asked whether and how the boy's fluctuations were related to the researcher's questions during the sessions.

Table 6: Change of Frequencies over Time

	n	false	correct	correct (close-ended)	level 1	level 2	level 3	level 4
Session 1	31	8	23**	8	1**	1**	6	7
Session 2	48	11	37***	12*	3**	13**	2**	7
Session 3	47	10	37***	7	7	4*	9	10
Difference 2-1	17**	3	14*	4	2	12***	-4	0
Difference 3-1	16*	2	14*	-1	6**	3	3	3
Difference 3-2	-1	-1	0	-5	4	-9**	7**	3
Total	126	29	97	27	11	18	17	24

Note. * *p* < .1, ** *p* < .05, and *** *p* <.01 for session frequencies indicate whether the frequency was significantly higher (in bold) or lower (in italics) than expected based on the total number of answers in that session. For the differences between the sessions, *p*-values indicate whether an increase (in bold) or decrease (in italics) of a frequency was significant.

5.7.2 Intertwining of person and context

5.7.2.1 Hypothesis 1: in- and decreases of complexity levels of researcher and child are related during session 1

To capture the general trends in both the boy's and researcher's complexity levels, we smoothed their complexity levels during session 1 (see Figure 9) using a Loess technique. In addition, for both the researcher and the boy, a linear trend line was fitted with a very slight positive slope, indicating a slight increase in complexity level over the task. Throughout the session, the researcher's graph was positioned above the boy's graph. The question remains, however, if changes in the researcher's complexity level were directly related to those of the boy. Figure 10—session 1 displays a re-scaled normalized Loess curve, in which the peaks in the complexity levels of the researcher mostly precede the peaks in the boy's level (peaks A-D). The offset between the researcher's and the child's peak is the biggest for peak B. Right before peak C the symmetry is restored and the boy's curve follows the peaks and drops of the researcher's curve again. As of yet, we can conclude that the in- and decreases of the interaction partners seemed related during session 1, albeit in a nonlinear fashion (see for example the dissymmetry at B). The researcher seemed to take the lead in this session, that is, most of her peaks (A, C and D) in complexity level precede the boy's peaks in complexity level (see also the covariance analyses for hypothesis 2).

5.7.2.2 Hypothesis 2: The interplay between boy and researcher shifts over time

Figure 10 also displays two graphs with normalized Loess curves of the second and third visits. In Figure 10—session 2, the first three peaks are more or less simultaneous. After that, the boy's level goes down during a relatively lengthy episode (point A), while that of the researcher shows two peaks, and then goes down. At the end, the symmetry seems to be restored. While during the first visit the boy generally followed the researcher in in- and decreases in complexity level, the offset of some of his peaks now also starts earlier. In the third session, the first peak of the researcher coincides with a bumpy peak of the boy on the line of increase. This is followed by a peak in the boy's understanding, right before a second peak of the researcher (point A). The researcher's and the boy's peaks in the middle occur in an asynchronous way (point B). Toward the end, the two peaks coincide again (point C).

We counted the numbers of simultaneous in- and decreases in the smoothed normalized data series, and used a Monte Carlo procedure to determine who first started to in- or decrease before the other joined (see Table 7). Over time, the boy initiated more simultaneous in- and decreases, whereas the researcher initiated less. The overall *p*-value for the proportional in/decrease of the boy and researcher across all three sessions was .002. While there was a significant difference between the initiations of the researcher and the boy during session 1 ($p < .01$), this difference disappeared in the next sessions. For session 3, a significant increase in the proportion of the boy's initiations occurred ($p < .01$ compared to session 1, and $p < .1$ compared to session 2). At the same time, the boy showed an increase in his mean understanding level ($p < .1$).

The last column in Table 7 displays how many seconds the researcher's graph has to shift to produce the most overlap with the boy's graph (highest covariance level). It shows that the researcher's graph has to shift 15 points upward in session 1 to form the most overlap (i.e., she shows in/decreases in complexity level before the boy does this 15 seconds later). In session 2 the highest covariance can be found if we leave the graphs in exactly the same position as they are. In session 3 the most overlap can be found when we move the graph of the researcher 15 seconds steps back, meaning that the boy is now 15 seconds ahead.

Closing this section on the person-context dynamics, we can indeed observe covariation within a temporal range. During session 1, the peaks of the researcher usually preceded those of the boy. The researcher was about 15 seconds ahead and initiated significantly more simultaneous in- and decreases (hypothesis 1; see Table 7). In the two subsequent sessions, the interplay between the boy and researcher shifted to a situation in which the boy took more initiative (hypothesis

2). He showed a significant increase in his initiations, and was about 15 seconds ahead of the researcher in the third session.

Table 7: Numbers and Proportions of Simultaneous Increases or Decreases Started by Researcher and Boy.

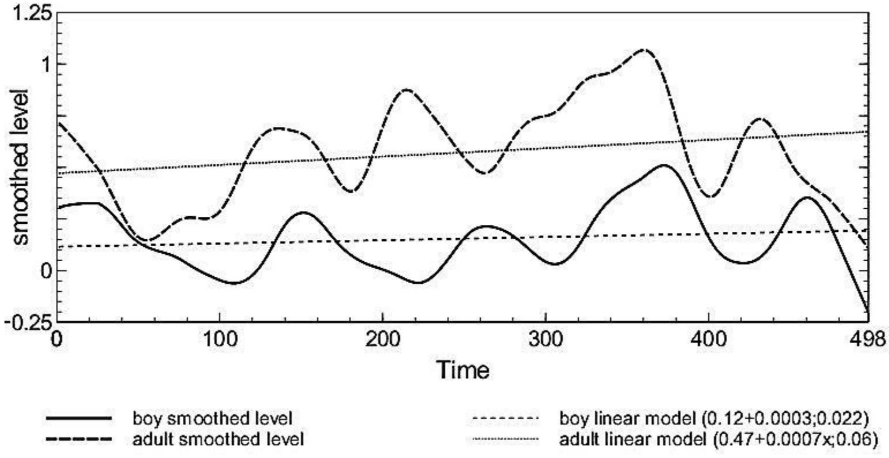
	Nr of simult. in/decreases	Prop. started by researcher	Prop. started by boy	Boy's mean underst. level	Optimal shift in Data points
Session 1	8	.88**	.12**	1.71	15
Session 2	13	.69	.31	1.43	0
Session 3	12	.42	.58	1.82	-15
Difference 2-1	5	-.19	.19	-.28	
Difference 3-1	4	-.46**	.46**	.11	
Difference 3-2	-1	-.27*	.27*	.39*	
Total	33	.64	.36	1.65	

Note. * $p < .1$, ** $p < .01$ for sessions indicate the significance level of the difference between researcher and boy. The *p*-values for session differences indicate the significance levels of within-person in/decreases over 2 sessions. The overall *p*-value for the proportional in/decrease of respectively the boy and researcher across all three sessions is .002 (not displayed in the table). The delay column displays how many seconds the researcher's graph has to shift to get the most overlap (the highest covariance) with the boy's graph.

5.8 Discussion

With this study, we showed how the development of understanding can be studied using a microgenetic method. In terms of the in-depth characteristics of learning distinguished by Siegler and colleagues (Siegler, 2006; Flynn & Siegler, 2007), this case study investigated the path (changes in understanding), and variability of understanding scientific tasks about air pressure. We focused not only on the understanding process of the boy, but also on the complexity of the questions asked by the interaction partner (the researcher), and how these related to one another over time. The complexity of questions and answers was measured on the same scale, thereby facilitating the comparison.

Figure 9. Loess smoothing of the time-serial graph of the complexity levels measured in the boy's answers (black line) and the researcher's questions (dashed line) during session 1.



Note. The Y-axis depicts the smoothed instead of the raw complexity levels.

The results show that the boy had multiple fluctuations in his understanding, which were not clustered in either the first or the second half of the interaction. Over the course of the three sessions, an increase in the number of (right) answers occurred, and the frequencies of the complexity levels shifted: The boy's level 2 answers increased during the second session, while his level 3 answers decreased. In session 3, this was exactly the other way around. These preliminary analyses gave us an idea of how the boy's complexity levels were organized over time, as well as how his understanding fluctuated during the first session.

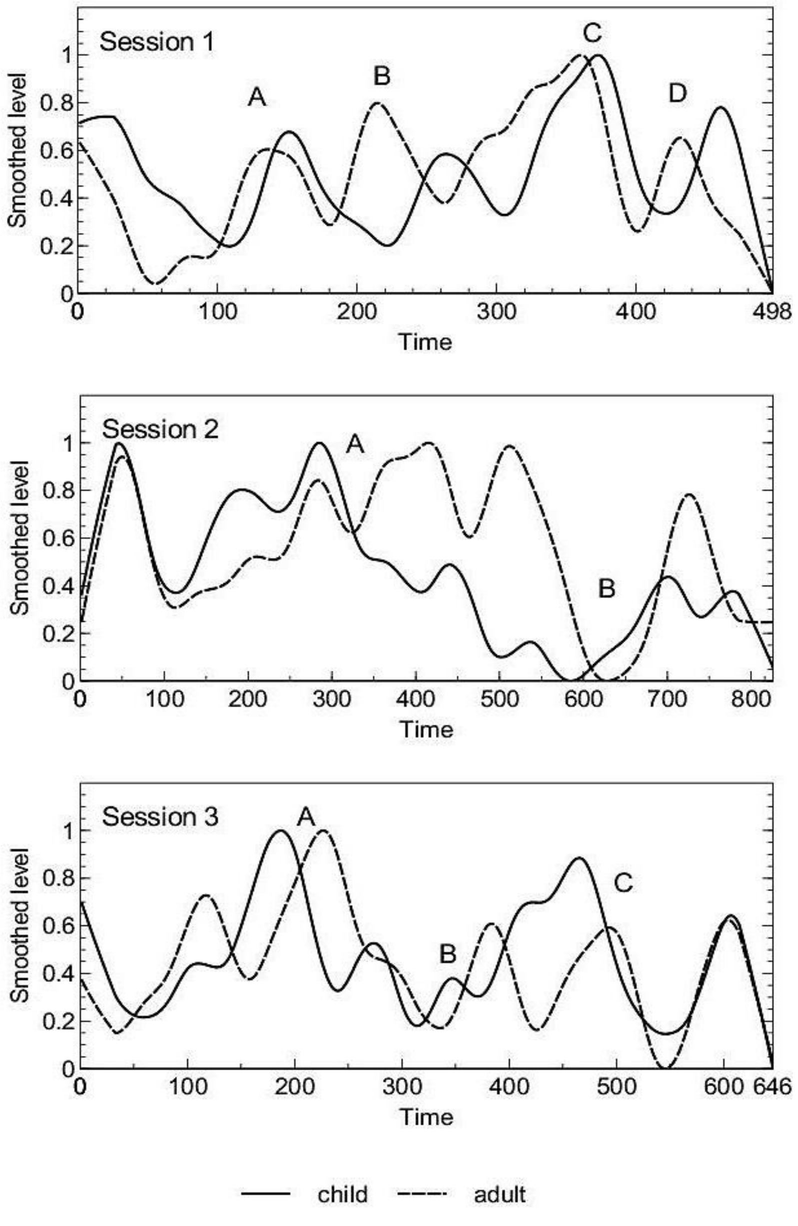
The underlying dynamics of the variability in understanding levels becomes visible when looking at the interplay between the boy and the researcher. In the first session, the boy usually followed the researcher's in- and decreases in complexity level. Over time, the boy initiated significantly more simultaneous in- and decreases, whereas the researcher initiated less. During session 3, a significant increase in the proportion of the boy's initiations occurred, and a significant increase in his mean understanding level at the same time. While the covariance of session 1 was highest when we moved the researcher's graph 15 seconds forward, the covariance of session 3 was highest when we moved the researcher's graph 15 seconds back, indicating that the boy was about 15 seconds ahead during session 3.

As it occurs in this study, learning is not only just answering questions, but also taking more responsibility for generating lines of thought, which is important for developing critical thinking skills (Bailin, 2002). Parallels can be drawn with studies focusing on self-regulated learning. Self-regulated learning is a process by which students are metacognitively, motivationally and behaviorally active in selecting and structuring their own learning process, which enhances their academic success (Zimmerman, 1990). In this study, a comparable result was found, since taking more initiative co-occurred with a significant increase in the boy's mean understanding level in the third session.

While the boy takes more initiative in the mutual in- and decreases in complexity level, the researcher takes less over time. This behavior (taking a step back) strongly resembles the concept of fading in the definition of scaffolding. Scaffolding is an intrinsically dynamic process in which a teacher provides adaptive support while the student carries out a learning task, and gradually reduces this support (fading) as the student progresses (Van De Pol, Volman, & Beishuizen, 2010; Van Geert & Steenbeek, 2005b). Fading of support provided by a teacher, and the accompanying increase in initiative (or self-regulation) from the end of the student seem to occur automatically, in a smooth fashion, suggesting that this mutual process emerges from the interaction dynamics, and not from the need or preferences of one interaction partner.

This study suggests some important indications for both research and educational practice. First, using tests to determine students' understanding at one point in time by aggregating test scores into one average score, might not accurately reflect their capacities in that domain, as students may fluctuate tremendously in the complexity of their reasoning. Microgenetic studies, on the other hand, enable a close examination of variability in students' understanding, which is a reflection of their learning process (Siegler, 2006). Having an indication of the score bandwidth of a student may help teachers to tune in at various levels, in order to shift their bandwidth gradually upward.

Figure 10. Normalized loess curves of the complexity levels measured in the boy's answers (black line) and the researcher's questions (dashed line) of three sessions.



The ways and complexity levels at which tests, research materials, or teachers interact with students have an important influence on their learning that is not always immediately clear. A visual inspection of Figure 9 shows that the researcher usually asks questions on a higher complexity level. However, after applying a smoothing technique and normalizing the graphs, we see a clear connection between in- and decreases in complexity level of the two interaction partners. That is, when the researcher increases the complexity of her questions, chances are the boy shows an increase in complexity as well, albeit on a lower level. A microgenetic approach thus enables researchers and teachers to look at interaction patterns like these, which influence the learning process tremendously, and might otherwise be missed.

Situated in an educational setting, this study contributes to the current need for classroom studies to back up findings from laboratory studies (Zimmerman, 2005). In this way, it can help to support changes in educational science programs, and help defining scientific concepts that have not yet been clearly defined.[25] Although this chapter is a case study, combining the data from multiple longitudinal case studies can answer important developmental questions. For example, do children who take more initiative learn faster than children who do not? Does variability in understanding contribute to long-term development? Once we have answered these questions and know more about learning processes in real time, we can fully implement the findings in classroom settings.

---

[25] Energy is one of the scientific concepts that require a more clear and accurate definition (see for example Coelho, 2009).

Chapter 6: How to Characterize the Development of Children's Understanding of Scientific Concepts: A Longitudinal Microgenetic Study[26]

Using a longitudinal study on children's understanding of scientific concepts, we compare the relative importance of general (e.g., standardized math and language learning achievement scores) and microgenetic measures (real-time interaction patterns) to characterize the development of scientific understanding over 1.5 years. A researcher worked five times with 31 children (3-5 years old, from regular and special primary schools) on scientific tasks about air pressure and gravity. The researcher's scaffolding behavior and the child's understanding were coded per utterance. Furthermore, children's standardized learning achievement scores and information on their home environment were obtained. A cluster analysis distinguished three developmental trajectories, which could best be predicted by interactions between the child and his/her proximal environment. In the discussion we consider the use of context-dependent versus context-independent measures when assessing children's understanding.

---

[26] This chapter is submitted as: Van Der Steen, S., Steenbeek, H., Van Dijk, M., & Van Geert, P. (submitted). How to Characterize the Development of Children's Understanding of Scientific Concepts: A Longitudinal Microgenetic Study.

Children’s academic achievements are frequently evaluated and the outcomes highly influence their prospective school admissions, further career, and future position in society (Haladyna & Downing, 2004; OECD, 2004). In the last decade, children’s achievements in STEM areas (i.e., science, technology, engineering and mathematics) have received considerable attention, because these “permeate nearly every facet of modern life, and also hold the key to meeting many of humanity’s most pressing current and future challenges” (National Research Council, 2011, p.1). Yet, European and American organizations warn that both the number of students choosing STEM fields for further study, as well as the STEM knowledge of the general student population are insufficient to guarantee future technological advancement (National Research Council, 2011; Roberts, 2002; Van Langen & Dekkers, 2005).

Given the personal and societal importance of children’s academic performance in STEM fields, social scientists are searching for its underlying predictive factors. These studies can be broadly divided in two lines of research: First, macro-studies of general (isolated) factors or characteristics that influence children’s academic achievement (e.g., their working memory or gender), and second, a smaller number of microgenetic studies investigating the real-time interaction dynamics between children and their proximal environment that affect their academic performance (e.g., real-time measures of how individual children respond to teaching or educational materials). Although both approaches have improved our understanding of children’s academic achievements, their methods have never been coupled in a single study to compare the relative strength of the associations between the general factors versus the microgenetic interaction dynamics to predict the long-term developmental patterns of children’s academic achievement. In the current chapter, we compare the relative strength of these associations using a longitudinal microgenetic study on young children’s performance in STEM fields. We focus on their conceptual understanding of scientific phenomena, that is, we study changes in their understanding while repeatedly working on scientific tasks in which the scientific concepts gravity and air pressure are embedded.

Scientific concepts can be defined as ideas about phenomena in the domains of chemistry, physics, and biology (Baartman & Gravemeier, 2011; OECD, 2003). During their school years, children develop several of these concepts that become increasingly more complex or veridical (Linn & Eylon, 2006; Zimmerman, 2005), for example about gravity (Novak, 2005; Palmer, 2001; Sharp & Sharp, 2007) and air pressure (Tytler, 1998; She, 2002). In combination with inquiry skills such as tool use and formulating hypotheses, children use these concepts for scientific reasoning (Zimmerman, 2005), which is required in academic STEM areas.

6.1 Macro-studies of general factors that influence children's academic achievement

The majority of the studies focusing on academic achievement in STEM areas are of the first type we distinguished, and can thus be characterized as (macro-) studies searching for predictors of academic achievement that are independent of immediate child-environment interactions. While some of these predictors can be characterized as psychological, such as learning style, personality, and working memory (e.g., Bull, Espy, & Wiebe, 2008), developmental psychology has also been concerned with the role of demographic variables, such as age, gender, and type of schooling. Indeed, since the early studies of developmental psychology, (neo-) Piagetian theories of cognitive development associate a child's increasing age with better developmental outcomes (Piaget, 1947/2001; see Fischer & Bidell, 2006 for a more recent account). Studies investigating children's understanding of scientific concepts for example, suggest that older children reason at a more advanced level than younger children when presented with density tasks (e.g., Rappolt-Schlichtmann, Tenenbaum, Koepke, & Fischer, 2007).

In several other studies, gender has been considered as a potential predictor of academic achievement in STEM areas (Baker, 2002). In the fields of science and technology, gender differences showing a male advantage are often reported, although some studies report an absence of these gender differences (Brotman & Moore, 2002). In a study on the development of astronomical science concepts for example, Bryce and Blown (2007) found 3 studies reporting no gender differences and 7 with gender differences favoring boys. Despite the inconclusive evidence for the existence of a substantial gender gap, the differences found in several studies have stimulated researchers to further investigate the role of gender in science (see Brotman & Moore, 2002 for an extensive literature review).

A smaller number of studies have considered school type (reflecting students' characteristics) as a factor related to academic performance, for example by identifying differences between children enrolled in regular and special schools for e.g., children with externalizing and internalizing behavioral disorders. Earlier research has consistently found negative academic outcomes for these special needs students (see Reid, Gonzalez, Nordness, Trout, & Epstein, 2004 for a meta-analysis) that do not seem to improve over time (Lane, Barton-Arwood, Nelson, & Wehby, 2008), and sometimes result in a 3-year lag compared to children from regular schools (Steenbeek, Jansen, & Van Geert, 2012). For example, a study by Lane et al. (2008) revealed that elementary students in special education score well below the 25th percentile on math and other academic subjects. The emotional and/or behavioral problems these children have seem to interfere with their ability to perform well on tests of learning achievement.

Although proximal contextual influences are usually acknowledged, most macro-studies focusing on general predictors of academic performance do not specifically assess the continuous intertwining of person and context (Richardson, Marsh, & Schmidt, 2010). That is, the child's performance is measured in a standardized environment, usually at one specific moment in time. The same applies to the tests used within schools to measure children's academic performance. Although the child-context interaction is important in all areas of education, and especially in STEM areas, it is generally assumed, both in research and practice, that the relative context-independence of standardized tests provides an objective measure of children's skills that has high predictive value across contexts.

6.2 Microgenetic studies investigating real-time interaction dynamics

In contrast, recent studies using a (dynamic systems) complexity approach to investigate person-context interactions suggest that understanding is formed from continuous child-context interactions, and cannot be assessed independently (Fischer & Bidell, 2006; Granott & Parziale, 2002; for one of the first accounts, see Thelen & Smith, 1994). This means that the social (e.g., the teacher) and/or the material environment (e.g., materials used in class) play an active part in the formation of (e.g.,) scientific understanding. The child's current understanding of a scientific concept is the child's continuously changing cognitive state, as he or she picks up and reacts to whatever goes on in the current dynamic interaction (Van Geert, 2011b). Hence, according to this view it would be virtually impossible to assess or predict performance independently and across all contexts (Van Gelder, 1998). One could, however, perceive much of the current cognitive state by carefully watching the verbalizations and actions that reflect the child's thinking during his/her interactions with the proximal environment (see chapter 3). To conclude, although macro-studies generally focus on child characteristics that are independent of the immediate child-context interaction, the microgenetic approach assumes an ongoing person-context construction of skills.

Studies applying a microgenetic approach observe and analyze learning processes that unfold during a short time span (Granott & Parziale, 2002). These studies have investigated the process of forming scientific concepts during the interaction with the material environment, e.g., by studying changes in children's understanding while building miniature bridges (Parziale, 2002), or solving balance scale problems (Philips & Tolmie, 2007). In addition, several microgenetic studies have been conducted to investigate the real-time transactional dynamics between the child and his/her proximal social environment, for example how child-teacher interactions contribute to learning processes by focusing on the teacher's scaffolding (Steenbeek & Van Geert, 2013). This construct describes how a student's level of knowledge changes as a result of the temporary support of a child's learning process by a more capable person, for example by giving instructions, asking questions, and providing assistance and encouragement (Van de Pol, Volman, & Beishuizen, 2010). Teacher and student are engaged in a mutual process of co-construction, in which the level of the student influences the scaffold, and vice versa (Renninger & Granott, 2005). Microgenetic studies have shown that scaffolding improves scientific understanding, particularly when aimed at a level that is somewhat higher than that of the student (Granott, 2005), and while preserving opportunities for children to take the initiative. In contrast, less optimal scaffolding, such as frequent mismatches between the child and the teacher's responses or too many self-iterations of the teacher, are associated with negative academic outcomes (Steenbeek et al., 2012).

6.3 Research questions and hypotheses

Despite the insights derived from macro studies focusing on general factors and microgenetic studies focusing on interaction patterns, these methods have never been compared in one study. Such a combination study would allow us to investigate the relative predictive value of both the macro ("context-independent") factors and the micro ("context-dependent") processes contributing to the long-term development of understanding scientific concepts. Such a study requires in depth measures of child-context interaction patterns over a longer period of time, while also obtaining demographic information and general measures that may contribute to the development of children's understanding of scientific concepts.

The research question of this chapter was twofold: First, how can we characterize the developmental patterns of children's understanding of scientific concepts over the course of 1.5 years, in terms of their shape? To study this, a researcher worked 5 times with individual children (3-5 years old, from both regular and special primary schools) on scientific tasks about air pressure and gravity. During these visits, the understanding levels of the children were coded per utterance and for each child the proportion of higher understanding levels was calculated per visit to provide a picture of children's performance. We examined how many distinct developmental patterns we could distinguish with regard to children's understanding over the course of 5 visits.

Our second question was: How can we characterize the distinct developmental patterns in terms of their associations with a number of microgenetic and macro predictive factors (see Table 8 for all measures)? These factors were either derived from the interactions during the tasks, questionnaires filled out by the parents of the children, or from children's learning achievement test scores obtained from their schools. We distinguished four types of measures: The so-called *interaction variables* were based on microgenetic coding of the child's and researcher's behavior during the visits (e.g., the proportion of child's initiatives during a visit, or the proportion of the researcher's follow-up questions). The macro factors we distinguished could be divided into *demographic variables* (e.g., the child's gender or age), and *school variables* (e.g., school type, or standardized learning achievement scores). Lastly, we distinguished *home environment variables*, comprising both macro measures of children's characteristics as indicated by their parents (e.g., child's motor/language development as rated by their parents), as well as micro measures of children's interactions at home (e.g., whether the parents encourage playing with educational toys, whether the family talks about school experiences).

Given that this is the first study combining general factors and microgenetic measures over a longer period of time, we did not have clear a priori hypotheses about the shape of the developmental trajectories (first research question) and the variables with the highest associations with these developmental trajectories (second research question). Instead, we adopted a thorough "bottom-up" strategy. If the general "context-independent" variables could best predict children's developmental patterns over time, we would observe the developmental trajectories to differ with regard to the proportion of 1) boys and girls, 2) the different age groups, 3) children from special and regular schools, and 4) children with low and high standardized learning achievement test scores. However, in concordance with the view of understanding as a complex process depending on person-context interactions, we would best predict the developmental trajectories by means of the measures derived from the interaction between the child and the researcher.

6.4 Method

6.4.1 Participants

The participants consisted of 31 Dutch primary school students, of which 17 (10 boys, 7 girls) were enrolled in regular primary schools, and 14 (12 boys, 2 girls)[27] in schools for special education. Each group consisted of three cohorts recruited at the start of the study: 3-year olds (*n* = 11, M_{age} = 40 months, *SD* = 3.7), 4-year olds (*n* = 9, M_{age} = 53 months, *SD* = 3.7), and 5-year olds (*n* = 11, M_{age} = 65 months, *SD* = 4.7). The two oldest cohorts attended kindergarten at a regular or special primary school, whereas the youngest cohort attended a regular or special daycare center at the beginning of the study. Within these schools and centers, parents were asked if their children could participate in a longitudinal study on the development of scientific concepts. All children whose parents provided a written consent were included in the study.

The special needs' student population in the Netherlands is quite diverse, that is, both children with internalizing (autism spectrum disorders, anxiety disorders) as well as externalizing problems (attention deficit hyperactivity disorder, oppositional defiant disorder) are enrolled in special schools and daycare centers, and are taught in the same classrooms. In our study, 64% of the special needs students had externalizing problems, and 36% internalizing problems. Most children in the regular schools had no emotional or behavioral problems, apart from one 3-year old boy with internalizing problems.

---

[27] The lower percentage of female students in the special education group is in concordance with the current trend of an overall lower percentage of female students in special schools (Reid et al., 2004).

6.4.2 *Materials*

6.4.2.1 Tasks

Each visit (5 in total) the children worked on two scientific tasks about air pressure and gravity. To simulate a series of science lessons, and to prevent simple testing effects, each subsequent task required extra manipulations or more elaborate thinking to explain the mechanism and embedded scientific concepts of the task. The air pressure task sequence started with a toy frog that could jump by means of squeezing a balloon attached to its inflatable legs; the task of the second visit involved the connection of two syringes of the same volume by a tube, whose pistons moved in opposite directions when manipulating; in the third visit syringes with different volumes were connected to explore differences in the functioning of the previous task; in the fourth visit syringes were used to operate a miniature version of an elevator, and in the fifth visit the carrying capacity of this syringe elevator was explored using air and water as content.

The gravity tasks started with an open marble track in which marbles fell down at the end of each trail to the next; in the second visit a different marble track with a stair-like mechanism to lift the marbles was explored; the task of the third visit was a ball-run, in which balls of different textures and weights were released and slid down a path to determine which would come the farthest; in the fourth visit, the effects of three ball runs with a different surface (wood, a smooth and a coarse carpet) were compared, and in the fifth visit children had to construct a working marble track with a looping, varying the distance and height of the track.

6.4.2.2 Questionnaires

The parents of the children were asked to fill out questionnaires after each visit. These questionnaires were constructed for this study and focused on demographic variables, as well as home environment characteristics that may influence STEM skills. The demographic questions (only asked in the first questionnaire) concentrated on the child's age, gender and diagnosis, the family composition, nationality and educational level of the parents, and the physical, motor, language and emotional development of the child so far. The remaining questions were included in all five questionnaires and mainly focused on the child's interactions with the home environment, such as: 1) parents' perception of the child's problem solving skills, curiosity and exploratory behavior; 2) the child's play behavior at home, the use of educational toys, cooperative play with parents and sports, and 3) parental stimulation in the form of household chores, stimulation of early arithmetic skills such as counting and recognizing numbers, and stimulation of construction toy play. Table 8 contains a list of all variables included in this study, and also indicates how items that were included in multiple questionnaires were combined to form a single variable.

6.4.2.3 Standardized learning achievement scores

Next to the questionnaire data and observational data derived from the visits, data on the academic performance of the children were obtained (see Table 8). Most Dutch schools take part in an ongoing national assessment program (Cito) and regularly evaluate their students' progress on several core subjects, such as (early) math and language skills. The scores included in this study were obtained from tests administered in kindergarten. On both tests, children could get a score from A (25% highest-scoring students) to E (10% lowest-scoring students). Although these standardized learning achievement tests do not cover science topics per se,[28] several general abilities important for scientific understanding are needed to perform well on these tests, such as executive functions and emotional self-regulation skills (Leseman, 2004). Given that children in kindergarten have limited spelling or number skills, the early math and language tests administered in kindergarten are mostly focused on mathematical and language *reasoning* (Wijnstra, Ouwens, & Béguin, 2003). That is, they address the ability to phrase words, understand questions (Cito, 2009), classify objects, and to measure and

---

[28] To this date, no ongoing assessment of performance in STEM areas is administered, apart from the math test included in our study.

observe differences and similarities (Koerhuis, 2011), which also comprised an important part of the hands-on science tasks administered in this study.

6.4.3 Procedure

6.4.3.1 Visits

During the 5 visits (one every 3 months), researcher and child were involved in a natural hands-on teaching-learning interaction. An adaptive protocol was constructed for each task, which guaranteed that all children were asked the basic questions reflecting the core building blocks of the task and the incorporated scientific concepts. At the same time, the protocol left enough space for children to take initiative and show their understanding spontaneously, and for the researcher to provide scaffolding when needed. For each task, the researcher showed the child the material and asked about its purpose and functioning. Afterward—regardless whether the child answered the previous questions right—he or she was encouraged to explore the material. The researcher asked questions about the task's functioning and underlying mechanisms, such as "Why does the piston of the other syringe go up when you push the piston of this syringe?" The researcher's scaffolding consisted of asking follow-up questions related to the child's (verbalized) level of understanding, encouraging the child to think about the task and to try out his/her ideas using the material, and summarizing the child's findings or previous answers. Even though children's answers were challenged sometimes, the feedback never included statements indicating whether the child was right or wrong. When the child could not give an answer, the researcher proceeded with another question or subject. Each task took approximately 20 minutes. All interactions took place within the schools or daycare centers and were recorded on video.

6.4.3.2 Questionnaires

Depending on the parents' preference, the questionnaires were either sent by e-mail, or given to the children to pass to their parents after the visit. One questionnaire was provided per child, which was filled out by the same parent each time, who was not informed about the child's development or performance during the tasks. The parent was instructed to simply fill out the questions and to send the questionnaires back. Parents were sent an electronic reminder if they did not return the questionnaire after a week. On average, 24.2 questionnaires (78%) were returned after each visit.

6.4.3.3 Coding of observational data

Using the computer program MediaCoder (Bos & Steenbeek, 2006), the recordings of the 5 visits were coded per utterance. The first step in the coding procedure was the determination of the moment when utterances started and ended. The second step involved the classification of all utterances of the child and the researcher into several categories. The researcher's utterances were coded as descriptive, predictive, and explanatory questions; encouragement; follow-up questions; compliments; clarifications; procedural remarks; directing the student's focus, and off-task utterances. The student's utterances were classified into descriptive, predictive, and explanatory answers/remarks; initiatives; content-related questions, and off-task utterances. After this initial classification, we combined children's coherent descriptive, predictive, and explanatory answers into meaningful units. The unit ended when the next utterance of the student fell into another category, or when the researcher interrupted the student (e.g., by asking another question). However, if the researcher simply encouraged the student to tell more about the same topic, the unit would continue.

In the fourth and final step, the complexity of students' answers within a unit were determined using a scale based on skill theory (Fischer & Bidell, 2006). Skill theory focuses on the complexity and variability of children's skills, which consist of actions and thinking abilities, embodied in verbal and non-verbal behavior in a specific context. The scale enables researchers to extract the complexity (of e.g., utterances) from content, which makes it possible to compare understanding across multiple seconds, contexts, and persons (Parziale & Fischer, 1998). According to this theory, learning is defined as building sets of skills, which are hierarchically ordered in 10 levels grouped into three tiers. The first tier consists of sensorimotor skills: simple connections of perceptions to actions or utterances. The second tier consists of representational skills; these are understandings that go beyond current perception-action couplings. The third and last tier consists of abstractions, which are general nonconcrete rules that also apply to other situations (Schwartz & Fischer, 2004). Within each tier, three levels can be distinguished: single sets, mappings (a relation between two single sets), and systems (a relation between two mappings). The levels assigned in our study (see Table 9) ranged from single sensorimotor sets (level 1) to single abstractions (level 7). We did not assign levels to incorrect answers or remarks, and kept these out of the analysis.

In order to make sure that the codings were reliable, a standardized codebook was used. For each round of coding (categories, units, and understanding levels), 10 raters went through a training by coding 3 video fragments of 15 minutes. The codings of the third fragment were compared to the codings of the first author (who constructed the codebook) and percentages of agreement were calculated. On average, these were: categories: 83% (range 80-93%; $p < .01$), units: 87% (range 80-100%; $p < .01$), and level of understanding: 84% (range 78-92%; $p < .01$).[29] Although the gender, age and school type could be inferred from the videos while coding, the raters were not aware of the child's standardized learning achievement test scores.

After coding we calculated the proportions of the researcher's content-related questions, follow-up questions, clarifications, as well as the child's content-related answers/remarks, initiatives, and off-task utterances over the five visits (these interaction variables are also listed in Table 8). As an outcome measure, children's frequencies of the understanding levels per task and per level were calculated. Since children frequently vary in understanding levels during the task, not the mean level, but the proportion of high understanding levels per task (level 4–single representations—and higher, divided by the total number of understanding levels) was calculated for each child. This would give us an

---

[29] *P*-values are calculated using Monte Carlo permutation tests (see below for an explanation).

indication of the highest possible levels the child can reach. We calculated 10 proportions of these high understanding levels: 2 (one for each task) for all 5 visits.

6.4.4 Analysis

The analysis consisted of three phases. First, using their 10 proportions of high understanding levels as input (1 for each task per visit), the children were divided in clusters using a hierarchical agglomerative clustering (HAC) analysis in Tanagra 1.4.18 (Rakotomalala, 2005).[30] The progress of the clusters over time was explored visually, and by comparing the mean difference in the proportion of high understanding levels between visit 1 and 5, using Monte Carlo permutation tests (Todman & Dugard, 2001). This statistical procedure can be easily applied to small samples and skewed distributions, while still producing reliable statistical results.

Taking the sample distribution into account, a Monte Carlo test measures the probability that a difference or statistic is caused by chance alone. This is done by drawing 5000 random samples from the original data, after which one can determine how often the observed or a bigger difference occurs in these random samples (positive cases). The number of positive cases is divided by the number of samples (5000), which produces a *p*-value. We decided to discuss all interesting results, defined as those having a *p*-value below .1.

To first examine the associations between the clustering and the general variables Gender, Age, School type, and Cito learning achievement test scores (chosen because their predictive value has been examined in other studies before), we performed a group characterization analysis in Tanagra to obtain comparative descriptive statistics. For each of these variables we obtained a test value,[31] which is a statistical comparison of the cluster and overall mean (for discrete variables a comparison of the cluster and overall proportion). High

---

[30] We asked the program to detect the most optimal number of clusters. Using a K-means clustering in the same program, we checked the validity of the HAC clustering. The K-means clustering was an almost perfect copy of the HAC clustering ($\chi^2 = 55.11$, $p < .001$).

[31] An explanation of the test value statistic of the Tanagra program can be found here: http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/en_Tanagra_Comprendre_La_Valeur_Test.pdf

absolute test values indicate a high predictive value of this particular variable for a specific cluster. Since the group characterization only yields a ranking of variables based on the test value, but no *p*-values, we also performed a series of Monte Carlo permutation tests to further examine the relative predictive value of the general variables Gender, Age, School type and Cito scores.

Next, to assess which macro and micro-variables would be most associated with the HAC clustering, an improved CHAID algorithm (Belaïd, Moinel, & Rangoni, 2010; cf. Kass, 1980) was used in Sipina 3.11 (Zighed & Rakotomalala, 1996) to build supervised learning decision trees. The input variables used for the decision trees were derived from both the questionnaires and recordings (see Table 8). Separate decision trees were made for each group (demographic, school, home environment, and interaction variables), and for all variables combined. The nodes of the decision tree were manually split using the variable that would contribute most (i.e., the variable with the highest Tschuprow's T and goodness of split).[32] We continued splitting until all children within a node were from the same cluster, or when only one child of another cluster was left in the same node. In two instances, the program indicated no further split was possible because of the limited predictive value of the remaining variables. The rules of each decision tree were evaluated using four statistics: Confidence (the percentage of children within a node that belong to a specific cluster); lift (the confidence divided by the overall percentage of children that belong to that cluster); support (the total number of children within a node divided by *N*), and strength (a test statistic comparing the cluster mean with the overall mean, or, in case of discrete variables, the cluster proportion with the overall proportion).

---

[32] If the program indicated that the predictive value of two or more variables was roughly equal, we used the variable that resulted in the most unambiguous split, meaning that it would lead to an isolation of a single cluster as much as possible.

6.5 Results

6.5.1 How can we characterize the developmental patterns of children's understanding of scientific concepts, in terms of their shape?

Using the 10 mean proportions of high understanding levels, a hierarchical cluster analysis yielded three clusters as the best statistical solution. Focusing on the shape of the trajectories, cluster 1 (*n* = 14) scored higher than the other two clusters on both the air pressure (Figure 11A) and gravity tasks (Figure 11B) during each visit. This indicates that this cluster had the highest proportion of high understanding levels (level 4 and higher) over all visits. During the air pressure task of visit 1, for example, 30% of the skill levels of cluster 1 were at least at the single representations level, whereas this was 5% and 4% for cluster 2 and 3 respectively (see Figure 11A). A Monte Carlo analysis revealed that this cluster's progress over 5 visits (the difference in proportion of high understanding levels between the first and the fifth visit) was marginally significant on the air pressure tasks (*p* = .1) and significant on the gravity tasks (*p* = .002).

Cluster 2 (*n* = 6) varied mostly with regard to the proportion of high understanding levels, especially on the air pressure tasks (see Figure 11A). They reached the level of cluster 1 during the second visit, but scored lower than the other two clusters during the third visit. Cluster 2 made significant progress over 5 visits on the air pressure tasks (*p* = .004), but not on the gravity tasks (*p* = .13). Cluster 3 (*n* = 11) scored below the overall average on almost all tasks, except for the first gravity task (see Table 10 and Figure 11B). While cluster 1 and 2 mostly alternated their in- and decreases, cluster 3 showed a brief increase on the air pressure tasks, which stabilized after the second visit (Figure 11A). For the gravity tasks (Figure 11B), cluster 3 showed a decrease from the first to the fourth visit, which was followed by a sharp increase. Similar to cluster 2, cluster 3 also made significant progress on the air pressure tasks (*p* = .0004), but not on the gravity tasks (*p* = .23).

6.5.2 How can we characterize the developmental patterns in terms of their associations with microgenetic and macro predictive factors?

To see whether we could explain the clustering by the general variables Gender, Age, School type and Cito scores, we first looked at the group characterization statistics in Table 10. High absolute test values indicate a high predictive value of that particular variable for a specific cluster. In this case, however, most test values did not exceed the absolute value of 2, apart from the percentage girls in cluster 2 (test value = 2.22), which was higher than in the other two clusters. A Monte Carlo permutation test revealed that this difference was marginally significant (67%, *p* = .07). The variable Age had a high absolute test value for cluster 1 (2.8), indicating that children in this cluster were slightly older than the overall group. Monte Carlo permutation tests confirmed that cluster 1 had a higher percentage of 5-year olds (64%, *p* = .03), and a lower percentage of 3-year olds (14%, *p* = .07). In addition, cluster 2 had a lower percentage of 5-year olds (0%, *p* = .08). No other significant differences were found. Thus, although cluster 1 could be characterized as slightly older, and cluster 2 as slightly more 'feminine', there was no distinct distribution of the variables Age and Gender across all three clusters. In addition, no significant associations between the clustering and children's Cito scores and School type were found.

Subsequently, we used an improved CHAID algorithm to build 5 decision trees to assess which variables best predict the cluster compositions (see Table 11). Using the demographic variables, the decision tree best predicting the clustering only used the diagnosis (internalizing, externalizing, or none) and the age of the child. No additional split was possible given the low predictive value of the remaining demographic variables, and the rules had low support, and a moderate strength and lift. With regard to the school variables, the only variable that could be used to split the tree is whether the child was born early, in the middle, or late in the academic year. The rules, however, had low confidence, lift, and strength, and again no additional split was possible given the low predictive value of the remaining school variables. Taking the home environment variables resulted in a tree in which the splits were based on parental encouragement to play with construction toys, sharing school experiences, the child’s exploratory behavior, and interest in numbers and counting. The rules had a moderate to high confidence, lift, support and strength. Lastly, the interaction variables yielded a decision tree with rules high in confidence, lift, support and strength. The most important interaction variables that predicted the clustering were the proportion of the researcher’s follow-up questions during various visits and the proportion of the child’s initiatives during visit 3. Figure 12 shows the decision tree for all variables combined. The rules had the highest confidence, lift, support and strength, and did not contain any of the demographic and school variables. Instead, the interaction and home environment variables determined the distribution of children across the three clusters, with the decisive variables being the child’s off-task behavior during visit 2, the researcher’s clarifications and follow-up questions during visit 1, and parental encouragement of playing with construction toys and sharing school experiences. In general, cluster 1 was characterized by low parental encouragement of construction toys (< 2.56), and a low proportion of off-task behavior during visit 2 ($< .29$). Most children in cluster 3 got more parental encouragement to play with construction toys (> 2.56) and shared less school experiences (< 2.08). To predict membership of cluster 2, more variables were needed. Just like cluster 3, these children received more parental encouragement to play with construction toys (> 2.56), but they also shared more school experiences (> 2.08), had a higher proportion of clarifications during visit 1 (> 0.01), and were asked less follow-up questions (< 0.17).

To summarize, a few demographic and school variables contributed to the prediction of the clustering, but their rules had low lift, support and strength, and their predictive value disappeared when using all variables as input. The decision tree with the highest confidence, lift, support and strength used a combination of interaction and home environment variables to determine the distribution of the children across the 3 clusters.

Table 8: Description of the variables used in this study (predictors for the supervised learning decision tree in phase 2 of the analyses).

Variable	**Type**	**Description**
Age	Demographic (continuous)	Child's age, rounded to whole numbers
Gender	Demographic (discrete)	Child's gender
Diagnosis	Demographic (discrete)	Child's diagnosis (internalizing, externalizing, none)
Family composition	Demographic (discrete)	Two-parent, single-parent, or other family composition
Mother's nationality	Demographic (discrete)	Mother's nationality: Dutch or foreign
Father's nationality	Demographic (discrete)	Father's nationality: Dutch or foreign
Mother's education level	Demographic (discrete)	Highest education level attained: secondary school, vocational education, (applied) university
Father's education level	Demographic (discrete)	Highest education level attained: secondary school, vocational education, (applied) university
School type	School (discrete)	Child's school: regular or special education
Cito score language	School (discrete)	Standardized test on language skills, scoring between A (high) and E (low), administered in kindergarten
Cito score math	School (discrete)	Standardized test on early math skills, scoring between A (high) and E (low), administered in kindergarten
Born early/mid/late in the academic year	School (discrete)	Child's date of birth early/mid/late in the academic year
Member of sports club	Home (discrete)	Is child member of a sports club (yes/no) over 5 visits
Motor development	Home (discrete)	Child's motor development rated by the parents as below average, average, or above average
Physical development	Home (discrete)	Child's physical development rated by the parents as below average, average, or above average
Language development	Home (discrete)	Child's language development rated by the parents as below average, average, or above

		average
Emotional development	Home (discrete)	Child’s emotional development rated by the parents as below average, average, or above average
Sum exploratory behavior	Home (continuous)	Based on part 1 of the questionnaires (child’s problem solving, curiosity and exploratory behavior), 14 items answered on a 7-point scale. A factor analysis revealed 1 clear axis with an Eigenvalue of 6.6 containing 8 items. Per child, the mean score of these 8 items over 5 visits were added.
Preference for playing with educational toys	Home (continuous)	Parents noted their child’s favorite toys over 5 visits. We divided the wide range of toys into 2 categories: educational toys (e.g., construction toys, puzzles) and “just for fun toys” (e.g., dressing up). We calculated the average number of reported educational toys.
Frequent playing with educational toys	Home (discrete)	Indicating whether the child plays at least once per week with educational toys (yes/no), as reported by the parents after visit 4 and 5.
Child’s interest in counting and numbers	Home (continuous)	How often the child is practicing counting and number recognition, indicated by the parents on a 5-point scale after visit 1-3. We took the average.
Parental encouragement counting and numbers	Home (continuous)	How often the parents encourage practicing counting and number recognition on a 5-point scale after visit 1-3. We took the average.
Parental encouragement construction toys	Home (continuous)	How often the parents encourage playing with construction toys on a 5-point scale after visit 1-3. We took the average.
Cooperative parent/child play educational toys	Home (continuous)	The average number of reported educational toys used during cooperative parent/child play over 5 visits.
Household chores	Home (continuous)	Average number of child’s household chores, reported by the parents after visit 3-5.
Sharing school experiences	Home (continuous)	Whether the child shares school experiences at home on a 4-point scale, indicated by the parents after visit 3-5. We took the average.
Spontaneity of child sharing school experiences	Home (continuous)	Whether the child shares school experiences spontaneously on a 3-point scale, indicated by the parents after visit 3-5. We took the average.

Researcher’s follow-up questions	Interaction (continuous)	Mean proportion of researcher’s follow-up questions (asking the child to elaborate on his/her previous answer) for each visit (5 variables).
Researcher’s content-related questions	Interaction (continuous)	Mean proportion of researcher’s content-related questions (descriptive, predictive, exploratory questions; excluding follow-up questions) for each visit (5 variables).
Researcher’s clarifications	Interaction (continuous)	Mean proportion of researcher’s clarifications (when the researcher summarized the child’s findings or previous answers) for each visit (5 variables).
Child’s off-task behavior	Interaction (continuous)	Mean proportion of child’s verbalized off-task behavior (utterances not related to the task) for each visit (5 variables).
Child’s initiatives	Interaction (continuous)	Mean proportion of child’s verbalized initiatives (request to perform an action or explore another part of the task) for each visit (5 variables).
Child’s content-related utterances	Interaction (continuous)	Mean proportion of child’s content-related utterances (answers to questions and remarks about the task content) for each visit (5 variables).
Proportion high understanding levels on air pressure tasks	Outcome (continuous)	The number of high understanding levels (level 4 and higher) divided by the total number of understanding levels, for each child on each air pressure task. These proportions were used as input for the cluster analysis.
Proportion high understanding levels on gravity tasks	Outcome (continuous)	The number of high understanding levels (level 4 and higher) divided by the total number of understanding levels, for each child on each gravity task. These proportions were used as input for the cluster analysis.
Cluster number (after cluster analysis)	Outcome (discrete)	Number indicating the membership of a specific cluster, used as outcome measure for the decision tree analysis.

Table 9. Description and examples of the understanding levels assigned in our study, based on skill theory (e.g., Fischer & Bidell, 2006).

Level	**Description of answer (example)**
1. Single sensorimotor set	Single characteristics of the task (“This ball is fast”)
2. Sensorimotor mapping	Links between single task characteristics, simple comparisons (“This ball rolls faster than that one”)
3. Sensorimotor system	Observable causal relations (“If I push the piston of this syringe, then the piston of the other one moves”)
4. Single representation	Coupling two causal relations; not directly observable relations and simple predictions (“Air causes the piston of the syringe to move”)
5. Representational mapping	Predictions and explanations in terms of a relation between two single representations (“The piston pushes the air, which travels through the tube to the other piston, which then gets pushed out by the air”)
6. Representational system	A coupling between two representational mappings (“This syringe contains air, and if I push its piston, the air goes through the tube to the other syringe, and pushes that piston upward. When I push that piston, the same mechanism causes the first one to go up”)
7. Single abstraction	A general (immaterial) concept that goes beyond (representations of) the material (“Gravity causes the marbles to go down when we release them”)

Table 10. Characterization of clusters in terms of the outcome variable (proportion high levels), and the variables age, gender, type of school, Cito scores.

Cluster HAC 1 (*n* = 14; 45.2%)				**Cluster HAC 2** (*n* = 6; 19.4%)				**Cluster HAC 3** (*n* = 11; 35.5%)			
Continuous variable	Test value	Group mean (SD)	Overall mean (SD)	Continuous variable	Test value	Group mean (SD)	Overall mean (SD)	Continuous variable	Test value	Group mean (SD)	Overall mean (SD)
Air 1	3.7	0.30 (0.2)	0.16 (0.2)	**Air 2**	1.68	0.55 (0.1)	0.43 (0.2)	**Gravity 1**	0.67	0.35 (0.1)	0.33 (0.1)
Air 4	3.5	0.43 (0.1)	0.32 (0.2)	**Gravity 2**	0.07	0.39 (0.2)	0.39 (0.2)	**Gravity 3**	-0.15	0.21 (0.1)	0.22 (0.1)
Air 3	3.5	0.38 (0.1)	0.27 (0.2)	**Gravity 4**	-0.44	0.24 (0.04)	0.26 (0.1)	**Gravity 5**	-0.98	0.39 (0.2)	0.43 (0.2)
Air 2	3.1	0.55 (0.1)	0.43 (0.2)	**Air 5**	-0.83	0.26 (0.09)	0.30 (0.1)	**Air 3**	-1.03	0.24 (0.1)	0.27 (0.2)
Air 5	3.0	0.38 (0.1)	0.30 (0.1)	**Gravity 1**	-1.31	0.26 (02)	0.33 (0.1)	**Age (mo)**	-1.47	48.73 (11.2)	52.94 (11.7)
Age (mo)	2.8	59.5 (10.1)	52.94 (11.7)	**Gravity 5**	-1.42	0.35 (0.1)	0.43 (0.2)	**Air 4**	-2.35	0.23 (0.1)	0.32 (0.2)
Gravity 4	2.8	0.32 (0.1)	0.26 (0.1)	**Air 4**	-1.59	0.22 (0.1)	0.32 (0.2)	**Air 5**	-2.48	0.22 (0.1)	0.30 (0.1)
Gravity 2	2.5	0.46 (0.1)	0.39 (0.2)	**Air 1**	-1.6	0.05 (0.1)	0.16 (0.2)	**Gravity 4**	-2.53	0.19 (0.09)	0.26 (0.1)
Gravity 5	2.1	0.49 (0.1)	0.43 (0.2)	**Age (mo)**	-1.75	45.33 (8.6)	52.9 (11.7)	**Air 1**	-2.57	0.04 (0.06)	0.16 (0.2)
Gravity 3	1.7	0.26 (0.1)	0.22 (0.1)	**Gravity 3**	-2	0.13 (0.07)	0.22 (0.1)	**Gravity 2**	-2.63	0.29 (0.2)	0.39 (0.2)
Gravity 1	0.4	0.34 (0.1)	0.33 (0.1)	**Air 3**	-3.14	0.10 (0.1)	0.27 (0.2)	**Air 2**	-4.61	0.20 (0.1)	0.43 (0.2)

Cluster HAC 1 (*n* = 14; 45.2%)				**Cluster HAC 2** (*n* = 6; 19.4%)				**Cluster HAC 3** (*n* = 11; 35.5%)			
Discrete variable	Test value	Cluster % (% of total)	Overall %	Discrete variable	Test value	Cluster % (% of total)	Overall %	Discrete variable	Test value	Cluster % (% of total)	Overall %
% Spec. ed.	-0.2	43% (43%)	45%	**% Girls**	2.22	44% (67%)	29%	**% Spec. ed.**	1.51	50% (64%)	45%
% Girls	-0.8	33% (21%)	29%	**% Spec. ed.**	-1.54	7% (17%)	45%	**% Girls**	-0.97	22% (18%)	29%
Cito math A	0.9	55% (55%)	44%	**Cito lang E**	1.78	100% (17%)	4%	**Cito lang C**	1.97	75% (38%)	16%
Cito lang. A	0.9	55% (55%)	44%	**Cito math E**	1.78	100% (17%)	4%	**Cito math D**	1.46	100% (13%)	4%
Cito lang. B	0.03	44% (36%)	36%	**Cito math A**	0.33	27% (50%)	44%	**Cito math C**	1.34	67% (25%)	12%
Cito math B	0.03	44% (36%)	36%	**Cito lang A**	0.33	27% (50%)	44%	**Cito lang B**	0.11	33% (38%)	36%
Cito math C	-0.4	33% (9%)	12%	**Cito math B**	-0.15	22% (33%)	36%	**Cito math B**	0.11	33% (38%)	36%
Cito lang. C	-0.8	25% (9%)	16%	**Cito lang B**	-0.15	22% (33%)	36%	**Cito math E**	-0.69	0% (0%)	4%
Cito math E	-0.9	0% (0%)	4%	**Cito math D**	-0.56	0% (0%)	4%	**Cito lang E**	-0.69	0% (0%)	4%
Cito lang. E	-0.9	0% (0%)	4%	**Cito math C**	-1.02	0% (0%)	12%	**Cito lang A**	-1.29	18% (25%)	44%
Cito math D	-0.9	0% (0%)	4%	**Cito lang C**	-1.2	0% (0%)	16%	**Cito math A**	-1.29	18% (25%)	44%

Note. Continuous variables are listed in the upper part, discrete variables in the lower part. Order of table entries depends on the test value, which can be used as a criterion for ranking the importance of the variables. High absolute test values indicate high predictive values of variables. Specific *p*-values cannot be calculated because of the interdependence between the clustering and the group characterization, but as a rule of thumb an absolute test value > 2 indicates a considerable difference

Table 11: Decision tree rules for the separate groups of variables and the variables combined

Variable type	Decision tree rules	Confidence	Lift	Support	Strength
Demographic	IF diagnose = internalize, THEN cluster = 3	100	2.86	0.19	3.62
	IF diagnose = none or externalize, AND age > 4.5, THEN cluster = 1	100	2.22	0.29	3.86
	IF diagnose = none or externalize, AND age < 4.5, THEN cluster = 2	38	2.00	0.19	2.60
School	IF born in academic year = mid, THEN cluster = 1	63	1.40	0.86	2.49
	IF born in academic year = early or late, THEN cluster = 2	42	2.21	0.83	2.46
	IF born in academic year = early or late, THEN cluster = 3	42	1.20	0.45	0.56
Home environment	IF parental encouragement construction toys < 2.56, THEN cluster = 1	85	1.89	0.79	3.69
	IF parental encouragement construction toys > 2.56, AND sharing school experiences < 2.08, THEN cluster = 3	100	2.86	0.64	3.99
	IF parental encouragement construction toys > 2.56, AND sharing school experiences > 2.08, AND sum exploratory behavior > 36.9, AND interest in counting and numbers > 2.59, THEN cluster = 2	83	4.37	0.83	4.35
Task environment	IF proportion follow-up questions M2 < 0.1, AND proportion follow-up questions M1 > 0.06, THEN cluster = 2	100	5.26	0.67	4.30
	IF proportion follow-up questions M2 > 0.1, AND proportion follow-up questions M3 > 0.04, AND proportion follow-up questions M5 > 0.08, THEN cluster = 1	92	2.04	0.79	4.07
	IF proportion follow-up questions M2 > 0.1, AND proportion follow-up questions M3 > 0.04, AND proportion follow-up questions M5 < 0.08, AND proportion initiatives M3 < 0.09, THEN cluster = 3	90	2.57	0.82	4.31
All	IF parental encouragement construction toys < 2.56, AND proportion off-task behavior M2 < 0.29, THEN cluster=1	100	2.22	0.79	4.48
	IF parental encouragement construction toys > 2.56, AND sharing school experiences < 2.08, THEN cluster = 3	100	2.86	0.64	3.99
	IF parental encouragement construction toys > 2.56, AND sharing school experiences > 2.08, AND proportion clarifications M1 > 0.01, AND proportion follow-up questions M1 < 0.17, THEN cluster = 2	100	5.26	1.00	5.48

Note. The rules describe how cluster compositions can be best predicted. Confidence represents the percentage of children within a node that belong to that cluster; lift is calculated by dividing the confidence by the overall percentage of children that belong to that cluster; support is calculated by dividing the total number of children in a node by *N*, and strength is a test statistic comparing the cluster mean/proportion to the overall mean/proportion. The higher the values of these statistics, the more relevant the rule.

Figure 11: Patterns of development for each cluster over 5 visits (M1-M5), split out by task sequence: Air pressure (A) and gravity (B). The Y-axis displays the cluster mean proportion of high understanding levels (the proportion of levels 4 and higher) during a visit.

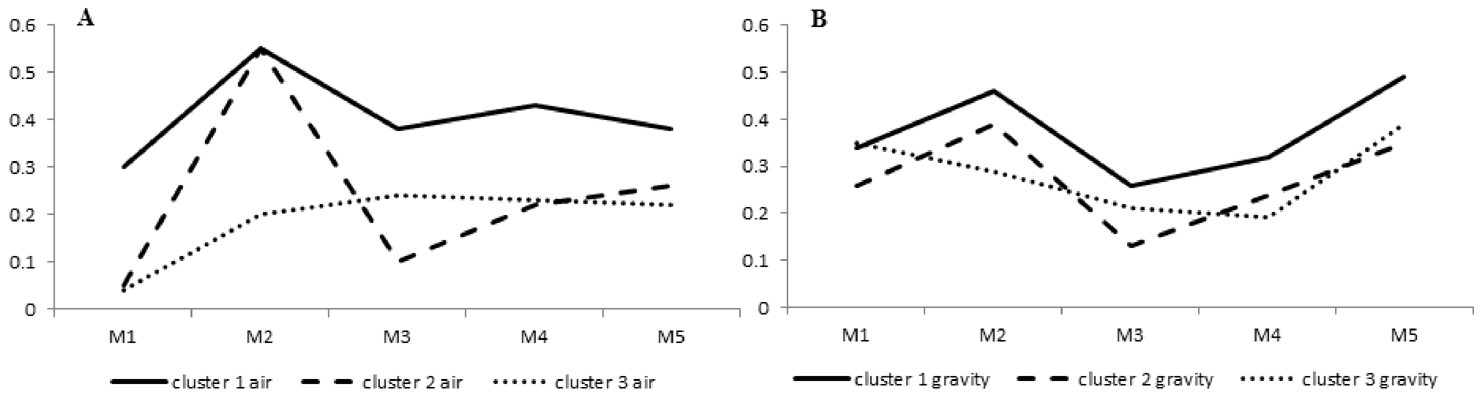
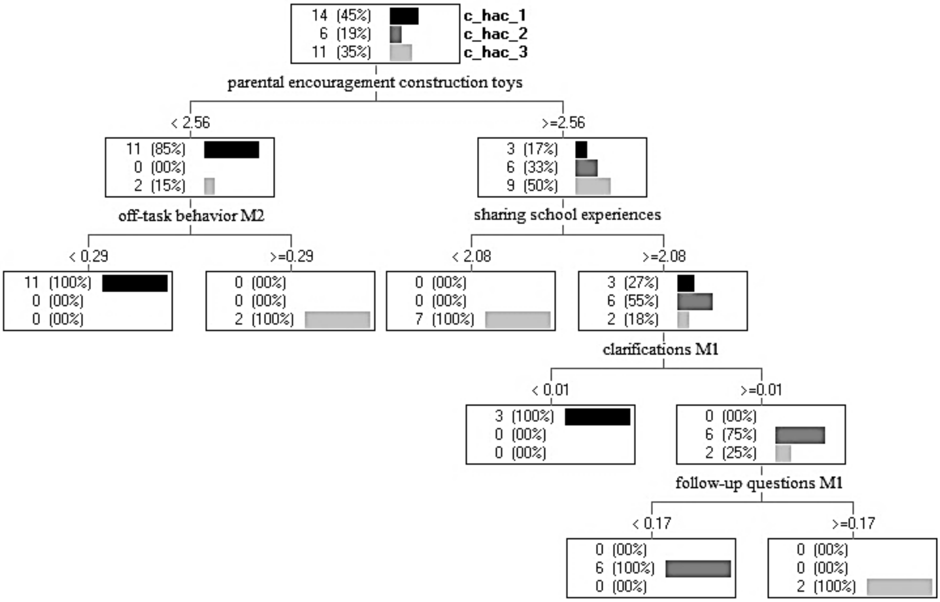


Figure 12: supervised learning decision tree with all variables (demographic, school, home environment, and task environment) as input. The rules (listed in Table 11) can be obtained by following the notes containing children from a specific cluster.



6.6 Discussion

6.6.1 Overview of findings

The aim of this study was to characterize the development of understanding the scientific concepts air pressure and gravity over the course of 1.5 years, in terms of both the shape and the predictive factors of the developmental trajectories. A cluster analysis yielded three groups with distinct developmental patterns over time, and provided a developmentally oriented differentiation on children’s developmental trajectories of scientific thinking. Cluster 1 could be characterized as the highest-scoring cluster, with proportions of high understanding levels positioned well-above the other two clusters. Cluster 2 was more variable: During some visits, it scored similar to cluster 1, while on other visits it scored similar to cluster 3. The third cluster usually scored below the other two clusters, showing lower proportions of high understanding levels, which remained fairly stable over time.

As it turns out, the predictive value of the demographic and school factors to determine the distribution of children across the 3 clusters was negligible. Using only these variables to predict the clustering yielded decision trees with rules low in strength, and their predictive value was overruled in combination with the home environment and interaction variables. Although we could not find a clear distribution of the variable age across all clusters, the best scoring cluster (1) was slightly older than the rest, indicating that age does play a role, although a marginal one in this age range (3-5 years). No clear distribution was found for the variable gender as well, although cluster 2 had a significantly higher percentage of girls (67%). This cluster was highly variable, alternating in high and low proportions of high understanding levels from visit to visit. This is noteworthy, as research has pointed out that especially boys are more variable in their performance on science tasks, which is assumed to have a large effect on their future ability to excel (Burkam, Lee, & Smerdon, 1997; Hedges & Nowell, 1995). Earlier studies commented on the gender gap being small at a young age (OECD, 2004), and increasing over time due to subtle social influences (Spelke, 2005), which might be an explanation for the absence of gender differences. Another possible explanation could be the use of hands-on tasks in this study, as research has shown that girls' performance tends to be better when they have the opportunity to manipulate materials (Burkam, et al., 1997).

While the predictive values of the demographic and school variables were low, the variables with the highest predictive values were the interaction variables (child's off-task behavior, researcher's clarifications, and follow-up questions during the first visits), and the home environment variables (parental encouragement of construction toys, and sharing of school experiences). These variables—yielding a decision tree with rules high in lift, support and strength—are not just fixed factors, but the product of ongoing interactions between the child and its home and task environment, which become stabilized patterns over time. For instance, the researcher does not decide to ask follow-up questions independently of the interaction, but because the current interaction elicits this behavior (e.g., she gets the idea that she has not fully grasped the child's understanding of the concept at issue). If the child indeed responds to this follow-up question by explaining the previous answer more in depth, chances are that the researcher will do this again. An iterative sequence like this ultimately leads to stabilized patterns within the interaction, called attractor states (Thelen & Smith, 1994; Van Geert, 1994). Hence, the results do not suggest that parents should encourage playing with construction toys, or that follow-up questions should be asked, but rather that the characteristics of the real time bi-directional interactions between children and their environment are associated with their developmental patterns over time (shown by their membership of a specific cluster). The child- and context variables are coupled, that is, they have a dynamic relationship, and should not be treated as single independent variables predicting children's development over time.

6.6.2 Special needs students and testing

The nature of the interaction is of particular importance when looking at the performance of the special needs students. Although numerous studies have found that these children show significant academic delays (Reid et al., 2004), in this study they did not perform worse than their peers enrolled in regular schools. To the contrary, the special needs students were divided over all 3 clusters, and School type was in none of the decision tree analyses a predictive variable, even when only the demographic variables were used.[33] Previous studies measured special needs students' performance using standardized tests. It is well-known that standardized tests not only measure the constructs they claim, but also some construct-irrelevant 'noise' (Haladyna & Downing, 2004). A famous example is reading comprehension (Messick, 1989). Students with reading difficulties score lower on math and science tests that require reading, with non-valid test scores

---

[33] Using the demographic variables only, the decision tree actually contained the variable Diagnosis. However, this yielded a node with 6 children with internalizing problems (all from cluster 3), consisting of 5 special school students and 1 regular student. The remaining 9 special school students were divided over cluster 1 and 2.

as result (Almond et al., 2010). While reading difficulties can be overcome by using nonverbal tests or text to speech software, most standardized test situations might be unable to meet the difficulties of many special needs students, which include attention problems, test anxiety, motivation problems, or difficulties to verbalize their thoughts (Cooper et al., 2004). The condition of optimal scaffolding in the current study, in which the researcher continuously draws the student's attention, changes the wording of questions if necessary, and uses follow-up questions[34], might be better suited to assess the understanding levels of special needs students.

From a practical point of view, the results of this study cast doubt on the assumption that individually made paper-and-pencil tests are indeed objective context-independent predictors of all academic performances. Children's standardized learning achievement test scores were not predictive of their long-term developmental patterns on the science tasks administered in this study. We already discussed that according to a complexity view, child and context characteristics are intertwined, making it impossible to assess or predict performance independently and across all contexts. Although more research is needed, with other age groups and in other academic fields, this viewpoint is supported by this study.

6.6.3 Limitations and future directions

It should be noted, however, that this study has some limitations. Even though many variables are included, some variables that are considered to be important for academic performance in STEM fields were not explicitly taken into account, such as working memory, executive functioning measures or general intelligence. That said, given that earlier research found that these variables significantly correlated with standardized math test scores (Bull et al., 2008), it is plausible that the standardized learning achievement test scores in our study

---

[34] The condition of optimal scaffolding we propose (ongoing use of a variety of scaffolding techniques directed at the needs of the individual child) is different from a dynamic testing method in which repeated formal test sessions are alternated with an intervention, or standardized forms of feedback (Grigorenko & Sternberg, 1998).

provided a rough measure of these characteristics. In addition, although the standardized learning achievement tests of math and language in this study did not cover science topics, they did address the ability to phrase words, understand questions (Cito, 2009), classify objects, and to measure and observe differences and similarities (Koerhuis, 2011). These skills also comprised an important part of the hands-on science tasks administered in this study.

Another possible limitation is that the sample used for this study was quite small, which might hinder the generalization of the findings. Given that we measured the interactions during the visits microgenetically—which is quite time-consuming—a small sample was, however, inevitable. It has to be noted, however, that this sample was quite representative for the Dutch school population, including boys and girls, three separate (young) cohorts, and children from both regular and special schools.

This study showed that the most important variables to characterize the long-term development of understanding scientific concepts were the product of ongoing interactions between the child and its home and task environment. These results question the predictive value of general predictors for the performance on scientific tasks, such as children's school type and standardized (learning achievement) scores. Our hope is that this study will be extended to other academic areas. Ultimately, new methods to evaluate academic achievements can be constructed, to satisfy both the need to evaluate children's academic performance, and to take on the issue of the intertwining of child-context dynamics driving this.

Chapter 7: Summary and General Discussion

This chapter gives an overview of this dissertation, and addresses some general discussion points related to the previous chapters.

In this study we used a longitudinal microgenetic method to examine young children’s understanding of scientific concepts in a prospective way. The main research question was: “How does children’s understanding of scientific concepts develop over the course of 1.5 years in interaction with the social and material context, and are special needs students equally able to acquire these skills?” We visited 32 children (3-5 years old) from both regular and special educational facilities, and provided them with hands-on scientific tasks about air flow/pressure and gravity/inertia/acceleration (see appendix A).

Children were visited each academic trimester for 3 years, and worked individually on the two tasks guided by the researcher’s questions. For this dissertation we used the data of 5 visits. Each visit was recorded on video, coded, and analyzed. The task situation was set up to reflect an inquiry-based individual science lesson. We used an adaptive protocol, which guaranteed that all children were asked the same basic questions reflecting the core building blocks of the tasks and the incorporated scientific concepts. At the same time, the protocol left enough space for children to take initiative and show their understanding spontaneously, and for the researcher to provide scaffolding when needed.

Given that we specifically view the understanding of scientific concepts as an ongoing process distributed across child and context, our goal was to examine this process while the children worked on the tasks. We therefore adopted a process approach (cf. Van Geert & Steenbeek, 2005a) in combination with a microgenetic method (cf. Granott & Parziale, 2002). That is, we studied the process of children’s understanding by means of frequent observations of both the child’s and the researcher’s utterances while working on the tasks. In this way, we obtained important information on how changes in understanding occur in interaction with the material and social context. Using a coding system developed for this study, we divided the utterances of the child and researcher into categories, and used skill theory to assess the complexity level of children’s task-related answers and remarks.[35] Additional information on children’s home environment was collected by means of questionnaires, filled out by their

---

[35] For chapter 5 we also coded the complexity of the researcher’s task-related descriptive, predictive, exploratory, and follow-up questions.

parents. We also obtained children's standardized (Cito) learning achievement test scores on early language and math tests, administered by their schools to keep track of children's progress. The early math and language tests administered in kindergarten are mostly focused on mathematical and language reasoning (Wijnstra, Ouwens, & Béguin, 2003). That is, these tests address children's ability to phrase and understand words, classify objects, and to measure and observe differences and similarities (Cito 2009; Koerhuis, 2011).

This dissertation is focused on a line of research in which each chapter is either related to a specific part of this longitudinal study, or takes a specific perspective on the data. Together, these parts give us an idea of how children's understanding develops while working on hands-on tasks guided by an adult, and of their developmental trajectories on the long term. Chapter 2 is focused on the set-up of the study, discussing its theoretical and practical foundations. In chapter 3, the first qualitative data of this study is used to build a theoretical model of understanding. Chapter 4 is focused on a cross-sectional comparison of regular and special needs students' understanding during one visit. Chapter 5 is centered on a case study, in which the microgenetic data of several sessions are coupled to get a picture of the development of understanding on a longer time scale. Lastly, in chapter 6 the developmental patterns of scientific understanding over the course of 5 visits are characterized in terms of both their shape, and the predictive factors underlying these patterns. To summarize this dissertation, we first focus on the findings of this study, organized by chapter. After this, we will proceed to a discussion covering the practical and theoretical consequences of this study's process approach and its outcomes.

7.1 Findings of this study

7.1.1 Chapter 3: Using the dynamics of a person-context system to describe children's understanding of air pressure

The purpose of this chapter was to lay the theoretical foundations of this dissertation's line of research. Our conceptual model captures understanding as a process that unfolds by means of the bi-directional interactions with the proximal environment. The advantage of using this conceptual model is that it makes the development of understanding more transparent and no longer limited to an invisible process inside the individual learner (Thelen, 1992; Van Geert & Fischer, 2009). Instead, it enables us to closely monitor interactions between child and environment to determine how a child's current understanding is constructed in real time. The model is based on 4 dynamic systems' properties: Intertwining person-context dynamics, iterativeness, intra-individual variability, and interacting time scales. Skill theory formed the basis of the coding system we constructed to determine the complexity of children's understanding levels.

We will now proceed by briefly illustrating the components of the model. First, from a dynamic view, understanding of scientific concepts can be seen as a process of intertwining person-context dynamics. That is, during a learning interaction, the child's verbal and nonverbal actions continuously affect the researcher (or another interaction partner for that matter) and the researcher's actions affect the child, creating the condition under which both components operate during the next moment in time (Fogel & Garvey, 2007; Steenbeek, 2006). Within such an interaction, understanding emerges through iteration, that is, every step in understanding is based on the previous one and embedded in the current context. Driven by these bi-directional iterative interactions with the environment, the complexity of children's understanding fluctuates during a task (intra-individual variability). It can increase, but also temporally decrease (see also Yan & Fischer, 2002 on the notion of scalloping), for example when the task difficulty increases, or when the teacher's support decreases. The iterative processes on the short term time scale influence processes on the long-term time scale (Lewis, 2000). In addition, the large-scale patterns also influence the short-term processes, by shaping the structure and function of the interaction on the short term (Lewis & Granic, 2000; Smith & Thelen, 2003; Steenbeek, 2006; Van Geert & Steenbeek, 2005a). For more detail and a visual interpretation of this conceptual model, see chapter 3.

7.1.2 Chapter 4: A comparison between young students with and without special needs on their understanding of scientific concepts

This chapter focuses on a cross-sectional comparison of regular and special needs students' understanding of the scientific tasks during one visit. The question was how the special needs students (i.e., children with behavioral or psychological problems) would develop their understanding of the scientific concepts air pressure and gravity during our hands-on tasks, guided by the researcher who worked with an adaptive protocol with room to provide scaffolding, if the student needed this. Earlier studies have shown that special needs students' academic performance is worse than that of regular students, probably because their behavioral and/or emotional problems interfere with their academic performance (Trout, Nordness, Pierce, & Epstein, 2003). However, the focus of these previous studies was primarily on academic achievement in traditional domains, such as arithmetic and spelling, measured with standardized tests. The question was how these children would perform on our scientific tasks, and if they would benefit from this study's setting, including the guidance (scaffolding) provided by the researcher during the tasks.

The results of this chapter show that although the special needs students made more mistakes, their mean level of understanding and mean number of answers did not differ from that of the regular students. If the special needs students would perform worse than the regular students, we would expect them to show more of the lower complexity levels (level 1 and 2). Indeed, they showed a significantly higher proportion of level 1 (sensorimotor action) answers compared to the regular group, but the regular students outperformed the special needs students on level 2 (sensorimotor mapping). The regular students had a significantly higher proportion of level 3 answers (sensorimotor system), but this was mostly caused by a significant difference between the 3-year-old special needs and regular students. The 4- and 5-year old cohorts did not differ in their proportion of level 3 answers. In addition, all age groups had a roughly equal proportion of the highest levels found during this visit (level 4 and 5—single representation and representational mapping). These results are in contrast with the significant gap we found between the standardized tests scores of the regular and special needs students included in this study. Scaffolding techniques adapted to the level of the student might therefore be of crucial importance for children's understanding of scientific tasks.

7.1.3 Chapter 5: A process approach to children's understanding of scientific concepts: A longitudinal case study

After looking at group differences in the previous chapter, this chapter concentrated on a case study by focusing on the interaction between a 4-year old boy and the researcher, while working on air pressure tasks during 3 subsequent visits. The central research question was: How can we characterize the interaction dynamics—the boy's and the researcher's fluctuations in complexity levels—during a single session? To see how the interaction dynamics would change over a longer period of time, the microgenetic data of several short-term interactions were coupled (Steenbeek & Van Geert, 2013).

The results show the boy had multiple fluctuations in his understanding during the first session, which were omnipresent. Over the course of the three sessions, an increase in his number of (right) answers occurred, and the frequencies of the complexity levels shifted: The number of level 2 (sensorimotor mapping) answers increased during the second session, while the level 3 (sensorimotor system) answers decreased. In the third session, this was exactly the other way around. Regarding the interaction dynamics, the boy usually followed the researcher's in- and decreases in complexity during the first visit. Over time, the boy initiated significantly more of these simultaneous in- and decreases, whereas the researcher initiated less. During session 3, a significant increase in the proportion of the boy's initiations occurred, as well as a significant increase in his mean understanding level. In this study, the boy's increases in understanding thus accompanied an increase in taking the initiative during the interaction, which illustrates the ideas of self-regulated learning (Zimmerman, 1990).

7.1.4 Chapter 6: How to characterize the development of children's understanding of scientific concepts: A longitudinal microgenetic study

After looking at group differences during a single session (chapter 4), and thoroughly focusing on the interaction dynamics in an individual developmental trajectory over the course of several visits (chapter 5), the main question of this chapter was: How can we characterize the developmental patterns of children's understanding of scientific concepts over the course of 1.5 years (5 visits), in terms of both the shape and the predictive factors underlying those patterns? To study this, we coded the behavior and complexity levels of understanding of 31 children on the gravity and the air pressure tasks, and the scaffolding behavior of the researcher (such as the number of encouragements and follow-up questions). In addition, children's standardized math and language test scores were used for the analyses, as well as the questionnaire data.

Using 10 proportions of high understanding levels for each child—the proportion of high understanding levels (level 4 and higher) divided by the total number of understanding levels per task—a cluster analysis yielded three groups with distinct developmental patterns over time. Cluster 1 had the highest proportion of high understanding levels on all tasks. Cluster 2 was highly variable, alternating in high and low proportions of high understanding levels over the visits. Lastly, cluster 3 showed lower proportions of high understanding levels that remained somewhat stable over time. All groups showed significant progress on the air pressure tasks. Children made considerably less progress on the gravity tasks, and this was only significant for cluster 1. When making an attempt to predict the clustering (i.e., the trajectories over time), the demographic (e.g., age, gender, diagnosis) and school variables (e.g., school type, standardized test scores) had low predictive values. In contrast, the variables with the highest predictive values were those variables derived from the interaction (children's off-task behavior, researcher's clarifications, and follow-up questions), and the home environment variables derived from the questionnaires (parental encouragement of construction toys, and sharing school experiences at home). These variables represent the product of ongoing interactions between the child and its surroundings, both at home and during the tasks. This illustrates our earlier claim that the context plays a vital part in the development of understanding, and cannot merely be seen as a temporary outside influence.

7.2 Discussion points inspired by the background and results of this study

7.2.1 What are scientific concepts and how do people construct their understanding of these?

Scientific concepts are ideas about phenomena in the fields of science, technology, engineering, and mathematics (Baartman & Gravemeier, 2011; OECD, 2003). One of the interesting aspects of scientific concepts is that these are multi-faceted. There is no simple discrete state of understanding a particular concept, but a continuum ranging from understanding the basic elements to a full scientific understanding of a particular concept based on a particular theory, such as Newtonian physics, or relativity theory.[36] Furthermore, hands-on learning of scientific concepts accompanies a number of domain-general procedural skills (Zimmerman, 2000), such as hypothesizing, measuring, and observing. This makes learning about scientific concepts a challenging, though pleasant activity for children and adults with different knowledge backgrounds. For most of us there is always something to learn, as long as the teacher (or the educational material) has a more sophisticated view of this particular scientific concept, and is able to share this by connecting to the current knowledge base of the student (King, 1994; National Research Council, 2005; 2007; Resnick, 1987). This connecting does not necessarily get harder when the student has a more advanced understanding. For experts in a certain field, it can be easier to connect to other experts or advanced students than to the level of real novices, such as young children.

---

[36] Note that relativity theory provides a more complete form of understanding of particular physical principles than Newtonian physics. In this sense, even highly developed forms of scientific understanding are always amendable.

In this dissertation, learning is studied as an ongoing process, distributed across child and context, in which the complexity of children's understanding of scientific concepts grows over time. Within this process, many mechanisms are at work that influence in- and decreases in understanding. The emergence of a particular type of understanding occurs on the short-term time scale of hands-on learning, as well as on the long-term time scale of development. On the short-term time scale (during a task), the child's understanding can be represented as a bandwidth in which he/she fluctuates (Siegler, 2007; cf. Vygotsky, 1934/1986). Viewed as a process, learning on the long-term time scale seems to consist of two mechanisms: First, it consists of reaching the upper regions of the personal bandwidth of understanding, and second, getting the upper limit of the bandwidth up (Van Der Veer & Valsiner, 1991; Van Geert, 1998).

Researchers have suggested that intra-individual variability drives this learning process (Goldin-Meadow & Alibali, 2002; Siegler, 1995; Van Geert & Van Dijk, 2002). This variability is an intrinsic part of the learning process and usually occurs naturally in interaction with the context. A well-known example is the 'Wuggle' study (Granott, 1993), in which students' understanding of Lego robots repeatedly fluctuates, for example after a discussion with a peer, when something unexpected happens, or when the student is challenged by a complex question about the material. These contextual influences can, in turn, break down a dominant understanding level or learning strategy, meaning that there is room to try out other strategies, and to re-explore the educational material. If this leads to a more advanced understanding, learning has occurred (Siegler, 2007). Moreover, when the new strategy has fruitful consequences on a longer term, this can eventually become the dominant consolidated strategy, which can be conceived as an attractor state, that is, a relatively stable state the system tends to hold (e.g., Thelen & Smith, 1994).

Like the 'Wuggle' study, the setting of the current study is based on the principles of inquiry-learning. That is, a setting in which students are "actively engaged using both science processes and critical thinking skills as they search for answers" (Gibson & Chase, 2002, p. 694), while being assisted by a teacher or other facilitator. This setting may be well suited for detecting intra-individual variability in students' strategies, verbalized thought processes, and actions, given that it enables researchers to closely look at students' active learning attempts. At the same time, this setting may as well promote intra-individual variability by enabling students to actively try out different things with the educational materials while constructing their knowledge. Indeed, inquiry-based science programs may have positive effects on students' understanding of scientific concepts (Gibson & Chase, 2002; Kanter, Smith, McKenna, Rieger, & Linsenmeier, 2003; Lehrer, Schauble, & Lucas, 2008). However, note that unguided or poorly guided inquiry-based science programs might not be effective in helping students to construct scientific knowledge (Alfieri, Brooks, Aldrich & Tenenbaum, 2010; Kirschner, Sweller & Clark, 2006; Van Geert, 2011a).

There are many different types of knowledge generation processes, one of which is the co-constructed process between a researcher, child, and task discussed in this dissertation. Although no researcher would claim the context-independence of psychological, behavioral or cognitive constructs, it is everyday practice to standardize the context in studies to find 'genuine' effects that can be generalized to other contexts, for example by means of standardized tests (Richardson, Marsh, & Schmidt, 2010). Nonetheless, when a child is individually assessed, the child is asked to construct knowledge without the help of an adult, but always in interaction with the context, for example when reading a question of a standardized test, and using a piece of paper to draw or write the answer on. Given the view that understanding can be considered as a process of co-construction distributed across child and context, context-independence can be considered a myth, even in those standardized paper-and-pencil tests (cf. Richardson et al., 2010; Marsh, Johnston, Richardson & Schmidt, 2009). After all, while the teacher's help can be set to 0, it will never be possible to set contextual influences to a 0 level, because the assessment always requires a particular context of activity. Furthermore, if it would be possible to find such a context-independent situation,[37] the ecological validity is in question: does measuring in

---

[37] Although finding a *situation* that is *context-independent* could be considered a contradictio in terminis.

this context-independent situation generalize to other learning situations, for which we agree that these are context-dependent?

In Chapter 5 of this dissertation, we have seen that standardized math and language learning achievement test scores have low predictive values for children's developmental trajectories of scientific understanding during the tasks of this study. Although these standardized tests did not explicitly cover science topics, they did address the ability to phrase words, understand questions (Cito, 2009), classify objects, measure and observe differences and similarities (Koerhuis, 2011), which are skills that are also highly needed for the science tasks in this study. In addition, standardized math tests have shown a high correlation with working memory or executive functioning measures (Bull et al., 2008), two skills also needed for the scientific tasks in the current study. Yet, the variables with the highest predictive value for children's trajectories of understanding were either derived from the interaction between the child, task, and researcher, or from the interaction between the child and his/her parents at home. This signals the importance of the interaction with the proximal environment for learning, and the role of the support provided by the interaction partner during learning activities, for example in the form of scaffolding.

7.2.2 Our take on the role of scaffolding in the understanding process

During teaching-learning interactions, like the one discussed in this dissertation, tuning into the student's initial understanding is key (National Research Council, 2007). After this, a teacher's responsiveness to the ongoing interaction and changing understanding of the student is needed (National Research Council, 2007; cf. Steenbeek, Jansen, & Van Geert, 2012), as well as the use of adaptive scaffolding techniques whenever the student needs help or guidance. Scaffolding consists of the teacher's adaptive temporary support that helps the student forward in his/her learning process (Van Geert & Steenbeek, 2005b; Granott, Fischer, & Parziale, 2002).[38] The three key characteristics of scaffolding are: *Contingency*—adapting to the current level of the student,

---

[38] Note that there is also something that we could call "material scaffolding", e.g., the way textbooks are set up from easier to more complex items (Van Geert & Steenbeek, 2005b).

Fading—gradually decreasing the offered support, and *Transfer of responsibility*—from teacher control to student control (Van De Pol, Volman, & Beishuizen, 2010). Wood, Bruner, and Ross (1976) define six ways of scaffolding: recruitment of the child's interest (evoking enthusiasm), reduction in degrees of freedom (breaking down questions in terms of their complexity), maintaining goal orientation (directing the student's attention), highlighting critical task features, controlling frustration (affective support), and demonstrating idealized solution paths (modeling). In the adaptive protocol used for this study for example, the researcher had the opportunity to break down questions in terms of their complexity, change the wording of questions, ask follow-up questions, encourage the student to share his/her thoughts, give compliments, clarify the child's explanations, and guiding his or her attention.

Scaffolding is by definition a dynamic, idiosyncratic construct (Van Geert & Steenbeek, 2005b). This means that the ideal form of scaffolding does not exist. There only exists an ideal form at a *specific moment* in time for a *specific teacher-student pair* (cf. Ensing, Van Der Aalsvoort, Van Geert & Voet, 2014; Van De Pol et al., 2010). In our study, for example, one high-performing 5-year old seemed to benefit most from open-ended complex questions. He seemingly enjoyed exploring the various aspects of the question, and carefully formulated his answer afterwards. One of his peers enrolled in a school for special education, on the other hand, seemed to get confused when these broad open-ended questions were asked, and lost track while exploring the material and thinking about an answer. Soon the interaction became focused on small details, away from the most important task mechanisms. His learning process seemed to benefit from short close-ended questions and frequent clarifications of the researcher, which seemed to give him guidance and direction.

Note that a teaching-learning interaction is never a one-way street (Stone, 1998; Van de Pol et al., 2010). Student, teacher and task are dynamically intertwined, and should not be treated as separate components that contribute to the teaching-learning process in the classical sense of the word. Thus, to create and facilitate a genuine teaching-learning process, not only the teacher, but also the student actively participates (Rogoff, 1993), and expresses his/her understanding, or lack of understanding for that matter. For this, trust, motivation (Christophel, 1990) and exploration are the key ingredients. The student has to feel supported by the teacher to show what he or she knows, and needs to be motivated by the teacher and the task to explore what he or she does not know in order to get most out of the interaction (cf. Kupers, Van Dijk, & Van Geert, 2013; Steenbeek & Van Geert, 2013; see also the self-determination theory of Ryan & Deci, 2000). This dissertation's task situation is in particular one in which trust can be cultivated, especially because it resembles a natural learning situation, in which scaffolding is permitted, and not a formal testing situation.

Although scaffolding is an idiosyncratic phenomenon (i.e. its form depends on characteristics of a specific teacher-student pair in a specific situation), we want to highlight some specific points related to our findings. In Chapter 5, we used a case study to illustrate the reciprocal nature of the teaching-learning process over the course of three interactions, showing how the up- and down-trends in the boy's and researcher's complexity level were related. While the boy initiated more of the mutual in- and decreases in complexity level over time, the researcher took a step back, and initiated less. This co-occurred with a significant increase in the boy's mean understanding level. In terms of the three key characteristics of scaffolding suggested by Van de Pol et al. (2010), real-time fading and transfer of responsibility seemed to occur in a smooth fashion, which has led us to suggest that this mutual process emerges from the interaction dynamics, and not from the need or preferences of a single interaction partner. Being focused on the interaction and the child's understanding during individual instruction sessions might therefore be enough to realize fading and the transfer of responsibility, without being explicitly aware of it.

The second point we want to address here, is the importance of scaffolding to characterize children's development over time. In chapter 6, the variables that were best at predicting students' progress over time in terms of their cluster membership were mostly variables that were derived from the person-context dynamics, and specifically the researcher's scaffolding techniques. For example, the cluster that varied most in its mean proportion of high understanding levels over the five visits could also be characterized by a lower number of follow-up questions and a higher number of clarifications from the researcher, variables derived from the codings of the scaffolding behavior of the researcher during the tasks.

A last point we want to address is related to our findings in chapter 4 of this dissertation, illustrating the positive effect of adaptive scaffolding techniques, which is in line with the results of several other studies (Van de Pol et al., 2010). The chapter shows that special needs students' performance on the scientific tasks was similar to the performance of the regular students, even though their standardized learning achievement test scores (math and language) were significantly lower. This means that the gap that exists between regular and special needs students' academic performance almost completely disappears when assessing students' understanding in a situation in which scaffolding and monitoring the student's learning process is allowed. This result is in accordance with the notion of functional and optimal levels of performance (Fischer, 1980; Kitchener, Lynch, Fischer, & Wood, 1993), of which the latter reflects performance under adequate help and assistance. Below we explore possible reasons why standardized tests might not serve the special needs student population that well.

7.2.3 Linking this study's results to trends in society: The pitfalls of standardized testing

When it comes to academic performance, special needs students almost always score below their peers in regular education in various domains (Steenbeek, Jansen, & Van Geert, 2012). This has led researchers to suggest that these schools, now heavily focused on student's behavioral problems, should invest more in their content-related educational program (Trout et al., 2003; Van der Worp-Van Der Kamp, Pijl, Bijstra, & Van Den Bosch, 2013). While there is no doubt that an optimal balance between a focus on behavior and a focus on academic skills would be beneficial, the instruments to measure the academic performance of this population may not be flawless, given the so-called issue of construct-irrelevant variance in standardized testing. This means that the interpretations of test scores and the implications attributed to these are likely to be ‘contaminated’ by certain characteristics of the student population taking the test (Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999; Messick, 1989). In other words: The standardized test does not only measure the construct it claims to measure (such as math or science skills), but the final score is highly influenced by certain other factors that may interfere. These factors include students’ concentration and attention problems, as well as communication problems, such as difficulties to interpret questions, or to verbalize or write down answers (Cooper at al., 2004). Although this affects a wide range of students, the validity of these assessment tools is particularly in question in the case of special needs students (Dolan, Hall, Banerjee, Chun, & Strangman, 2005; Haladyna & Downing, 2004), who in particular suffer from attention and communication problems. Another source of construct-irrelevant variance in this population are social and test anxiety symptoms, which interfere with their ability to score optimally on standardized tests. For instance, one of the special needs students in our study said: “I’m dumb you know, that’s why I go to this [special] school”. To conclude, the fact that the special needs students score low on standardized tests in multiple studies (Trout et al., 2003; Lane et al., 2008) might mean nothing more than the mere fact that they score low *on these particular tests*, and might say less about their academic capabilities than researchers, policy advocates, and educators assume.

In the Netherlands, all regular primary schools are currently obliged to participate in a ‘pupils monitoring system’ to regularly evaluate students’ academic performance (math and language skills) with standardized tests. From August 2014 on, special educational settings will also be obliged to evaluate their students in the same way (Van Bijsterveldt-Vliegenthart, 2011). While it is in itself a good idea to track children’s academic development, the question is if these tests completely measure what they claim, despite the attempts of the test provider to adjust the standardized tests to the special student population (i.e. adding fewer questions; grouping questions about similar topics together—Cito, 2012).

The claimed advantage of these standardized tests is that they provide an “unbiased” record of the student’s progress over time, which helps us notice when the student falls behind (Cito, 2012). In addition, the tests would help teachers by signaling what is important to teach, would motivate students and teachers to work harder, and make instruction better (Amrein & Berliner, 2002). The disadvantages of standardized testing, however, might rule out these advantages. Besides the construct-irrelevant variance that accompanies them, there is also the danger that students will be taught specifically to these tests, especially if test scores become an important tool to assess individual students or schools (Amrein & Berliner, 2002; Kohn, 2000; Koretz, 2009). If schooling starts to resemble test-training, the material students encounter is limited, and it becomes questionable whether the students—and society in general would benefit from this. Amrein and Berliner (2002) call this the Heisenberg effect: “The more important any quantitative social indicator becomes in social decision-making, the more likely it will be to distort and corrupt the social process it is intended to monitor” (p.5). For example, if schools’ average test scores are made public, they can eventually become stigmatized as low-performing schools. This may lead them to find ways to improve students’ scores, for example by intensive test-training in the classroom (Kohn, 2000), or even preventing students with learning disabilities or a low socioeconomic status from admission.

It is clear that the standardized tests aimed to track students’ progress have important negative consequences (Kohn, 2000; Koretz, 2009). Still, testing is an integral part of schooling nowadays, and strongly supported by public policy. If we want to eliminate their disadvantages as much as possible, we might be better served with adaptive, universally designed testing methods. (Note, however, that this would not completely solve the problem posed by dynamic systems theory that it is doubtful whether a particular sampling context in the form of a standardized test tells us something worth knowing about other sampling contexts, such as a children’s cognitive functioning in their actual school context.) The term universal design comes from architecture (Mace, Hardie, & Place, 1991), and stands for the design of buildings that are usable for all people, without the need for additional adaptations (e.g., for people in a wheelchair). In education, the term is used to describe an educational view in which the presentation of information and the options for students to demonstrate their knowledge and skills is made flexible. It reduces barriers in educational material and instruction by providing accommodations and supports for all students, including students with disabilities or developmental delays (Pisha & Coyne, 2001; Rose & Meyer, 2002). In this way, universally designed education materials closely resemble the researcher's position during the scientific tasks in our study, who provided (e.g.,) clarifications, encouragement, and broke down questions in terms of complexity when needed.

Applying the universal design principles to tests would still enable us to track students' performance within a certain field, but under a condition that profoundly diminishes construct-irrelevant variance (Dolan et al., 2005). Temporal support structures would be available for all students, minimizing the chance of failure due to the testing circumstances. For example, computerized universally designed tests would contain text to speech software (cf. Dolan et al., 2005), a build-in dictionary to help students understand the wording of the questions (cf. Thompson, Johnstone, & Thurlow, 2002), and the option to break down the questions into smaller components.[39] Students would be able to either type in their answers, or record their verbal answers or even their actions with a build-in camera. Subsequently, teachers can track students' progress while they work through the program. In the case of multiple choice questions, it is even possible to let the computer program assess students' performance automatically, and use this to adaptively select the following item. That is, if the student's answer is wrong, the program can select a less complex question; see for example the computer program Math Garden (cf. Gierasimczuk, Van Der Maas, & Raijmakers, 2012).

---

[39] Note that this differs from a dynamic testing method, which aims to measure students' learning potential in a particular domain by testing repeatedly and giving feedback *after* each test (Lidz, 1991; Sternberg & Grigorenko, 2002). In the universal design situation, temporal support structures will be available *during* the test.

7.3 Possible limitations of this study, its implications and future directions

7.3.1 Limitation 1: Have we truly studied talent?

This dissertation is part of the Curious Minds research project, which focuses on young children’s talent for STEM fields. To be more specific, this dissertation concentrates on the development of young children’s understanding of the scientific concepts air pressure and gravity, in interaction with hands-on tasks and with guidance of an adult. This means that this study is explicitly focused on a small part of the extensive Curious Minds definition of talent (Van Geert & Steenbeek, 2007): “Talent is a child’s capacity to (ultimately) reach a high level of performance in a specific domain. Characteristics are: a high learning potential; the ability to elicit high-quality support from the (social) environment; in-depth processing of domain-specific information; creativity; belief in one’s own competence; enthusiasm, and a strong intrinsic motivation to learn” (p. 4). Time constraints prevented us from incorporating all aspects of this definition to our data coding and analysis. We could therefore conclude that the answer to the question “Have we studied talent?” is negative, if we focus on the complete definition of talent. That said, although the outcome measures of this study do not fully reflect this definition, the setup of the research project does incorporate more of its aspects. First, the study focused on the real-time ongoing process of understanding over a longer period, rather than on specific outcomes of children’s learning processes, such as grades or test scores. Second, we studied children while interacting with the proximal environment, that is, the tasks, and the researcher who administered these. In this way, we tried to establish the zone of excellent functioning in individual children by providing them with scaffolding that also increased their motivation, and by asking questions that were aimed at examining their creativity in creating explanations of new phenomena. Lastly, the study was explicitly set up in a prospective way, by studying children at an age at which they are in the midst of developing their scientific skills, and have not necessarily reached an exceptionally high level of scientific reasoning yet.

By adopting a prospective approach, we have not specifically targeted ‘excellent’ children, that is, those children who already showed a high level of scientific understanding at a young age. One could therefore say that this study is not about talent development, but more on young children’s development of scientific skills *in general*, and in particular their understanding of air pressure and gravity. However, the opposite—recruiting children with high abilities in the domain of science—would not be in concordance with our dynamic emergent view of talent (see chapter 2). According to this view, talent is emergenic (based on the interaction of several personal properties), epigenetic (a different onset for the development of these properties, and inter-individual differences in the property configuration), and dynamic (depending on iterative child-context interactions and chance—Simonton, 1999; 2001). The interaction between the child’s and the environment’s characteristics may (or may not) cause an upward spiral, making it hard to predict when a relatively outstanding performance in a certain area becomes observable.

Humans are dynamic, open living systems (Yun Dai & Renzulli, 2008), and change over time in interaction with the context. In talent development, a relatively high performance at a young age does not determine the child’s further developmental trajectory. The performance can even decline over time, due to multiple interactions between child- and context characteristics, or other children can catch up (Simonton, 1999). Excellent performance or commitment at a young age is also not a prerequisite for the development of talent. For example, Moesch and colleagues (2011) found that in sports most of the elite (i.e., talented) athletes in their study specialized during their late teenage years, and trained less in early childhood. Hence, if talent scouts only select the high-performing committed young children to work with, other children equally capable of reaching a high level of performance at a later age may miss out. Another point of caution when it comes to talent scouting at a young age is the well-known relative age effect (Helsen, Starkes, & Van Winckel, 2000). This entails that for some students a high performance at a young age could very well be attributed to their relatively older age (being born early in the academic year) compared to their classmates, and not necessarily to a difference in capacity. At a young age, a difference of a few months between two dates of birth may cause considerable differences in terms of height, attention span, or emotion regulation, to name a few. This, in turn, may lead to a better (perceived) performance compared to the child's younger peers, and this contrast might be falsely attributed to a difference in capacity.

The notion that children are dynamic, changing systems who develop in ongoing interaction with the environment, and that early high performance is no prerequisite for the further development of talent, could serve as an advice for teachers, parents, and policy advocates. The question arises if recruiting children for talent programs at a very young age would serve the children and the program best, given that the recruited children may not develop in the way the program expects them to, as developmental pathways are highly idiosyncratic and variable. Spurts, drops, and stable periods can occur, and are in fact part of a healthy developing system (cf. Van Dijk & Van Geert, 2002). In addition, recruiting at a young age might mean that other capable children who are currently not showing a high performance, or experience a temporary drop in performance, are missing out (Van Geert, 2011a; see also Gladwell, 2009). A selection based on a single test score and talent programs with a single, small recruitment time window may therefore not be the best way to recruit all talented children.

7.3.2 Limitation 2: Is this study truly dynamic?

Some of the outcomes discussed in this dissertation do not explicitly relate to dynamic properties of the understanding process. Nevertheless, the results did emerge from the dynamic properties that are part of the understanding process as observed during the tasks. As outlined in chapter 3, these properties are characterized as intertwining person-context dynamics, iterativeness, intra-individual variability, and interacting time scales. Throughout the rest of this dissertation, these dynamic properties have been the underlying basis of the descriptions and analyses in the three empirical chapters. Let us illustrate this point.

In chapter 4, we compared the regular students with the special needs students in our study. Although this chapter does not zoom in on the microgenetic codings of understanding for individual children—but instead takes aggregated measures, such as the mean skill level—the fact that we microgenetically coded the data might have influenced the outcome of this chapter. That is, the lack of finding a considerable difference between the regular and special needs students might be due to the fact that we did not simply take one outcome measure at the end of the task, but took several microgenetic codings of children's understanding during the interaction with the researcher and task. In this way, the aggregated measures reflect the learning process and the interaction more than a single score would do. Chapter 5 specifically targets the intertwining person-context dynamics and the interacting time scales, albeit in an exploratory manner. It shows how the interaction between the complexity levels of the researcher and a 4-year old boy takes shape during the tasks, and how this changes over the course of 3 visits. The researcher and boy are engaged in a dynamic 'dance', in which the researcher not just directs the in- and decreases in complexity level of the dyad, but also reacts to whatever the boy is doing in response to what the researcher initiates, and eventually starts following the boy's lead.

Lastly, chapter 6 shows how we can characterize children's developmental trajectories of understanding the scientific tasks over time, in terms of both their shape and their predictors. For this, we used data mining techniques, adding a large number of variables derived from the interaction dynamics, information of the children's home environment, and other more general measures. The trajectories of the three distinct clusters we found in this chapter could not be sufficiently explained by the general measures, such as standardized test scores, the age, or the gender of the child. Instead, factors that did matter were the variables that reflected the person-context dynamics during the tasks as well as at home.

We could not have done this study without a coding system that enabled us to capture ongoing changes in understanding levels. Skill theory (Fischer, 1980; Fischer & Bidell, 2006) provided a ruler to measure each task-related utterance, capturing the child's understanding of the two different tasks in a comparable way. Measuring with this ruler means that we can extract the reasoning complexity from its content. That is, if coded in the right way, a sensorimotor mapping level on a gravity task is equal to a sensorimotor mapping level on the air pressure task, given that both require the child to couple two single characteristics of the task into a meaningful structure of understanding. In addition, the underlying principles that skills are dynamic and encompass both the characteristics of the person as well as the context; that they can be highly variable on the short-term time scale, and that they can be coupled to the longer time scale of development, made this theory well suited for this study's longitudinal setup and its microgenetic codings of understanding.

7.3.3 Limitation 3: Is this study representative?

The number of children that participated in this study was somewhat small: 32 children in total, divided in two different subgroups depending on their school type (regular, *n* = 17; special, *n* = 15), and then divided in 3 small age groups (*n* = 4 – 7). This small number of participants was inevitable, given that both the data collection as well as the microgenetic data coding was time-consuming. We have therefore used the term 'multiple case study approach' to describe our sampling. Since a long time, researchers have argued that such case studies lack scientific value. The main objections are that one cannot generalize from a small sample, and that case studies leave too much room for the researcher's interpretations and are therefore quite subjective. However, as Flyvbjerg (2006) notes, studies aimed to generalize may be overvalued in the social sciences as being the only source of scientific development, and the researcher's choices of the categories, variables, and questionnaires in large quantitative studies can be equally subjective. Given that in large quantitative studies the researcher does not get as close to the participants as in the case of small *N* studies, this subjectivity is less likely to be corrected by the researcher, colleagues, or by the participants themselves while interacting with the researcher.

This does not mean that research using large random samples is by definition flawed. To the contrary, these studies can answer important questions, and can reveal group characteristics, differences between groups, or the general effect of interventions. Small *N* studies on the other hand, like this study, enable us to study patterns, differences, or effects for individuals in depth. This point is related to the ergodicity problem, which says that statistical relationships captured from comparing data *between* large groups of individuals, are in general not directly applicable to statistical relationships concerning data *within* individuals (Molenaar & Campbell, 2009). Hence, large random samples of individuals can never be a replacement for process related studies that are individual-based, which are in practice usually limited to relatively small numbers of participants.

The fact that these studies focus on individual cases, does, however, not mean that we cannot generalize. Flyvbjerg (2006) claims that generalization is possible, if the cases are well-chosen, for example when these are extreme cases, prototype exemplars of the population, or when taking multiple cases that have different characteristics. This study is an example of the latter, in which our subsamples differ with respect to the age of the participants and their type of school.A last claim Flyvbjerg (2006) makes, is that case studies are an important source of concrete context-dependent knowledge, which is important in the social sciences, as general context-independent theories that explain human behavior are hard to come by. This argument is in concordance with what we argued before about how knowledge is always constructed in interaction with the context[40]. In other words, this context-dependence not only applies to the study of knowledge construction of individuals, as we did in our study, but also to the knowledge construction within the social sciences.

---

[40] This is what we called the dynamic embedded view of conceptual knowledge in chapter 2.

7.3.4 Limitation 4: Can we translate the findings and the setting of this study to practice?

When doing a study in the social sciences, it is an important question whether there is ecological and external validity, that is, whether we can translate the study's setting and its findings to the real world. Is it indeed possible to translate this study's setting—individual children working together with a researcher on hands-on scientific tasks for about 30 minutes per session—to the current educational practice? The answer is that it depends, partly on the organization of the classroom. Indeed, the number of hours per student are limited nowadays, as teachers have to attend to bigger groups of students (AOb/ITS, 2013). However, in the last decades, teachers in the Netherlands have started to divide the classroom into small groups that work together on projects, receive extra instruction for a particular subject, or get extra challenging materials (Terwel & Van den Eeden, 1992). Although the individual setting of the current study is probably not easy to translate to current educational practice, an adaptation to small group work is.

Other researchers have begun to study the effects of these Curious Minds small group settings. Using video feedback coaching and hands-on scientific tasks similar to the ones used in this study, they assist teachers in how to construct inquiry-based science lessons (Menninga, Van Dijk, Steenbeek & Van Geert, 2013; Van Vondel & Steenbeek, 2014; Wetzels, Steenbeek, & Van Geert, 2013). In these lessons, small groups of children work together on a hands-on task, after getting instructions from the teacher, or a worksheet with the most important points they have to address, usually corresponding to the empirical cycle (describing, predicting, testing, and explaining—De Groot, 1969). While the groups work together, the teacher walks around in the classroom, assists the students when needed, provides scaffolding, and asks additional questions. The video feedback coaching program is specifically directed at the ways teachers ask questions. They are advised to formulate open-ended questions, and to let the students explore their ideas without prompting them with the answer. So far, this Curious Minds setting seems well applicable to the classroom.

Can we translate the findings of this study to educational practice? We have already argued that a small sample size does not necessarily mean that we cannot generalize our findings. First, the finding that special needs students can reach similar levels of understanding while working on hands-on tasks under a condition of scaffolding, can have important implications for practice. Although earlier studies found a significant gap between these children and their peers in regular education (Trout et al., 2003; Van der Worp-Van Der Kamp et al., 2013), we found no meaningful differences when adapting the proximal context to their needs using scaffolding techniques. Of course one could argue that this does not prove the absence of a difference in the group's abilities. For example, the researcher could have provided more scaffolding to the children in the special education group. Although we have checked this and have not found a difference at the group level,[41] scaffolding is a dynamic idiosyncratic construct, and differences between the numbers of scaffolding instances for individual researcher-child dyads do exist. However, even if the number of scaffolding instances was higher for some students in the special education group, this does not disprove our findings that special needs students *can* have a similar performance when scaffolded to their individual needs. In general, special educational schools in the Netherlands have less students and more teachers per classroom (AoB/Its, 2013). Given that these students seem to benefit from a condition with scaffolding[42], it may be fruitful to examine how the Curious Minds small groups setting can be implemented in special education, maybe not only to teach science, but other academic areas as well. If we also provide tests that are adaptive and universally designed to test this teaching method (Rose & Meyer, 2002; Pisha & Coyne, 2001), we might get a better picture of special needs students' abilities.

A second finding stemming from this research is that teachers and parents have an important influence on children's performance. Although this view is not

---

[41] Using the scaffolding categories of chapter 6 (proportion of follow-up questions, content-related questions, and clarifications per task, per visit), we found no statistically significant differences in favor of either the regular or the special needs students.

[42] Note that the special needs students in this sample had significantly lower standardized test scores.

new, the influence may be bigger than teachers and parents think. Understanding does, according to our view, not reside in the head of the learner, but rather gets formed in a dynamic, bi-directional interaction with the context. Teachers and parents can structure this interaction by providing the right materials, scaffolding and tests when needed. Sensitivity to the child's needs and creating a supportive environment (Christophel, 1990), in which the child feels confident to demonstrate his or her (lack of) understanding, is therefore key.

7.4 Future directions

We want to end this dissertation by calling for future research. Part of this research has already begun, in the projects of Wetzels (2013), Van Vondel (2014), and Menninga (2013) and their colleagues, by studying the effects of Curious Minds small group settings, using video feedback coaching. There might be a possibility to implement the Curious Minds approach even further, by constructing universally designed educational computer programs that could help students to practice with STEM content, by simulating the effects of materials, with added scaffolding possibilities, such as text-to-speech, or feedback from a virtual teacher. The Center for Applied Special Technology in the United States has already begun to build and explore such universally designed media-rich learning environments to teach science (Rappolt-Schlichtmann et al., 2013), and in the Netherlands an adaptive computer program exists in which children work on math problems (Gierasimczuk et al., 2012). In line with this, we also want to call for research on the application of adaptive, universally designed tests, based on the work of Thompson and colleagues (2002) and Dolan and colleagues (2005). As we mentioned before, these tests would enable us to track students' performance, but under a condition that profoundly diminishes construct-irrelevant variance.

As pointed out in chapter 2, this dissertation adds to the knowledge we have about young children's understanding of scientific concepts, and may eventually help to construct effective inquiry-based science lessons for young children. These lessons, in turn, can possibly stimulate the STEM knowledge and careers of the future student population. This dissertation has shown that when it comes to children’s abilities in STEM fields, tasks that elicit children’s enthusiasm and the support of an adult (teacher, researcher, or parent) during these tasks are of tremendous importance. Or, in Albert Einstein’s (1879 - 1955) words: “The point is to develop the childlike inclination for play and the childlike desire for recognition, and to guide the child over to important fields for society. Such a school demands from the teacher that he be a kind of artist in his province.”

Chapter 8: Nederlandse Samenvatting (Summary in Dutch)

Dit proefschrift richt zich op de vraag: Hoe ontwikkelen jonge kinderen (3-5 jaar oud) hun begrip van wetenschappelijke concepten op de lange termijn in interactie met de sociale en materiele omgeving en zijn kinderen uit het speciaal onderwijs (cluster 4, met gedrags- en/of psychische problemen) in staat om hun begrip te ontwikkelen op hetzelfde niveau? Om specifiek te zijn, richtten we ons op hoe het begrip van individuele kinderen (*n* = 32) zich ontwikkelde tijdens praktische wetenschappelijke taken waarin de wetenschappelijke concepten zwaartekracht/inertie/snelheid en luchtstroming/luchtdruk waren geïntegreerd en waarbij we de interactie met de taak en de onderzoeker die de taak afnam, in de analyse meenamen. Tijdens de afname maakte de onderzoeker gebruik van een adaptief protocol, waarbij het kind op een natuurlijke manier door de taak geleid werd door middel van een aantal beschrijvings- voorspellings- en verklaringsvragen. Het protocol bood de onderzoeker ruimte om ondersteuning (‘scaffolding’) te bieden en gaf de kinderen de mogelijkheid om op een actieve onderzoekende manier te leren door verwachtingen op te stellen, bewijs te vergaren en de bevindingen te verklaren. In totaal werden de kinderen over het verloop van drie jaar 10 keer bezocht om steeds twee wetenschappelijke taken te doen. Dit proefschrift is gericht op de eerste anderhalf jaar van deze studie (5 taken).

Om een nauwkeurig beeld te krijgen van de ontwikkeling, hebben we gekozen voor een procesbenadering (process approach). Dat betekent dat we het begrip van kinderen op een microgenetische wijze gecodeerd hebben gedurende de interactie met de taak en de onderzoeker, waarbij we gebruik hebben gemaakt van een codeersysteem dat gebaseerd is op skill theory (Fischer, 1980) om de complexiteit van de uitingen te bepalen. Hierbij werd uitdrukkelijk de dynamiek tussen het kind en de context meegenomen. Naast de video-opnames en bijbehorende coderingen, hebben we ook achtergrondinformatie over de kinderen verzameld met behulp van vragenlijsten die door de ouders werden ingevuld en testuitslagen van de taal- en rekentoetsen uit het leerlingvolgsysteem van de scholen.

Deze dissertatie is gericht op een onderzoekslijn waarbij elk hoofdstuk ofwel gerelateerd is aan een specifiek gedeelte van dit longitudinale onderzoek, ofwel een specifiek perspectief op de gegevens biedt. Samen geven deze delen een idee van hoe het begrip van kinderen zich ontwikkelt terwijl zij werken aan praktische (‘hands-on’) wetenschappelijke taken met ondersteuning van een onderzoeker, zowel op de korte termijn (tijdens een taak) als op de lange termijn (over het verloop van meerdere taken). In deze samenvatting zullen we nu de afzonderlijke hoofdstukken en bijbehorende bevindingen bespreken.

8.1 Theoretisch kader en de eerste kwalitatieve data (h. 2 en 3)

Hoofdstuk 2 is gericht op de opzet van dit longitudinaal microgenetisch onderzoek. De theoretische en praktische grondslagen worden behandeld en we geven een uitgebreide beschrijving van de deelnemers, materialen, de wijze van dataverzameling en het coderen. Dit hoofdstuk dient als een overzicht en kan gebruikt worden als referentie bij het lezen van de andere hoofdstukken. In hoofdstuk 3 wordt een theoretisch model over de ontwikkeling van begrip van wetenschappelijke concepten besproken, gebaseerd op een aantal kenmerken van complexe dynamische systemen en skill theory (Fischer, 1980). In het model wordt het verkrijgen van begrip (van bijvoorbeeld wetenschappelijke taken) beschouwd als een proces dat vorm krijgt door een dynamische interactie met de proximale omgeving. Het model kan zowel onderzoekers en onderwijzers ondersteunen door expliciet te maken hoe kinderen hun begrip in interactie met de omgeving ontwikkelen, waardoor de ontwikkeling van begrip transparanter wordt.

Vanuit een dynamisch perspectief kan het begrip van wetenschappelijke taken beschouwd worden als een proces van verstrengelde dynamische interacties tussen de leerling en de leraar (of onderzoeker, of ouder). Dat betekent dat ieder moment het (non)verbale gedrag van het kind invloed heeft op dat van de leraar en andersom, waarbij ze samen het volgende moment creëren in het leerproces (Fogel & Garvey, 2007; Steenbeek, 2006). Binnen de interactie ontstaat begrip door iterativiteit. Dit betekent dat elke staat van begrip gebaseerd is op een voorgaande staat van begrip, ingebed in de huidige context. Door deze dynamische iteratieve interacties met de omgeving zal het begrip van het kind (of beter: de complexiteit daarvan) fluctueren. Het kan verbeteren, maar ook tijdelijk verslechteren, bijvoorbeeld als de taakmoeilijkheid omhoog gaat, of als de ondersteuning van de leraar vermindert. Met andere woorden, er is intra-individuele variabiliteit in begrip. De iteratieve processen op de korte termijn tijdens een taak vormen de ontwikkeling van begrip op de lange-termijn (Lewis, 2000). Echter, de lange-termijn ontwikkeling zal ook het korte-termijn proces tijdens een taak beïnvloeden, doordat het de basis is van de onderliggende structuur en functie van de interactie (Lewis & Granic, 2000; Smith & Thelen, 2003; Steenbeek, 2006; Van Geert & Steenbeek, 2005a). Meer details over dit conceptuele model en een visuele interpretatie zijn te vinden in hoofdstuk 3.

8.2 Een cross-sectionele vergelijking (h. 4)

Hoofdstuk 4 is gericht op een cross-sectionele vergelijking van leerlingen uit het regulier en speciaal onderwijs (cluster 4, voor kinderen met gedrags- en/of psychische problemen) tijdens één bezoek. Eerder onderzoek liet zien dat kinderen uit het speciaal onderwijs slechter presteren dan kinderen uit het regulier onderwijs, mogelijk omdat hun emotionele en/of gedragsproblemen een optimale academische prestatie in de weg staan (zie Trout, Nordness, Pierce, & Epstein, 2003 voor een uitgebreid literatuuroverzicht). De focus van dit eerdere onderzoek lag vooral op de scores van deze kinderen op gestandaardiseerde toetsen. De vraag was hoe de kinderen uit het speciaal onderwijs hun begrip tijdens de twee wetenschappelijke taken zouden ontwikkelen terwijl zij hieraan werkten met een onderzoeker, die vragen stelde vanuit een adaptief protocol waarin ruimte was voor ondersteuning (scaffolding). Zouden deze kinderen profiteren van deze setting?

Als de kinderen uit het speciaal onderwijs slechter zouden presteren dan de reguliere studenten, zouden we verwachten dat zij meer lage complexiteitsniveaus zouden laten zien. De resultaten laten zien dat hoewel de kinderen uit het speciaal onderwijs gemiddeld meer fouten maakten, hun gemiddeld aantal antwoorden en het gemiddelde complexiteitsniveau niet verschilde van dat van de kinderen uit het regulier onderwijs. De kinderen in het speciaal onderwijs lieten een hogere proportie van het laagste complexiteitsniveau (niveau 1 sensorimotor action) zien, maar de kinderen uit het regulier onderwijs hadden een hogere proportie van het op één na laagste niveau (niveau 2 sensorimotor mapping). De reguliere studenten hadden een hogere proportie van uitingen op niveau 3 (sensorimotor system), maar dit kwam vooral doordat er een significant verschil was tussen de driejarige reguliere en speciale studenten. De vier- en vijfjarige groepen verschilden niet van elkaar wat betreft de proportie antwoorden op niveau 3. Op de hoogste niveaus die gevonden werden in deze studie (niveau 4—single representation en 5—representational mapping) verschilden de twee groepen niet. Deze resultaten zijn in tegenspraak met wat eerder onderzoek heeft gevonden en in tegenspraak met het significante verschil tussen de twee groepen dat wij vonden op de taal- en rekentoetsen van het leerlingvolgsysteem. Ondersteuning in de vorm van scaffoldingtechnieken die aansluiten bij de studenten zou daarvoor van cruciaal belang kunnen zijn voor het begrip van kinderen tijdens de wetenschappelijke taken en wellicht ook voor hun academische prestaties in het algemeen.

8.3 Een diepte-analyse van drie opeenvolgende interacties (h. 5)

Nadat we hebben gekeken naar de groepsverschillen, is hoofdstuk 5 gericht op een diepte-analyse van de interacties tussen een vierjarige jongen en de onderzoeker tijdens de luchtdruktaken van drie opeenvolgende bezoeken. De focus van dit hoofdstuk is op de interactiedynamiek tussen de jongen en de onderzoeker, waarbij wij ons in het bijzonder richten op de fluctuaties in complexiteitniveau van zowel de uitingen van het kind als de vragen van de onderzoeker. De resultaten laten zien dat de jongen tijdens de eerste taak fluctueerde in zijn begrip. Deze fluctuaties waren gelijk verdeeld over de hele duur van de interactie, dat wil zeggen dat er geen verschillen werden gevonden in aantallen fluctuaties tussen de eerste en de tweede helft van de interactie. Over het verloop van 3 sessies zagen we een toename in het aantal goede antwoorden van de jongen en verschoven de frequenties van de complexiteitsniveaus: het aantal niveau 2 (sensorimotor mapping) uitingen steeg tijdens de tweede sessie, terwijl het aantal niveau 3 uitingen omlaag ging. In de derde sessie ging dit precies de andere kant op en waren de verhoudingen tussen de frequenties weer gelijk aan die van de eerste sessie. Wat betreft de interactie, zagen we dat de jongen tijdens het eerste bezoek meestal de stijgingen en dalingen in het complexiteitsniveau van de vragen van de onderzoeker volgde. Over verloop van tijd initieerde de jongen significant meer van deze gezamenlijke stijgingen en dalingen in complexiteitsniveau, terwijl de onderzoeker steeds minder initiatief hierin nam. In de derde sessie was het gemiddelde complexiteitsniveau in de uitingen van de jongen significant hoger, terwijl hij ook meer initiatief nam in de gezamenlijke stijgingen en dalingen in het complexiteitsniveau van de dyade. Dit illustreert de ideeën uit de theorie van 'self-regulated learning' (Zimmerman, 1990). Vanuit deze theorie zou een actieve houding van studenten wat betreft het selecteren en structureren van hun eigen leerproces hun academisch succes vergroten.

8.4 Ontwikkelingstrajecten in wetenschap en techniek (h. 6)

Na het bekijken van de groepsverschillen in hoofdstuk 4 en de diepte-analyse van de interactiedynamiek in een individueel ontwikkelingstraject in hoofdstuk 5, is de vraag die in hoofdstuk 6 wordt gesteld: Hoe kunnen we de ontwikkelingstrajecten van het begrip van kinderen van de wetenschappelijke concepten zwaartekracht en luchtdruk over het verloop van anderhalf jaar (5 bezoeken) karakteriseren? We kijken hierbij zowel naar de vorm van deze trajecten als naar de voorspellende factoren die met deze trajecten samenhangen. Een groot aantal variabelen werd bij dit onderzoek betrokken: allereerst het complexiteitsniveau van de taakrelevante uitingen van de kinderen, om te bekijken hoe het ontwikkelingstraject over tijd verloopt, maar daarnaast ook het gedrag van de kinderen tijdens de taken (initiatiefname, inhoudelijke uitingen en off-task uitingen), de ondersteuning van de onderzoeker (scaffolding-technieken als inhoudelijke vragen, follow-up vragen en verduidelijkingen), en informatie afkomstig uit de vragenlijsten en de taal- en rekentoetsen van het leerlingvolgsysteem. Een clusteranalyse gebruikmakend van 10 proporties per kind (één voor elke taak tijdens elk van de vijf bezoeken) van de hoogste complexiteitsniveaus (niveau 4 en hoger), leverde drie groepen met een verschillend ontwikkelingstraject op.

Cluster 1 had de hoogste proportie van hoogste complexiteitsniveaus op alle vijf luchtdruktaken en vijf zwaartekrachttaken. Cluster 2 was variabel en liet afwisselend hoge en lage proporties van deze hoge complexiteitsniveaus zien. Het laatste cluster had veelal lage proporties van de hoogste complexiteitsniveaus, die redelijk stabiel waren over tijd. Alle groepen lieten een significante vooruitgang zien op de luchtdruktaken. Hoewel er ook vooruitgang was op de zwaartekrachttaken, liet alleen cluster 1 op deze taken een significante stijging in de hoge complexiteitsniveaus zien. De variabelen die het minst samenhangen met de clusterindeling waren demografische (leeftijd, geslacht, diagnose) en schoolvariabelen (schooltype, prestaties op gestandaardiseerde taal- en rekentoetsen). De variabelen die de hoogste associatie hadden met de clusterindeling kwamen uit de interactie tussen de onderzoeker (scaffolding-technieken als inhoudelijke vragen, follow-up vragen en verduidelijkingen) en het kind (initiatiefname, inhoudelijke uitingen en off-task uitingen) terwijl zij aan de praktische wetenschappelijke taken werkten. Variabelen die daarnaast een sterke associatie hadden met de clusterindeling, reflecteerden de interacties van de kinderen en hun ouders in de thuissituatie (aanmoediging van de ouders met betrekking tot het spelen met constructiespeelgoed, het delen van schoolervaringen met elkaar). Deze zogenaamde interactie- en thuisvariabelen representeren de interacties tussen het kind en zijn of haar directe omgeving, zowel thuis als tijdens de taken. Dit illustreert ons eerdere punt dat de context een zeer belangrijke rol speelt in de ontwikkeling van begrip en niet gezien kan worden als een eenmalige of eenzijdige invloed van buitenaf.

8.5 Conclusie

In dit proefschrift wordt leren beschouwd als een voortdurend proces tussen een kind en zijn directe omgeving, waarbij de complexiteit van het begrip dat kinderen hebben van wetenschappelijke concepten groeit binnen deze specifieke context. In dit proces zijn er verschillende mechanismen die de toe- en afname in begrip beïnvloeden. Deze variabiliteit is een intrinsiek deel van het leerproces en komt op een natuurlijke manier voor in interactie met de omgeving. Deze voortdurende interactie met de omgeving is dus van groot belang voor het leerproces en kan hier niet los van worden gezien. Dit wordt, onder andere, geïllustreerd in hoofdstuk 4, waarin blijkt dat het onderwijzen in deze natuurlijke interacties samenhangt met een toename in de complexiteit van het begrip van kinderen uit een populatie die gekenmerkt wordt door emotionele en gedragsproblemen en lagere academische prestaties op gestandaardiseerde toetsen. Dit komt mogelijk doordat gestandaardiseerde toetsen niet alleen de prestaties meten waar zij voor gemaakt zijn, maar ook een aantal andere bijbehorende constructen, zoals aandacht en woordenschat.

Eén van de manieren waarop we de interactie met de directe omgeving kunnen bekijken tijdens het leerproces, is door te kijken naar de relatie tussen de ondersteuningstechnieken (scaffolding) van de docent en de veranderingen in het begrip van de leerling. Uit dit proefschrift blijkt dat het gebruik van deze technieken in leraar-leerling interacties op een natuurlijke manier uit de interactie voortvloeit en ook afgebouwd wordt naarmate de leerling minder behoefte heeft aan een specifieke vorm van ondersteuning. In de case study die we gedaan hebben (hoofdstuk 5), was zichtbaar dat het kind na verloop van tijd zelf meer initiatief nam en dat de onderzoeker, zonder vooropgezet plan, hierbij van een leidende naar een volgende rol ging. Daarnaast is gebleken dat het gebruik van deze technieken een zeer belangrijk verband heeft met de vooruitgang van kinderen op de lange termijn (hoofdstuk 6). In hoofdstuk 7 behandelen we deze en een aantal andere discussiepunten die geïnspireerd zijn op de achtergrond en de resultaten van deze studie.

8.6 Praktische implicaties van dit onderzoek

Dit onderzoek laat zien dat de interactie tussen leraar, taak en leerling vervlochten is en dat de afzonderlijke componenten niet los van elkaar gezien kunnen worden. Investeringen in wetenschap en techniekonderwijs zouden dus in de eerste plaats gericht moeten zijn op de professionalisering van leraren op dit gebied en het ontwikkelen van onderwijsmaterialen die kinderen in staat stellen op een praktische manier wetenschappelijke concepten te ontdekken. Hierbij moet worden opgemerkt dat de professionalisering van leraren niet per definitie gericht moet zijn op hun feitelijke kennis van wetenschap en techniek (mits zij basiskennis hebben van fundamentele wetenschappelijke concepten), maar op hoe zij kinderen binnen het wetenschap- en techniekonderwijs kunnen ondersteunen en hoe zij op de juiste manier kinderen vragen kunnen stellen tijdens het ontwikkelingsproces. Immers, het begrip van kinderen over wetenschap wordt gevormd door onder begeleiding met deze taken bezig te gaan. Aangezien de (wetenschappelijke) kennis van kinderen in interactie met de omgeving ontwikkelt, is het extra investeren in gestandaardiseerde toetsen om kennis te “meten” minder zinvol. Deze gestandaardiseerde situatie heeft weinig raakvlakken met de praktische situatie waarin kinderen leren. Resultaten van deze toetsen hebben daarom weinig samenhang met hoe kinderen over langere tijd zich ontwikkelen in een schoolcontext tijdens de praktische taken.

References

Albanese, A., Neves, M. C., & Vicentini, M. (1997). Models in science and in education: A critical review of research on students' ideas about the earth and its place in the universe. *Science & Education, 6,* 573-590. doi: 10.1023/A:1008697908361

Alfieri, L, Brooks, P. Aldrich, N., & Tenenbaum, H. (2010). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*, 1-18.

Allen, J. W., & Bickhard, M. H. (2013). Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate. *Cognitive Development, 28,* 96–133. doi: 10.1016/j.cogdev.2013.01.002

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention paired associates. *Journal of Verbal Learning and Verbal Behavior, 8*, 463-470. doi: 10.1016/S0022-5371(69)80090-3

Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & Lazarus, S. (2010). Technology-enabled and universally designed assessment: Considering access in measuring the achievement of students with disabilities—A foundation for research. *Journal of Technology, Learning, and Assessment, 10*. Retrieved 16 August 2013 from http://www.jtla.org.

American Psychological Association (2000). *Diagnostic and statistical manual of mental disorders, (4th ed., text rev.).* Washington, DC.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education policy analysis archives, 10*, 1-74.

AOb/ITS (2013). Groep XXL resultaten onderwijsblad- enquête 'Eén jaar Rutte II' over ontwikkeling klassengrootte. Retrieved from: http://www.aob.nl/doc/GroepXXL03102013.pdf

Ayboub, C. C., O'Connor, E., Rappolt-Schlichtmann, G., Fischer, K. W., Rogosch, F. A., Toth, S. L., & Cicchetti, D. (2006). Cognitive and emotional differences in young maltreated children: A translational application of dynamic skill theory. *Development and Psychopathology*, *18,* 679-706. doi: 10.1017/S0954579406060342

Baartman, L. K. J., & Gravemeijer, K. (2011). Science and Technology Education for the Future. In M. J. De Vries, H. Van Keulen, S. Peters, and J. Walma Van Der Molen (Eds.). *Professional Development for Primary Teachers in Science and Technology: The Dutch VTB-Pro Project in an International Perspective* (pp. 21-33). Rotterdam, NL: SensePublishers. doi: 10.1007/978-94-6091-713-4_3

Baguley, T. (2012). Can we be confident in our statistics? *The Psychologist*, *25*, 128-129.

Bailin, S. (2002). Critical thinking and science education. *Science & Education, 11*, 361-375. doi: 10.1023/A:1016042608621

Baker, D. (2002). Where is gender and equity in science education? *Journal of Research in Science Teaching*, *39*, 659-663. doi: 10.1002/tea.10044

Barnes, K. A., Howard, J. H. Jr., Howard, D. V., Gilotty, L., Kenworthy, L., Gaillard, W. D., & Vaidya, C. J. (2008). Intact implicit learning of spatial context and temporal sequences in childhood autism spectrum disorder. *Neuropsychology, 22*, 563-570.

Bassano, D., & Van Geert, P. (2007). Modeling continuity and discontinuity in utterance length: A quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science, 10*, 588-612. doi: 10.1111/j.1467-7687.2007.00629.x

Belaïd, A., Moinel, T., & Rangoni, Y. (2010). Improved CHAID algorithm for document structure modeling. *Spie Proceedings 7534*, Document Recognition and Retrieval XVII. doi:10.1117/12.839794

Blijd-Hogewys, E. M. A. (2008). *The development of theory-of-mind and the theory-of-mind storybooks*. Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Bos, J. & Steenbeek, H. (2006). *Mediacoder: A Simple Application For Coding Behavior within Media Files.* University of Groningen, Groningen, The Netherlands*.*

Brotman, J. S., & Moore, F. M. (2008). Girls and science: A review of four themes in the science education literature. *Journal of Research in Science Teaching*, *45*, 971-1002. doi: 10.1002/tea.20241

Bryce, T. K., & Blown, E. J. (2007). Gender effects in children's development and education. *International Journal of Science Education*, *29*, 1655-1678. doi: 10.1080/09500690701278420

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental neuropsychology*, *33*, 205-228. doi: 10.1080/87565640801982312

Burkam, D. T., Lee, V. E., & Smerdon, B. A. (1997). Gender and science learning early in high school: Subject matter and laboratory experiences. *American Educational Research Journal*, *34*, 297-331. doi: 10.3102/00028312034002297

Bybee, R. W. (2010). What is STEM education? *Science, 329*, 996-996. doi: 10.1126/science.1194998

Carrière, R. M. (2009). *The transition to early education: A dynamic theoretical framework.* Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Cheshire, A., Muldoon, K. P., Francis, B., Lewis, C. N. & Ball, L. J. (2007). Modelling change: New opportunities in the analysis of microgenetic data. *Infant and Child Development, 134*, 119-134. doi: 10.1002/icd.498

Chiu, M. H., Lin, J. W. (2005). Promoting fourth graders' conceptual change of their understanding of electric current via multiple analogies. *Journal of Research in Science Teaching, 42*, 429-464. doi: 10.1002/tea.20062

Christophel, D. M. (1990). The relationships among teacher immediacy behaviors, student motivation, and learning*. Communication Education, 39*, 323-340. doi: 10.1080/03634529009378813

Cito (2009). *Primair onderwijs, Leerling- en onderwijsvolgsysteem. Taal voor kleuters groep 1 en 2.* Internal publication.

Cito (2012). Primair en special onderwijs. *Toetsen voor speciale leerlingen. Voor functioneringsniveaus groep 3 tot en met groep 8*. Arnhem: Cito group.

Coelho, R. L. (2009). On the concept of energy: How understanding its history can improve physics teaching. *Science and Education 18*, 961-983. doi: 10.1007/s11191-007-9128-0

Cooper, C. R., Baum, S. M., & Neu, T. W. (2004). Developing scientific talent in students with special needs: An alternative model for identification, curriculum, and assessment. *Journal of Advanced Academics*, *15*, 162-169. doi: 10.4219/jsge-2004-456

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing. A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11*, 671-684.

Dai, D. Y., & Renzulli, J. S. (2008). Snowflakes, living systems, and the mystery of giftedness. *Gifted Child Quarterly, 52*, 114-130. doi: 10.1177/0016986208315732

De Berg, K. (1995). Student understanding of the volume, mass, and pressure of air within a sealed syringe in different states of compression. *Journal of Research in Science Teaching, 32,* 871-884

De Groot, A. D. (1969). *Methodology.* Mouton.

De Lange, J., Feijs, E.J.G. & Uittenbogaard, W. (2007). Ik ben alleen nieuwsgierig. *Talent, 9*, 26-28.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* New York: Oxford University Press.

Dixon, J. A., & Bangert, A. S. (2002). The prehistory of discovery: Precursors of representational change in solving gear system problems. *Developmental Psychology, 38*, 918-933. doi: 10.1037/0012-1649.38.6.918

Dolan, R. P. and Hall, T. E. (2001). Universal Design for Learning: Implications for Large Scale Assessment. *IDA Perspectives, 27,* 22-25.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment, 3*. Retrieved from: http://www.jtla.org

Dryden, J. (1885). Epistle to Congreve. In W. Scott and G. Saintsbury (Eds.). *The works of John Dryden.* Vol. 11, (pp. 57-60). Edinburgh: Paterson (original work published in 1693).

Dunlop, J. (2000). How children observe the universe. *Publications Astronomical Society of Australia, 17*, 194-206. doi: 10.1071/AS00194

Ensing, A., Van Der Aalsvoort, G. M., Van Geert, P., & Voet, S. (2014). Learning potential is related to the dynamics of scaffolding, an empirical illustration of the scaffolding dynamics of five year olds and their teacher. *Journal of Cognitive Education and Psychology.*

Epstein, M. H., Kinder, D., & Bursuck, B. (1989). The academic status of adolescents with behavioral disorders. *Behavioral Disorders, 14*, 157–165.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100*, 363-406. doi: 10.1037/0033-295X.100.3.363

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological review, 87*, 477-531. doi: 10.1037/0033-295X.87.6.477

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In R. M. Lerner (Ed.) & W. Damon (Series ed.), *Handbook of Child Psychology: Vol. I. Theoretical Models of Human Development* (pp. 313–399). New York: Wiley.

Fischer, K. W. & Corrigan, R. (1981). A skill approach to language development. In R. Stark (Ed.), *Language behavior in infancy and early childhood,* (pp. 245-273). Amsterdam: Elsevier.

Fischer, K. W., & Granott, N. (1995). Beyond one-dimensional change: Parallel, concurrent, socially distributed processes in learning and development. *Human Development, 38,* 302-314.

Fischer, K. W., & Rose, S. P. (1999). Rulers, models, and nonlinear dynamics: Measurement and method in developmental research. In G. Savelsbergh, H. Van Der Maas, and P. Van Geert (Eds.), *Nonlinear developmental processes* (pp. 197-212). Amsterdam: Edita-KNAW.

Fischer, K. W., Rose, L. T., & Rose, S. P. (2007). Growth cycles of mind and brain: Analyzing developmental pathways of learning disorders. In K. W. Fischer, J. Holmes Bernstein, and M. H. Immordino-Yang, *Mind, brain, and education in reading disorders*, (pp. 101-123). NY: Cambridge University Press.

Flynn, E., & Siegler, R. (2007). Measuring change: Current trends and future directions in microgenetic research. *Infant and Child Development*, *16*, 135-149.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry, 12*, 219-245. doi: 10.1177/1077800405284363

Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science.* Brighton, Sussex, U.K.: The Harvester Press.

Fogel, A. & Garvey, A. (2007). Alive communication. *Infant Behavior and Development, 30*, 251-257. doi: 10.1016/j.infbeh.2007.02.007

Gagné, F. (1985). Giftedness and talent: Reexamining a reexamination of the definitions. *Gifted Child Quarterly, 29*, 103-112. doi: 10.1177/001698628502900302

Gardner, H. (1993). The relationship between early giftedness and later achievement. In: G. R. Bock & K. Ackrill. *The origins and development of high ability* (pp. 175-186). West Sussex, England: Wiley & Sons.

Garnett, P. J., Garnett, P. J., & Hackling, M. W. (1995). Students' alternative conceptions in chemistry: A review of research and implications for teaching and learning. *Studies in Science Education, 25*, 69-96. doi: 10.1080/03057269508560050

Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In P. Mussen (Ed.), *Handbook of child psychology*, (Vol 4, pp. 167-230). New York: Wiley.

Gibson, H. L., & Chase, C. (2002). Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science. *Science Education, 86*, 693-705. doi: 10.1002/sce.10039

Gierasimczuk, N., Van Der Maas, H., & Raijmakers, M. (2012). *Logical and psychological analysis of deductive mastermind.* Proceedings of the 24th European summer school in logic, language and information, ESSLLI 2012, 1-13.

Gilmore, C. K. & Bryant, P. (2008). Can children construct inverse relations in arithmetic? Evidence for individual differences in the development of conceptual understanding and computational skill. *British Journal of Developmental Psychology, 26,* 301-316. doi: 10.1348/026151007X236007

Gladwell, M. (2009). *Outliers: The story of success.* Penguin UK.

Goldin-Meadow, S., & Alibali, M. W. (2002). Looking at the hands through time: A microgenetic perspective on learning and instruction. In N. Granott and J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning.* (pp. 80-105). NY: Cambridge University Press.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: using the hand to read the mind. *Psychological review,100*, 279. doi: 10.1037/0033- 295X.100.2.279

Granott, N. (1993). *Microdevelopment of co-construction of knowledge during problem solving: puzzled minds, weird creatures, and wuggles* (Doctoral dissertation, Massachusetts Institute of Technology).

Granott, N. (2002). How microdevelopment creates macrodevelopment: Reiterated sequences, backward transitions, and the zone of current development. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 213-242). NY: Cambridge University Press.

Granott, N. (2005). Scaffolding dynamically toward change: Previous and new perspectives. *New Ideas in Psychology, 23,* 140-151. doi: 10.1016/j.newideapsych.2006.07.002

Granott, N., Fischer, K. W., & Parziale, J. (2002). Bridging to the unknown: A transition mechanism in learning and development. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 131-156). NY: Cambridge University Press.

Granott, N., & Parziale, J. (2002). Microdevelopment: A process-oriented perspective for studying development and learning. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 1-28). NY: Cambridge University Press.

Greeno, J. G. (1989). A perspective on thinking. *American Psychologist, 44*, 134-141. doi: 10.1037/0003- 066X.44.2.134

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, *124*, 75-111. doi: 10.1037/0033-2909.124.1.75

Grotzer, T. A. (2004). Putting everyday science within reach. *Principal Leadership, 5*, 16-21.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*, 17-27. doi: 10.1111/j.1745-3992.2004.tb00149.x

Hale, T. S., Bookheimer, S., McGough, J. J., Phillips, J. M., & McCracken, J. T. (2007). Atypical brain activation during simple & complex levels of processing in adult ADHD. *Journal of Attention Disorders, 11*, 125-140.

Halley, J. D. & Winkler, D. A. (2008). Classification of emergence and its relation to self-organization. *Complexity, 13,* 10-15. doi: 10.1002/cplx.20216

Haselager, P., de Groot, A., & Van Rappard, H. (2003). Representationalism vs. anti-representationalism: a debate for the sake of appearance. *Philosophical Psychology, 16*, 5-24. doi: 10.1080/0951508032000067761

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability and numbers of high-scoring individuals. *Science*, *269*. doi: 10.1126/science.7604277

Helsen, W. F., Starkes, J. L., & Van Winckel, J. (2000). Effect of a change in selection year on success in male soccer players. *American Journal of Human Biology, 12*, 729–735. doi: 10.1002/1520- 6300(200011/12)12:6<729::AID-AJHB2>3.0.CO;2-7

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students*. The Journal of Educational Research, 93*, 113-125. doi: 10.1080/00220679909597635

Hirsch-Pasek, K., Golinkoff, R. M., Berk, L. E., & Singer, D. G. (2009). *A mandate for playful learning in preschool: Presenting the evidence*. New York: Oxford University Press.

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review, 70,* 33-40.

Hollenstein, T. (2007). State space grids: Analyzing dynamics across development. *International Journal of Behavioral Development, 31*, 384-396. Doi: 10.1177/0165025407077765

Humphries, T. (2007). Attention-deficit/hyperactivity disorder. In I. Brown, & M. Percy (Eds.), *A comprehensive guide to intellectual and developmental disabilities* (pp. 295-307). Baltimore, MD, US: Paul H Brookes Publishing.

Inhelder, B., & Piaget, J. (2001). *The growth of logical thinking from childhood to adolescence*. (A. Parsons & S. Milgram, Trans.) (Rev. ed.) New York: Basic Books. (Original work published in 1958).

Jacoby, W. G. (2000). Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies, 19*, 577–613. doi: 10.1016/S0261-3794(99)00028-1

Jansen, B. R. J., & Van Der Maas, H. L. J. (2001). Evidence for the phase transition from Rule I to Rule II on the balance scale task. *Developmental Review, 21*, 450-494. doi: 10.1006/drev.2001.0530

Kalff, A. C., de Sonneville, L. M. J., Hurks, P. P. M., Hendriksen, J. G. M., Kroes, M., Feron, F. J. M., Steyaert, J., Van Zeben, T. M. C. B., Vles, J. S. H., & Jolles, J. (2003). Low- and high-level controlled processing in executive motor control tasks in 5-6-year-old children at risk of ADHD. *Journal of Child Psychology and Psychiatry, 44*, 1049-1057.

Kanter, D. E., Smith, H. D., McKenna, A., Rieger, C., & Linsenmeier, R. A. (2003). *Inquiry-based laboratory instruction throws out the 'cookbook' and improves learning.* Paper presented at the annual meeting of the American Society for Engineering.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, 119-127.

Kello, C. T., Brown, G. D. A., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., Van Orden, G. C. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences 14*, 223-232. doi: 10.1016/j.tics.2010.02.005

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31,* 338-368. doi: 10.3102/00028312031002338

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86. doi: 10.1207/s15326985ep4102_1

Kitchener, K. S., Lynch, C. L., Fischer, K. W., & Wood, P. K. (1993). Developmental range of reflective judgment: The effect of contextual support and practice on developmental stage. *Developmental Psychology, 29*, 893-906. doi: 10.1037/0012-1649.29.5.893

Koerhuis, J. (2011). Van welk kind is deze schaduw? Nieuwe toets rekenen voor kleuters van Cito. *Volgens Bartjens, 30*, 4-7.

Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education, 51*, 315-327. doi: 10.1177/0022487100051004007

Koretz, D. M. (2009). *Measuring up: What educational testing really tells us.* Cambridge, USA: Harvard University Press.

Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Development, 60*, 1316-1327. doi: 1316–1327.10.2307/1130923

Kuenzi, J. J. (2008). *Science, Technology, Engineering, and Mathematics (STEM) Education: Background, Federal Policy, and Legislative Action*. Congressional Research Service Reports. Retrieved from: http://digitalcommons.unl.edu/docs/35

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science, 6,* 133-139.

Kuhn, D. (2002). A multi-component system that constructs knowledge: Insights from microgenetic study. *Microdevelopment: Transition processes in development and learning* (pp. 109-130). NY: Cambridge University Press.

Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how). In: W. Damon, R. M. Lerner, (Series Eds.), D. Kuhn & R. S. Siegler (Vol. Eds). *Handbook of child psychology: Vol. 2. Cognition, perception and language* (6th ed, pp. 953-993). Hoboken, NJ: John Wiley & Sons.

Küpers, E., Van Dijk, M., McPherson, G., & Van Geert, P. (2013). A dynamic model that links skill acquisition with self-determination in instrumental music lessons. *Musicae Scientiae.* doi: 10.1177/1029864913499181

Ladd, G., Birch, S., & Buhs, E. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, *70*, 1373-1400.

Lane, K. L., Barton-Arwood, S. M., Nelson, J. R., & Wehby, J. (2008). Academic performance of students with emotional and behavioral disorders served in a self-contained setting. *Journal of Behavioral Education, 17*, 43-62. doi: 10.1007/s10864-007-9050-1

Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). *STEM: Good Jobs Now and for the Future.* ESA (Economics and Statistics Administration) Issue 03-11. US Department of Commerce.

Langen, A. V., & Dekkers, H. (2005). Cross-national differences in participating in tertiary science, technology, engineering and mathematics education. *Comparative Education, 41*, 329-350. doi: 10.1080/03050060500211708

Lehrer, R., & Schauble, L. (1998). Reasoning about structure and function: children’s conceptions of gears. *Journal of Research in Science Teaching, 35*, 3-25. doi: 10.1002/(SICI)1098-2736(199801)35

Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development, 23*, 512-529. doi: 10.1016/j.cogdev.2008.09.001

Leseman, P. (2004). De toegevoegde waarde van vroeg testen. *Pedagogiek, 24*, 3-11.

Lewis, M. D. (2000). Emotional self-organization at three time scales. In M.D. Lewis and I. Granic (Eds.), *Emotion, Development, and Self-Organization* (pp. 37–69). Cambridge, MA: Cambridge University Press

Lewis, M. D., & Granic, I. (2000). *Emotion, Development, and Self-Organization*. Cambridge, MA: Cambrige University Press.

Lichtwarck-Aschoff, A., Van Geert, P., Bosma, H., Kunnen, S. (2008). Time and identity: A framework for research and theory formation. *Developmental Review, 28,* 370-400. doi: 10.1016/j.dr.2008.04.001

Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment.* New York: The Guilford Press.

Linn, M.C., & Eylon, B.S. (2006). Science education: Integrating views of learning and instruction. In P.A. Alexander, P.H.Winne (Eds.), *Handbook of Educational Psychology* (2nd ed, pp. 511–544). Mahwah, NJ: Lawrence Erlbaum Associates.

Luo, X., & Li, X. (2003). Memory and learning ability of children with ADHD. *Chinese Mental Health Journal, 17,* 188-190.

Mace, R. L., Hardie, G. J., & Place, J. P. (1991). Accessible environments: Toward universal design. In W. E. Preiser, J. C. Vischer, & E. T. White (Eds.). Design intervention: Toward a more humane architecture. New York: Van Nostrand Reinhold.

Marsh, K. L., Johnston, L., Richardson, M. J., & Schmidt, R. C. (2009). Toward a radically embodied, embedded social psychology. *European Journal of Social Psychology, 39*, 1217-1225. doi: 10.3758/PBR.17.3.342

Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

McConaughy, S. H., Volpe, R. J., Antshel, K. M., Gordon, M., & Eiraldi, R. B. (2011). Academic and social impairments of elementary school children with attention deficit hyperactivity disorder. *School Psychology Review, 40*, 200-225.

McKinsey & Company (2007). *How the world's best-performing school systems come out on top.* Retrieved on June 2011 from: http://www.mckinsey.com/clientservice/Social_Sector/our_practices/Education/Knowledge_Highlights/How%20School%20Systems%20Get%20Better.aspx

Meindertsma, H. B., Van Dijk, M. W. G., Steenbeek, H. W., & Van Geert, P. L. C. (2012). Application of Skill Theory to compare scientific reasoning of young children in different tasks. *Netherlands Journal of Psychology, 67*, 9-19.

Menninga, A., Van Dijk, M. W. G., Steenbeek, H. W. & Van Geert, P. L. C. (2013). *Language in science education.* Poster presented at the EARLI conference, München, Germany.

Merrell, K. W., & Boelter, E. (2001). An investigation of relationships between social behavior and ADHD in children and youth: Construct validity of the Home and Community Behavior Scales. *Journal of Emotional and Behavioral Disorders, 9*, 260-269.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*, 5-11. doi: 10.3102/0013189X018002005

Moesch, K., Elbe, A. M., Hauge, M. L., & Wikman, J. M. (2011). Late specialization: the key to success in centimeters, grams, or seconds (cgs) sports. *Scandinavian Journal of Medicine & Science in Sports, 21*, 282-290. doi: 10.1111/j.1600-0838.2010.01280.

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*, 112-117. doi: 10.1111/j.1467-8721.2009.01619.x

Mooney, P., Epstein, M. H., Reid, R., & Nelson, J. R. (2003). Status of trends in academic intervention research for students with emotional disturbance. *Remedial and Special Education*, *24*, 273-287. doi: 10.1177/07419325030240050301

National Research Council (2000). *Inquiry in the national science education standards: A guide for teaching and learning.* Washington, DC: National Academy Press.

National Research Council (2005). *Mathematical and Scientific Development in Early Childhood.* Washington D.C.: National Academic Press.

National Research Council. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Committee on Science Learning, Kindergarten Through Eighth Grade. R. A. Duschl, H. A. Schweingruber, and A. W. Shouse (Eds.), Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2010). *Exploring the Intersection of Science Education and 21st Century Skills: A Workshop Summary.* Margaret Hilton, Rapporteur. Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Noens, I. L., & Van Berckelaer-Onnes, I. A. (2005). Captured by details: Sense-making, language and communication in autism. *Journal of Communication Disorders, 38*, 123–141.

Novak, J. D. (2005). Results and implications of a 12-year longitudinal study of science concept learning. *Research in Science Education, 35*, 23–40. doi: 10.1007/s11165-004-3431-4

OECD Program for International Student Assessment (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science, and Problem Solving Knowledge and Skills*. Retrieved from OECD website: http://www.oecd.org/pisa/pisaproducts/pisa2003/pisa2003assessmentframeworkmathematicsreadingscienceandproblemsolvingknowledgeandskills-publications2003.htm.

OECD Program for International Student Assessment (2004). *PISA Learning for Tomorrow's World: First Results from PISA 2003* (Vol. 659). OECD Publishing.

OECD (2005). *Students with Disabilities, Learning Difficulties and Disadvantages: Statistics and Indicators*, OECD Publishing. doi: 10.1787/9789264009813-en.

OECD Program for International Student Assessment (2006). *PISA: Where Immigrant Students Succeed: A Comparative Review of Performance and Engagement in PISA 2003*. OECD Publishing.

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

Palmer, D. (2001). Students’ alternative conceptions and scientifically acceptable conceptions about gravity. *International Journal of Science Education, 23*, 691-706. doi: 10.1080/09500690010006527

Papadouris, N., Constantinou, C. P., & Kyratsi, T. (2008). Students' use of the energy model to account for changes in physical systems. *Journal of Research in Science Teaching, 45,* 444-469. doi: 10.1002/tea.20235

Parziale, J. (2002). Observing the dynamics of construction: children building bridges and new ideas. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 157-180). NY: Cambridge University Press.

Parziale, J. & Fischer, K.W. (1998). The practical use of skill theory in classrooms. In R.J. Sternberg and W.M. Williams (Eds.), *Intelligence, instruction and assessment* (pp. 96-100). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perkins, D., & Blythe, T. (1994). Putting understanding up front. *Educational Leadership, 51,* 4-7.

Philips, S., & Tolmie, A. (2007). Children's performance on and understanding of the balance scale problem: The effects of parental support. *Infant and Child Development, 16*, 95-117. doi: 10.1002/icd.504

Piaget, J. (2001). *The Psychology of Intelligence* (M. Piercy & D. E. Berlyne, Trans.) (Rev. ed.) London: Routhledge Classics (Original work published in 1947).

Pisha, B., & Coyne, P. (2001). Smart From the Start The Promise of Universal Design for Learning. *Remedial and special education, 22*, 197-203. doi: 10.1177/074193250102200402

Pollatou, E., Karadimou, K., & Gerodimos, V. (2005). Gender differences in musical aptitude, rhythmic ability and motor performance in preschool children. *Early Child Development and Care, 175*, 361-369. doi: 10.1080/0300443042000270786

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211-227. doi: 10.1002/sce.3730660207

Pressley, M., Hogan, K., Wharton-McDonald, R., & Mistretta, J. (1996). The challenges of instructional scaffolding: The challenges of instruction that supports student thinking. *Learning Disabilities Research & Practice, 11*, 138-146.

Rakotomalala, R., 2005. TANAGRA: a free software for research and academic purposes. In *Proceedings of EGC'2005, RNTI-E-3, vol. 2.* (pp 697-702).

Rappolt-Schlichtmann, G., Daley, S. G., Lim, S., Lapinski, S., Robinson, K. H., & Johnson, M. (2013). Universal Design for Learning and elementary school science: Exploring the efficacy, use, and perceptions of a web-based science notebook. *Journal of Educational Psychology, 105*, 1210-1225. doi: 10.1037/a0033217

Rappolt-Schlichtmann, G., Tenenbaum, H. R., Koepke, M. F., & Fischer, K. W. (2007). Transient and robust knowledge: Contextual support and the dynamics of children's reasoning about density. *Mind, Brain, and Education, 1,* 98-108. doi: 10.1111/j.1751-228X.2007.00010.x

Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education, 38*, 130-143. doi: 10.1177/00224669040380030101

Renninger, K., & Granott, N. (2005). The process of scaffolding in learning and development. *New Ideas in Psychology*, *23*, 111-114. doi: 10.1016/j.newideapsych.2006.07.003

Resnick, L. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2010). Challenging the egocentric view of coordinated perceiving, acting, and knowing. In B. Mesquita, L. Feldman Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 307-333). New York, NY: Guilford Press.

Roberts, G. (2002). *SET for success. The supply of people with science, technology, engineering and mathematical skills*. The report of Sir Gareth Roberts' Review.

Retrieved from: http://www.vitae.ac.uk/policy-practice/1685/Roberts-recommendations.html

Rogoff, B. (1993). Children's guided participation and participatory appropriation in sociocultural activity. In R. Woxniak & K. Fischer (Eds.), *Development in context: Acting and thinking in specific environments* (pp. 121-153). Hillsdale, NJ: Erlbaum.

Rohaan, E., & Van Keulen, H. (2011). What Everyone Should Know About Science and Technology. In M. J. De Vries, H. Van Keulen, S. Peters, and J. Walma Van Der Molen (Eds.), *Professional Development for Primary Teachers in Science and Technology: The Dutch VTB-Pro Project in an International Perspective* (pp. 35-47). Rotterdam, NL: SensePublishers. doi: 10.1007/978-94-6091-713-4_3

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68-78. doi: 10.1037110003-066X.55.1.68

Schwartz, M. S., & Fischer, K. W. (2004). Building general knowledge and skill: Cognition and microdevelopment in science learning. In A. Demetriou & A. Raftopoulos (Eds.), *Emergence and transformation in the mind: Modeling and measuring cognitive change* (pp. 157 – 185). Cambridge, UK: Cambridge University Press.

Scruggs, T. E., & Mastropieri, M. A. (1986). Academic characteristics of behaviourally disordered and learning disabled students. *Behavioral Disorders, 11*, 184–190.

Séré, M. G. (1986). Children's conceptions of the gaseous state, prior to teaching. *European Journal of Science Education*, *8*, 413-425. doi: 10.1080/0140528860080408

Sharp, J. G. & Sharp, J. C. (2007): Beyond shape and gravity: children's ideas about the earth in space reconsidered. *Research Papers in Education, 22*, 363-401. doi: 10.1080/02671520701497613

She, H. C. (2002). Concepts of a higher hierarchical level require more dual situated learning events for conceptual change: A study of air pressure and buoyancy. *International Journal of Science Education, 24*, 981-996. doi: 10.1080/09500690110098895

Shipstone, D. M. (1984). A study of children's understanding of electricity in simple DC circuits. *European Journal of Science Education, 6*, 185-198. doi: 10.1080/0140528840060208

Shrager, J. & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*, 405-410. doi: 10.1111/1467-9280.00076

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, *3*, 1-5.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*, 225-273. doi: 10.1006/cogp.1995.1006

Siegler, R. S. (1996). *Emerging minds. The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Volume 2: Cognition, perception, and language* (6th ed, pp. 464–510). Hoboken, NJ: Wiley.

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10*, 104–109. doi: 10.1111/j.1467-7687.2007.00571.x

Siegler, R. S., & Alibali, M. W. (2005). *Children's thinking*. NY/ US: Pearson/Prentice Hall.

Siegler, R. S., & Crowly, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46,* 606-620.

Simonton, D. K. (1999). Talent and its development: An emergenic and epigenetic model. *Psychological Review, 106*, 435-457. doi: 10.1037/0033-295X.106.3.435

Simonton, D. K. (2001). Talent development as a multidimensional, multiplicative, and dynamic process. *Current Directions in Psychological Science, 10*, 39-43. doi: 10.1111/1467-8721.00110

Simonton, D. K. (2005). Giftedness and genetics: The emergenic-epigenetic model and its implications. *Journal for the Education of the Gifted, 28*, 270-286. doi: 10.4219/jeg-2005-338

Simonton, D. K. (2011). Creativity and discovery as blind variation: Campbell's (1960) BVSR model after the half-century mark. *Review of General Psychology, 15*, 158-174. doi: 10.1037/a0022912

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in cognitive sciences, 7*, 343-348. doi:10.1016/S1364-6613(03)00156-6

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science: a critical review. *American Psychologist*, *60*, 950-958. doi: 10.1037/0003-066X.60.9.950

Steenbeek, H. W. (2006). *Modeling dyadic child-peer interactions: Sociometric status, emotional expressions, and instrumental actions during play*. Unpublished doctoral dissertation, University of Groningen, The Netherlands.

Steenbeek, H., Jansen, L., & Van Geert, P. (2012). Scaffolding dynamics and the emergence of problematic learning trajectories. *Learning and Individual Differences*, *22*, 64-75. doi: 10.1016/j.lindif.2011.11.014

Steenbeek, H. W. & Uittenbogaard, W. (2009). TalentenKracht brengt talent voor wetenschap en techniek van jonge kinderen in kaart. In H. Van Keulen and J. Walma Van der Molen, *Onderzoek naar wetenschap en techniek in het Nederlandse basisonderwijs* (pp. 131-144). Den Haag: programma VTB-Pro Platform Bèta Techniek.

Steenbeek, H. W., & Van Geert, P. (2005). A dynamic systems model of dyadic interaction during play of two children. *European Journal of Developmental Psychology, 2,* 105-145. doi: 10.1080/17405620544000020

Steenbeek, H. W., & Van Geert, P. (2007). 'Do you still like to play with him?' *Netherlands Journal of Psychology, 63,* 79-92. doi: 10.1007/BF03061069

Steenbeek, H., & Van Geert, P. (2013). The emergence of learning-teaching trajectories in education: A complex dynamic systems approach. *Nonlinear dynamics, psychology, and life sciences, 17*, 233-267.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic Tesing: the nature and measurement of learning potential.* Cambridge, UK: Cambridge University Press.

Stone, C. A. (1998). The metaphor of scaffolding, its utility for the field of learning disabilities. *Journal of Learning Disabilities, 31*, 344-364. doi: 10.1177/002221949803100404

Taber, K. S. (2001). Building the structural concepts of chemistry: some considerations from educational research. *Chemistry Education: Research and Practice in Europe, 2*, 123-158. doi: 10.1039/B1RP90014E

Terwel, J., & Van den Eeden, P. (1992). Differentiële effecten van het werken in kleine groepen: theorie, hypothesen en onderzoek. *Pedagogische Studiën, 69*, 51-66.

Thelen, E. (1989). Self-organization in developmental processes: Can systems approaches work? In M. Gunnar and E. Thelen (Eds.), *Systems and development: Vol. 22. The Minnesota symposium in child psychology* (pp. 77-117). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thelen, E. (1992). Development as a dynamic system. *Current Directions in Psychological Science, 1*, 189-193.

Thelen, E., & Corbetta, D. (2002). Microdevelopment and dynamic systems: Applications to infant motor development. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 59-79). NY: Cambridge University Press.

Thelen, E. & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: The MIT Press.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from: http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.htm

Todman, J. B. & Dugard, P. (2001). *Single-Case and Small-N Experimental Designs: A Practical Guide To Randomization Tests*, Mahwah, NJ, USA: Lawrence Erlbaum.

Trainor, L. J., & Unrau, A. (2012). Development of pitch and music perception. In L. Werner, R. R. Fay, and A. N. Popper. *Human Auditory Development* Vol. 42, (pp. 223-254). New York: Springer. doi: 10.1007/978-1-4614-1421-6_8

Trout, A. L., Nordness, P. D., Pierce, C. D., & Epstein, M. H. (2003). Research on the Academic Status of Children with Emotional and Behavioral Disorders A Review of the Literature From 1961 to 2000. *Journal of Emotional and Behavioral Disorders, 11*, 198-210. doi: 10.1177/10634266030110040201

Trumper, R. (1993). Children's energy concepts: a cross-age study. *International Journal of Science Education, 15*, 139-148. doi: 10.1080/0950069930150203

Tytler, R. (1998). The nature of students’ informal science conceptions. *International Journal of Science Education, 20*, 901-927. doi: 10.1080/0950069980200802

Understanding (1989). *Oxford English dictionary.* Oxford University Press. Retrieved June 24, 2011, from http://www.oed.com/view/Entry/212090.

U.S. Public Health Service (1999). *Report of the surgeon general’s conference on children’s mental health: A national action agenda*. Washington, DC: Department of Health and Human Services. Retrieved from: http://www.surgeongeneral.gov/topics/cmh/childreport.html

Van Bijsterveldt-Vliegenthart, M. (2011). *Kamerbrief over invoeringsmomenten verplicht lvs en centrale eindtoets in sbo en so.* De minister van Onderwijs, Cultuur en Wetenschap. Tweede Kamer der Staten Generaal. Retrieved from: http://www.rijksoverheid.nl/ministeries/ocw/documenten-en publicaties/kamerstukken/2011/11/25/kamerbrief-over-invoeringsmomenten-verplicht-leerlingvolgsysteem-en-centrale-eindtoets-in-sbo-en-so.html?ns_campaign=kamerstukken-ministerie-van-onderwijs%2C-cultuur-en-wetenschap&ns_mchannel=att.

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

Van Der Maas, H., Klinkenberg, S., & Straatemeier (2010). Rekentuin.nl: combinatie van oefenen en toetsen. [Mathgarden.com: a combination of practice and assessment]. *Examens*, 2010.

Van Der Veer, R., & Valsiner, J. (1991). *Understanding Vygotsky: A quest for synthesis*. Blackwell Publishing.

Van Der Worp-Van Der Kamp, L., Pijl, S. J., Bijstra, J. O., & Van Den Bosch, E. J. (2013). Teaching academic skills as an answer to behavioural problems of students with emotional or behavioural disorders: A review. *European Journal of Special Needs Education*.

Van Dijk, M. & Van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development, 16*, 7-33. doi: 10.1002/icd.506

Van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. Harvester Wheatsheaf.

Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review, 105*, 634-677. doi: 10.1037/0033-295X.105.4.634-677

Van Geert, P. (2002). Developmental dynamics, intentional action and fuzzy sets. In N. Granott and J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 319-343). NY: Cambridge University Press.

Van Geert, P. (2004). Dynamic Modelling of cognitive development: Time, situatedness and variability. In A. Demetriou, & A. Raftopoulos (Eds.), *Cognitive developmental change: Theories, models and measurement* (pp. 354-379). Cambridge, MA: Cambridge University Press.

Van Geert, P. (2008). Nonlinear complex dynamic systems in developmental psychology. In: S. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and Complexity in Psychology - The Theory of Nonlinear Dynamical Systems* (pp. 242-281). Cambridge: Cambridge University Press.

Van Geert, P. (2011a). *Talent and Excellence for Science and Technology in Young Children: A Review of the Literature.* Internal report. Groningen: Rijksuniversiteit Groningen.

Van Geert, P. (2011b). The contribution of complex dynamic systems to development. *Child Development Perspectives, 5*, 273-278. doi: 10.1111/j.1750-8606.2011.00197.x

Van Geert, P. & Van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data*. Infant Behavior & Development, 25,* 340-374. doi: 10.1016/S0163-6383(02)00140-6

Van Geert, P. & Van Dijk, M. (2003). Ambiguity in child language: The problem of inter-observer reliability in ambiguous observation data. *First Language, 23*, 259-284. doi: 10.1177/01427237030233001

Van Geert, P. L. C. & Fischer, K. W. (2009). Dynamic systems and the quest for individual based models of change and development. In J. P. Spencer, M. S. C. Thomas, & J. McClelland (Eds.), *Toward a New Grand Theory of Development? Connectionism and Dynamic Systems Theory Reconsidered* (pp. 313-335). Oxford: Oxford University Press.

Van Geert, P., & Steenbeek, H. (2005a). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review, 25*, 408-442. doi: 10.1016/j.dr.2005.10.003

Van Geert, P. & Steenbeek, H. W. (2005b). The dynamics of scaffolding. *New Ideas in Psychology, 23*, 115-128. doi: 10.1016/j.newideapsych.2006.05.003

Van Geert, P., & Steenbeek, H. W. (2007). *Waar ligt de kracht? Bijlage bij onderzoeksvoorstel satelliet Groningen TalentenKracht.* Internal report. Groningen: Rijksuniversiteit Groningen.

Van Geert, P., & Steenbeek, H. W. (2008). Brains and the dynamics of 'wants' and 'cans': A commentary on Immordino-Yang's tale of two cases. *Mind, Brain, and Education*, *2,* 62-66.

Van Geert, P. L. C., & Steenbeek, H. W. (2013). (P)(re) presentations: What are they and how do they develop? Commentary on “Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate” by J. Allen & M. Bickhard. *Cognitive Development, 28*, 138-143. doi: 10.1016/j.cogdev.2013.01.004

Van Geert, P., Steenbeek, H.W., & Kunnen, S. (2011). Monte Carlo Techniques: Statistical simulation for developmental data. In E. S. Kunnen (Ed.), *A dynamic systems approach to adolescent development.* (pp. 43-52) London-New York: Psychology Press.

Van Gelder, T. J. (1995) What might cognition be, if not computation? *Journal of Philosophy*, *91*, 345-381.

Van Gelder, T. J. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences, 21,* 615-628.

Van Langen, A. V., & Dekkers, H. (2005). Cross-national differences in participating in tertiary science, technology, engineering and mathematics education. *Comparative Education*, *41*, 329-350. doi: 10.1080/03050060500211708

Van Orden, G. C., Holden, J. G., Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132*, 331-350. doi:10.1037/0096-3445.132.3.331

Van Schijndel, T. J., Singer, E., Van Der Maas, H. L., & Raijmakers, M. E. (2010). A sciencing programme and young children's exploratory play in the sandpit. *European Journal of Developmental Psychology, 7*, 603-617. doi: 10.1080/17405620903412344

Van Vondel, S., & Steenbeek, H. W. (2014). Talentenkracht in de klas: wetenschappelijke redeneren in de bovenbouw. Een coachingsmodule voor leerkrachten in de bovenbouw van het reguliere basisonderwijs. Groningen: Rijksuniversiteit Groningen.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and instruction, 4,* 45-69. doi: 10.1016/0959-4752(94)90018-3

Vosniadou, S. (2007). The cognitive-situative divide and the problem of conceptual change. *Educational Psychologist, 42*, 55-66. doi: 10.1080/00461520709336918

Vosniadou, S., & Brewer, W. F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research, 57,* 51-67. doi: 10.3102/00346543057001051

Vygotsky, L. S. (1986). Thought and Language (A. Kozulin, Trans.). Cambridge, MA: MIT Press. (Original work published in 1934).

Wetzels, A. F. M., Steenbeek, H. W., & Van Geert, P. (2013). The effect of a coaching program on teachers’ practices in primary science teaching. *Article in preparation.*

Wijnstra, J., Ouwens, M., & Béguin, A. (2003). *De toegevoegde waarde van de basisschool.* Arnhem: Cito group (internal report).

Wilkening, F., & Sodian, B. (2005). Scientific reasoning in young children: Introduction. *Swiss Journal of Psychology, 64,* 137–139. doi: 10.1024/1421-0185.64.3.137

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89-100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Wood, D., Wood, H., & Middleton, D. (1978). An experimental evaluation of four face-to-face teaching strategies. *International Journal of Behavioral Development, 1*, 131-147.

Yan, Z., & Fischer, K. W. (2002). Always under construction. *Human Development, 45*, 141-160. doi: 10.1159/000057070

Yan, Z., & Fischer, K. W. (2007). Pattern emergence and pattern transition in microdevelopmental variation: Evidence of complex dynamics of developmental processes. *Journal of Developmental Processes, 2,* 39-62.

Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213, 21-87.

Zacharia, Z.C. (2007). Comparing and combining real and virtual experimentation: an effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning, 23*, 120–132. doi: 10.1111/j.1365-2729.2006.00215.x

Zednik, C. (2011). The Nature of Dynamical Explanation. *Philosophy of Science, 78*, 238-263. doi: 10.1086/659221

Zighed, D. A., & Rakotomalala, R. (1996). *Sipina-w (c) for windows: User's guide*. Technical report, ERIC laboratory, University of Lyon 2, France.

Zimmerman, B. J. (1990). Self-Regulated learning and academic achievement: an overview. *Educational Psychologist, 25*, 3-17. doi: 10.1207/s15326985ep2501_2

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*, 99-149. doi: 10.1006/drev.1999.0497

Zimmerman, C. (2005). The development of scientific reasoning skills: What psychologists contribute to an understanding of elementary science learning. *Final draft of a report to the National Research Council Committee on science learning kindergarten through eighth grade.*

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

Appendix A: Tasks

Table 1: incorporated aspects of air pressure sequence		**jumping frog**	**air squirt**	**the squirts**	**platform lift**	**air vs water**	**air canon**	**black box**	**balloon in syringe**	**straws & watering can**	**balloon & pop-pop boat**
main aspects	*air pressure*	x	x	x	x	X	x	x	x	x	x
	$P \times V = C$		x	x	x	X		x	x		
	piston system		x	x	x	X	x	x	x		
moving elements in task	*1*	x									
	2		x	x					x		x
	3 or more				x	X	x	x			
reactions in the system	*single*	x									
	double		x	x					x		
	threefold				x	X	x	x		x	x
effect input = output?	*symmetrical*	x	x		x			x			
	asymmetrical			x		X	x		x	x	x
visibility of effects	*visible*	x	x	x	x	X	x				x
	invisible							x	x	x	
driving mechanism	*movement*		x	x					x	x	
	driving object	x			x	X					x
	postponed						x	x			
miscellaneous	*liquid (pressure)*					X				x	x
	steam										x

Table 2: incorporated aspects in the gravity task sequence		**open marble track**	**stairs marble track**	**ball slide**	**nemo slide**	**looping**	**balance**	**marble track**	**slides & paradox**	**crater maker**	**electric cradle**
main aspects	*gravity*	x	x	x	x	x	x	x	x	x	x
	acceleration	x		x	x	x		x	x		
	inertia	x	x	x	x		x	x	x	x	x
friction	*no variation*		x			x			x		x
	variation in objects	x		x	x		x	x		x	
	variation in surface				x			x			
object weight	*stable*	x	x			x		x	x		x
	varies			x	x		x			x	
questions	*mostly representational*	x	x	x	x	x	x	x	x	x	x
	abstract questions				x		x	x	x	x	x
child’s manipulation	*fixed task*	x	x	x	x				x		x
	partly self-designed					x	x	x		x	
balance within task	*not present*	x	x	x	x	x		x		x	x
	balancing	x	x	x	x		x				
	center of gravity					x	x		x		
movement	*single effect*	x		x	x	x		x	x		
	movement transmission		x				x			x	x

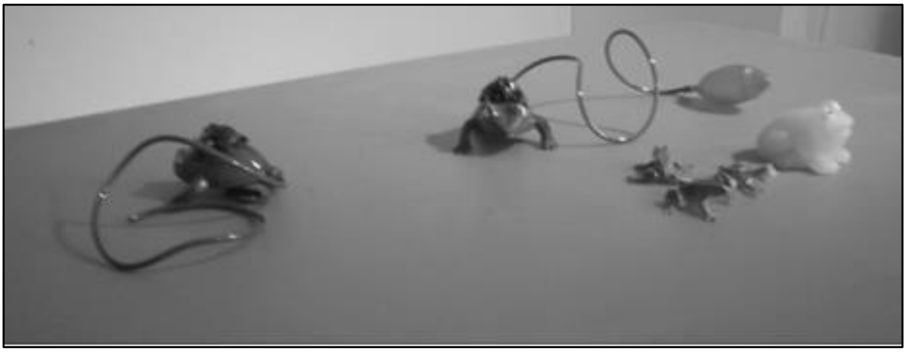
Air pressure sequence: Task 1

Name: Jumping frog (in Dutch: Het kikkertje)

Origin: Inspired on a task used by the *Curious Minds* team of the University of Utrecht.

Materials used: Set rubber or plastic toy frogs (toy store); two jumping toy frogs (toy store), one with balloon cut off.

Short description: The task starts by asking if children have ever seen frogs. Then the set of toy frogs is put on the table, asking children if they see differences and similarities between these. Subsequently, children are asked whether the toy frogs can jump, like real frogs. When the child realizes they cannot jump, the jumping frog is put on the table and children are asked if they can make this frog jump. Once children succeed, the researcher asks how this works. At the end of the task, a toy frog without a balloon is put on the table. The researcher asks why this frog, although highly similar to the other one, cannot jump.



Air pressure sequence: Task 2

Name: Air squirt (in Dutch: Luchtspuit)

Origin: Inspired on a task used by the *Curious Minds* team of the University of Utrecht (prototype task).

Materials used: Two syringes (used to be sold at the Dutch department store Hema in a science box, now available via e.g., websites selling medical products), a short and a long transparent plastic or silicone tube (sold at e.g., aquarium stores).

Short description: The tasks starts by asking children if they have ever seen a syringe and what these are used for. After comparing the two syringes ("are these the same?"), they are connected by a tube. Children are asked to predict what will happen if one of the syringes is pushed in, and are encouraged to try this out. The researcher asks for an explanation, and then asks what happens if you pull one of the pistons (instead of pushing). At the end of the task, a longer tube is connected to the syringes and differences in the functioning of the task are explored.

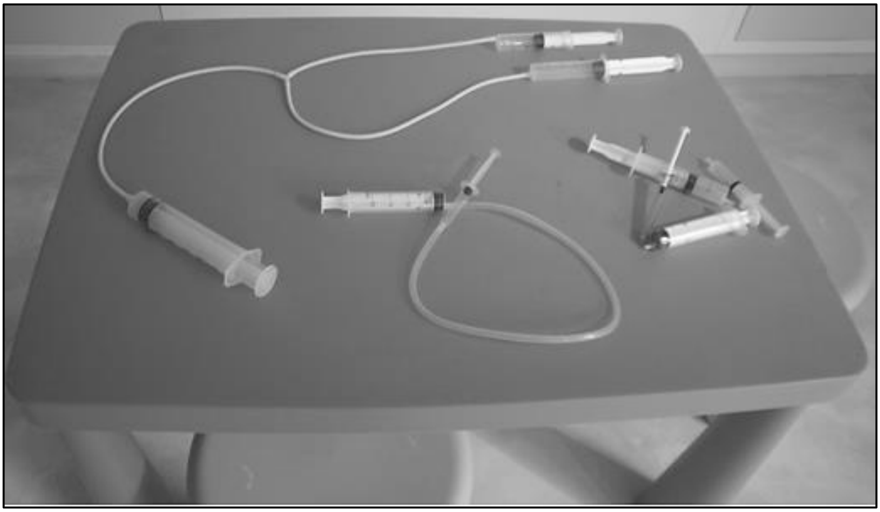
Air pressure sequence: Task 3

Name: The squirts (in Dutch: De spuitjes)

Origin: Own design, based on the administration of tasks so far.

Materials used: Syringes of different volume (available via e.g., websites selling medical products), transparent plastic or silicone tubes (sold at e.g., aquarium stores), triangle tube divider (sold at model building stores).

Short description: Children are asked if they remember the air squirt task from the second visit. If they don't, part of this task is repeated. Then the new task starts: combining different syringes: big ones with smaller ones, very thin ones, etc. Each time, children are asked what they think will happen (will the piston come out all the way or not?), and—after seeing the result—whether they can explain what just happened. Subsequently, two syringes of medium size are joined together by a tube with a cut in it (not immediately visible). Children are asked why the task does not work anymore. Near the end of the task, three syringes are connected (one large, two medium sized). Again, the child is asked for a prediction, and—after trying out his/her ideas with the material—for an explanation.



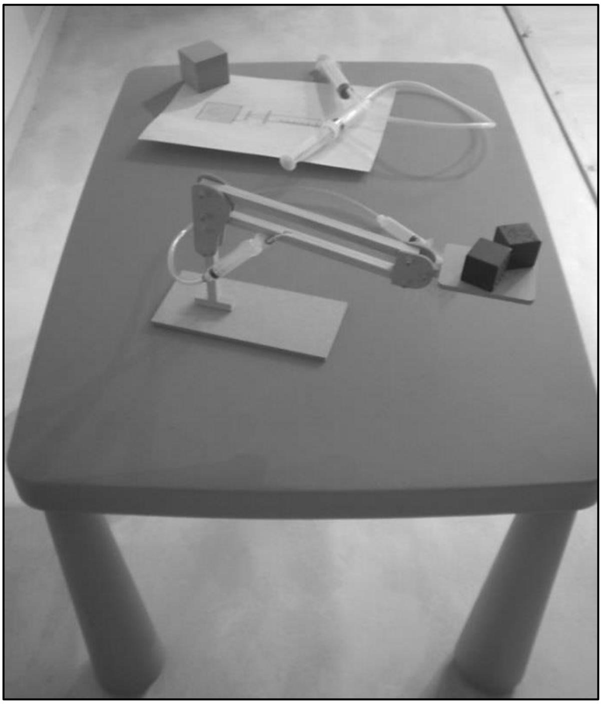
Air pressure sequence: Task 4

Name: The platform lift (in Dutch: De platformlift)

Origin: Inspired on a task used by the *Curious Minds* team of the University of Utrecht, and the administration of tasks so far.

Materials used: Construction box for platform lifts (available at the Dutch school supply store Heutink), two syringes and tube (see previous tasks), wooden and soft construction blocks (toy store), map (drawing).

Short description: Children are asked if they remember the air squirt task. Then the power of this system is explored by pushing in both pistons at the same time, and by trying to push away a soft construction block (the map is used as an aid if children cannot find out how the air squirt can push away blocks). Then the child is asked to use the air squirt to lift the same construction block. Subsequently, the platform lift is put on the table, without the syringes and tube attached. The child is asked to construct the lift in such a way that it can lift construction blocks. At the end of the task, the child is asked whether he or she thinks real lifts work in a similar way.



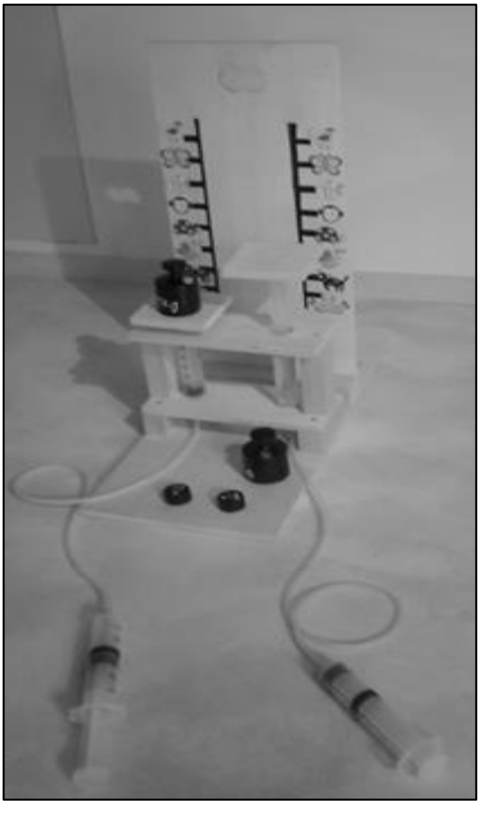
Air pressure sequence: Task 5

Name: Air versus water (in Dutch: Lucht versus water)

Origin: Own design, based on the administration of tasks so far.

Materials used: Wood (hardware store), syringes and tubes (see previous tasks), stickers (office supply store), weights (school supply store Heutink), water.

Short description: Children are asked what will happen if they push the syringe containing air halfway (until the blue line), and all the way until the end. Differences in the distance are explored. This is repeated for the other syringe, which contains water. After this, the weights are put on the platforms. First a small one (500 grams), then a heavy one (1 kg). When using the heavy weight, air compresses. Hence, the platform connected to the syringe with water comes further. Each time, children are asked to predict the distance on the scale, and to explain the result.



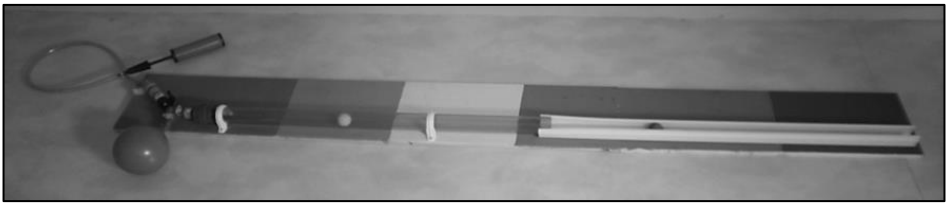
Air pressure sequence: Task 6

Name: Air canon (in Dutch: Luchtkanon)

Origin: Own design, based on the administration of tasks so far.

Materials used: Wood, garden sprinkler parts, transparent drainage tube, gutter made from a component used for room dividing (hardware store), ball pump, balloon, light and heavy table tennis balls (toy store).

Short description: There are tree (sprinkler) taps on this device, one to (dis)connect the air pump, one to (dis)connect the balloon, and one to (dis)connect the drainage tube. Children are first asked what they think the apparatus is for. Through questioning, they realize they have to open some taps to make the canon work. There are two ways to shoot a ball down the tube: 1) opening the taps connected to the pump and tube, and 2) by inflating the balloon first, and then releasing the air in the tube. The researcher asks for an explanation of the mechanism. When children figure out how to use the balloon, differences between 2 and 4 pumps of air in the balloon are explored. The colors on the wood serve as a measuring device to see how far the ball goes.



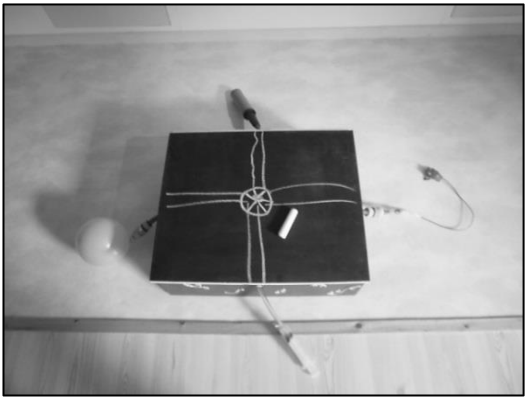
Air pressure sequence: Task 7

Name: The black box (Dutch name: The black box)

Origin: Own design, based on the administration of tasks so far.

Materials used: Wood, garden sprinkler parts, parts of a garden hose, chalkboard paint (hardware store), transparent tube (aquarium store), jumping frog, balloon, chalk, ball pump (toy store), syringe (medical store online).

Short description: In the box, out of the children's sight, are 4 taps: one leading to a jumping frog, one leading to a syringe, one leading to a balloon, and one leading to nothing. After exploring the outside of the box, the researcher changes something inside the box (opens one of the taps). The child is asked to predict what the researcher changed, then pumps the ball pump and observes the result. After all taps have been opened in this way one by one, the child is asked to draw what he/she thinks is inside the box, by using chalk. The box is then opened and differences between the drawing and the inside are discussed. Then it is the child's turn to manipulate the taps, trying to make the frog jump and the balloon inflate.



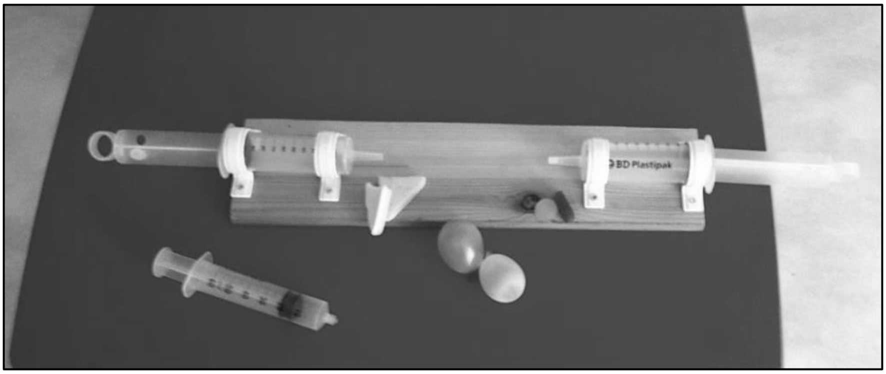
Air pressure sequence: Task 8

Name: Balloon in syringe (Dutch name: Ballon in spuit)

Origin: Own design, based on administration of tasks so far, inspiration from various websites (e.g., www.encyclopedoe.nl).

Materials used: Balloons, small (water) balloons (toy store), large syringes (medical store online), wood, construction handles (hardware store), wine gums, marshmallows.

Short description: Children are presented with two inflated balloons: a small and a big one. Questions are: “What causes balloons to be big or small?” “Can you make a balloon bigger or smaller without destroying it?” After children explore the material and conclude that this is not possible, small water balloons (containing air) are presented and put in a syringe. The question is what happens if the piston of the syringe is pushed in, while holding your finger on the syringe’s tip. This makes the balloon grow smaller when pushing the piston, and bigger when pulling the piston. The child is asked why this happens, and if this can be repeated using a marshmallow (yes), and using a winegum (no). Why does it only work with some objects?



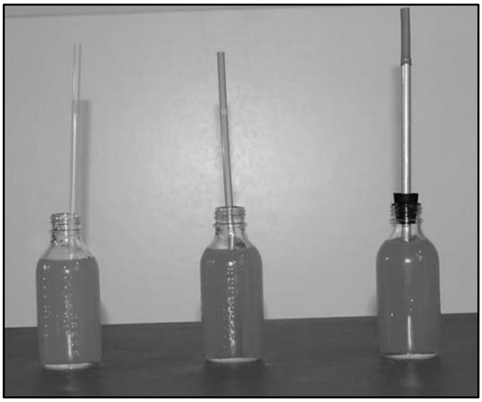
Air pressure sequence: Task 9

Name: Straws & special watering can (Dutch name: Limonadeflesjes en de bijzondere gieter)

Origin: Task with straws based on Tina Grotzer's course "Research and Evidence: Framing Scientific Research for Public Understanding" (Harvard Graduate School of Education), the special watering can experiment was found on the website www.proefjes.nl.

Materials used: Small bottles (medical store online), lemonade, drinking straws, lemonade bottle with cap (supermarket), cork, bowl (kitchen supply store), water.

Short description: Children are asked if they want to try some lemonade, and to drink from a straw. Subsequently, the child is asked to draw what happens when drinking from a straw. Then another straw (with a hole) is presented. What happens if you drink from it with the hole in the lemonade? And what happens if you turn the straw upside down, with the hole sticking out of the bottle? The child is asked to explain why no lemonade comes out of the straw in the latter case. Finally, the child is asked to drink from a straw going through a cork enclosing the top of the bottle. Why is this not working? Then the next part of the task is presented: a watering can made out of a lemonade bottle, with holes in the bottom. It only works with the cap off. The child is asked to explain this.



Air pressure sequence: Task 10

Name: Balloon and pop-pop boat (Dutch name: Ballonboot en stoomboot)

Origin: Own design, based on administration of tasks so far, inspiration from various websites (e.g., www.encyclopedoe.nl).

Materials used: Balloon boat, balloons, pop-pop boat (online toy stores), baby bath (baby shop), lighter, small candles (kitchen supply store), small syringe to inject water in the boat's tubes (medical store online), water.

Short description: The children are first presented with the balloon boat (without balloon attached). The question is how it can sail by itself. When children cannot think of a way, a balloon is presented and they are asked whether they could use this to make the boat sail. Children attach an inflated balloon to the boat's chimney and the boat sails. Questions such as "how does it work?" and "can you make it sail for a longer time?" are asked. Then the pop-pop boat is presented. Again, the question is how to make it sail, using a small candle. Children try to predict, and then the researcher makes the boat sail, by squirting some water in the tubes of the boat, and lighting the candle. Children are asked questions such as: How come it starts sailing, what is the driving mechanism? How come it takes a while? How do you make it stop?

Gravity sequence: Task 1

Name: Open marble track (Dutch name: Jodelbaan)

Origin: Inspired on a task used by the *Curious Minds* team of the University of Utrecht (prototype task).

Materials used: Wooden marble track from the brand NiC or Fagus and objects for it (toy store), short wooden sticks and big wooden beads to attach to the end of the track (craft's store), doormat (hardware store), markers next to the doormat to see how far marbles reach (e.g., toy blocks).

Short description: Children are asked if they ever saw something like this, and how they think it works. Then the marbles are presented and children are asked to make these roll down the track. The researcher asks what happens if two marbles are rolling behind one another, and which one reaches the end of the track first (at the end of each track, the marbles switch, which is why the marble that starts first, is the last at the end of the track). Other objects (a disc, a little doll) are also rolled down the track, and differences are observed. At the end of the task, the child is asked how far the marbles reach when they get the opportunity to roll out on a doormat. The marbles never come further than the middle of the mat. The child is asked why he/she thinks this is the case (because marbles lose their speed at the end of every level, it does not matter at which level the marble is released).

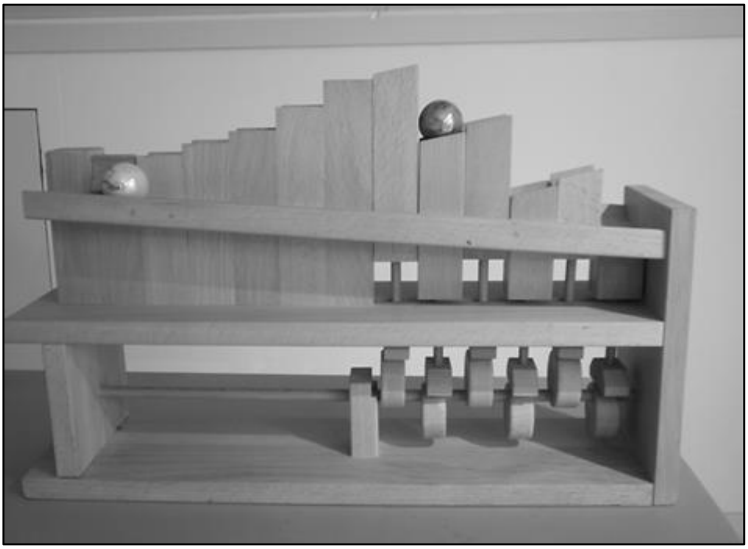
Gravity sequence: Task 2

Name: Stairs marble track (Dutch name: Trapkogelbaan)

Origin: Inspired on a task used by the *Curious Minds* team of the University of Utrecht.

Materials used: Wooden marble track with a stair-wise mechanism (made in Germany, sold online through various toy store websites), big marbles (toy store), big nails used as obstacles on the slope of the track (hardware store).

Short description: This task does not look like a marble track, so children are first asked what they think it is. Then the marble is presented, and the child is asked how the track works. After they find out, the child is asked how it is possible that the marble goes upward, and how it is possible that the marbles alternate tracks (left/right) when going down the slide (the final step of the stairs works as a switch, taking the marbles to either the left or the right slide). Then obstacles (nails) are put in one of the slides, at various points down the slope, and the child is asked why the marbles cannot overcome the obstacles at the beginning of the slide, but can at the end of the slide (due to their increased speed).



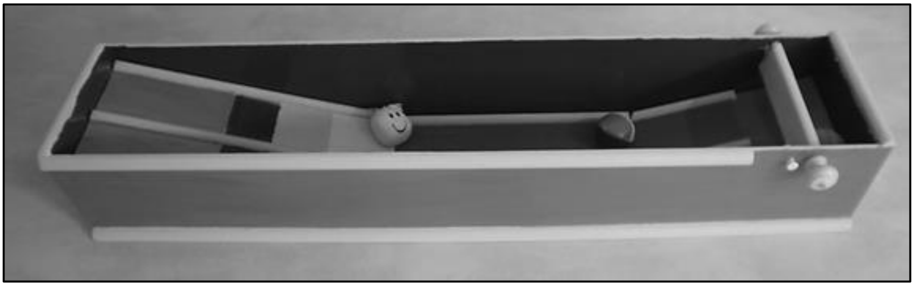
Gravity sequence: Task 3

Name: Ball slide (Dutch name: Ballenbak)

Origin: Own design, based on the administration of tasks so far.

Materials used: Wood, nails/screws, paint (hardware store), various balls differing in size, weight, and texture; scale (toy store).

Short description: Children are asked what they think the slide is for. When they mention they need a ball, various balls are presented. The first ball causes a lot of friction and does not come far. The child is asked for an explanation. The second ball is put on the scale together with the first one, to determine which one is heavier. The second ball (heavier) comes further on the slide. Then a soft light tennis ball is introduced. After using the scale, children conclude that this ball is lighter. Still, it comes further, due to the fact that it has less friction. Weight does not matter that much, friction does. Then a hard heavy ball is put on the slide. This one causes the least amount of friction and comes the furthest. At the end of the task, child and researcher participate in a race. The child is asked which ball he/she chooses to race and why.



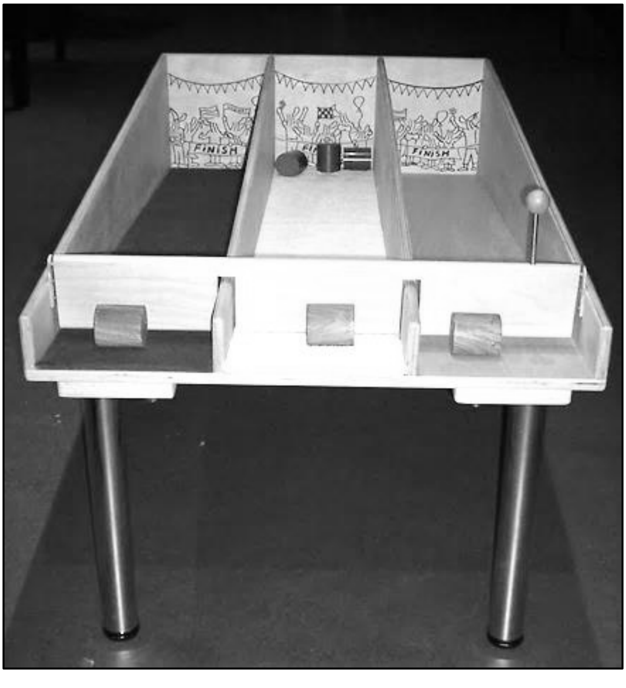
Gravity sequence: Task 4

Name: Nemo slide (Dutch name: Nemobak)

Origin: Task borrowed from science center Nemo, Amsterdam. Has earlier been used by the University of Amsterdam (UvA) and Nemo for a study.

Materials used: Nemo slide (wood, metal legs, different types of fabric, wooden and heavy plastic cylinders).

Short description: First 3 wooden cylinders are used to see which track is fastest (the one without fabric, as opposed to the ones with a smooth and coarse type of fabric). Subsequently, the heavier (grey) cylinders are used, which leads to the same effect. Then the lighter and heavier cylinders are paired on one of the slides. They reach the end of the track at the same time. This is counter-intuitive. Most children think that heavier weights go faster. Gravity, however, works on both cylinders in an equal manner. Children are explicitly asked if they have ever heard of gravity and how gravity might be at work in this task.



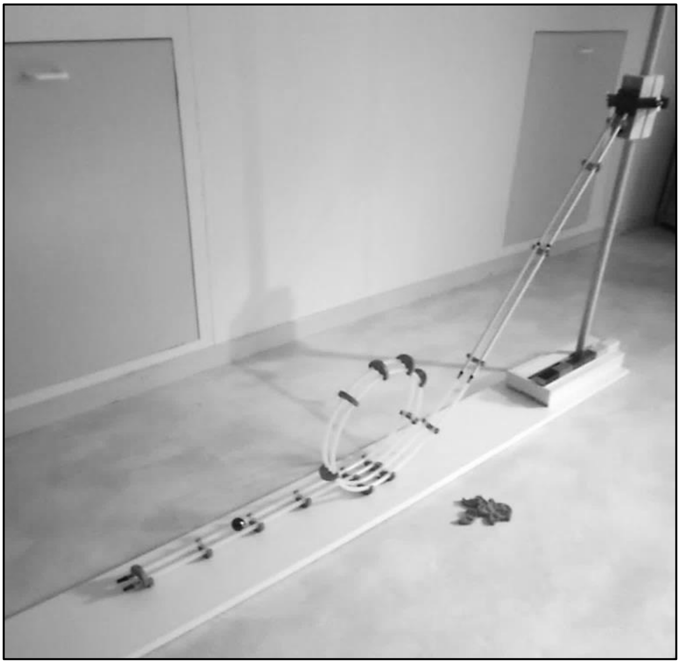
Gravity sequence: Task 5

Name: The looping (Dutch name: The Looping)

Origin: Own design, based on the administration of tasks so far.

Materials used: Flexible marble track from brand Mabro (online toy store), wood, broomstick, wooden rail for broomstick, adjustable hinge system on broom stick (hardware store).

Short description: The height of the track can be varied using the hinge system attached to the starting point of the track, and by varying the position of the broom stick. Depending on the height, the marble either rolls through the looping successfully, or falls down. At the beginning of the task, the marble falls down, and children are asked how this is possible. Through a series of adjustments (strengthening the track, putting the hinge system up, bringing the broomstick to the front, adjusting the hinge system once more) children can make the looping work. Throughout the task, children are asked to predict what will happen, and to explain their observations and actions.



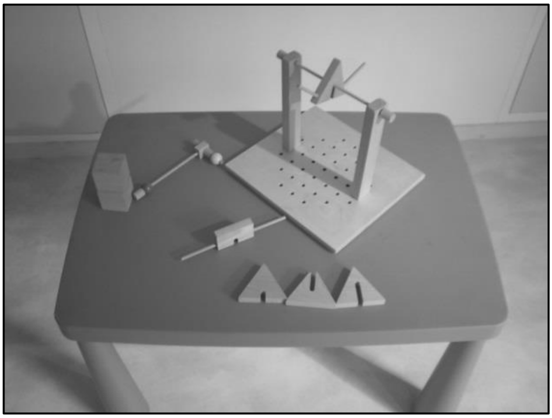
Gravity sequence: Task 6

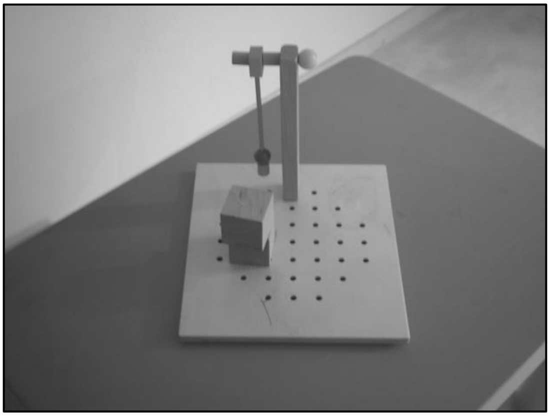
Name: The balance (Dutch name: Bijzondere balans)

Origin: School supply store Heutink.

Materials used: Balance scale/pendulum set (Dutch school supply store Heutink), wooden building blocks (toy store).

Short description: This task is made up of two parts. First we build a horizontal bar and let objects balance on this. Children are asked to investigate when there is a balance and when not. Children realize that it does not necessarily depend in the form of the objects (triangle, square), but on the distance between the point on which we try to balance, and the object's center of gravity. After this, the task is changed into a pendulum, and we explore the transfer of energy from the pendulum to wooden building blocks (stacked, next to one another, lighter and heavier blocks, etc.).





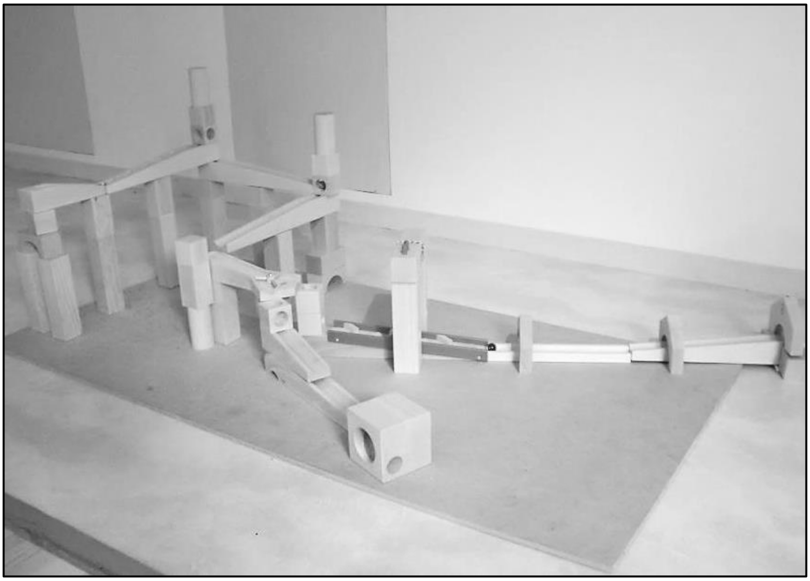
Gravity sequence: Task 7

Name: The marble track (Dutch name: De knikkerbaan)

Origin: Own design, based on the administration of tasks so far.

Materials used: Haba marble track and several extension sets (specialized toy store), wooden plank, glue (hardware store), marbles (toy store).

Short description: The marble track is pre-build before the task, but it does not work properly: the children have to solve the flaws one by one. First, the marble cannot not reach the first slope. Children are asked to make the first part of the track work properly. In the middle of the track there is an intersection, causing the marbles to take one of the tracks. This mechanism is explored by the children. Subsequently, children have to fix the track leading to a small looping, as the marbles cannot complete the full looping in the pre-build version of the track. Then we explore the other half of the track, trying to make the marbles reach the three goals (also impossible in the pre-build version). We finish by asking some questions about gravity and acceleration.



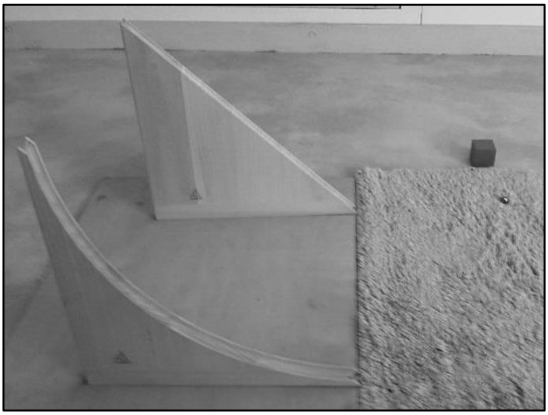
Gravity sequence: Task 8

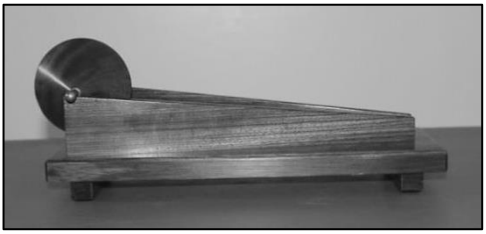
Name: Slides and paradox (Dutch name : Glijbanen en paradox)

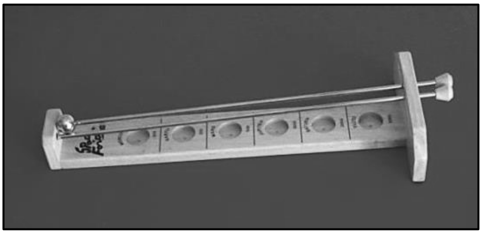
Origin: Own design, based on administration of tasks so far.

Materials used: Two slides with different slope, small marbles, paradox, adjustable paradox (Arabesk educational toy store—out of business), doormat (hardware store), markers next to the doormat to see how far marbles reach (e.g., toy blocks).

Short description: Two different slides are presented, and the child is asked on which slide the marble will come furthest. After trying this out, the child is asked to explain which slide works best (the curved slide). Subsequently, the child is asked to try out if one of the slides can be used to make the marble roll to the furthest marker. The second part of the task consists of the paradox. The question is what the object is for, and in which direction the object will roll. When this is counterintuitive (the object appears to roll upward), the object is explored using a ruler and an adjustable paradox to vary the distance between the two rails of the paradox. The child is asked to explain why the object appears to roll upward.







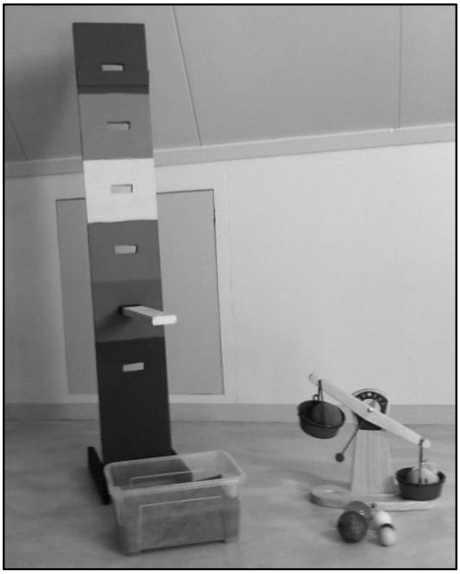
Gravity sequence: Task 9

Name: Crater maker (Dutch name: Kraters maken)

Origin: Own design, based on administration of tasks so far.

Materials used: Wood, paint, small transparent box (hardware store), toy scale, balls of different size and weight (toy store), sand.

Short description: The aim of this task is to compare the impact of balls of different weights, which are released from different heights by means of an adjustable diving board. The child is first asked what the object is for, and to release a variety of balls. The deeper the crater, the heavier the ball, but also the higher the point of release, the deeper the crater. The child is asked to predict differences between the craters and to explain these. The child is asked how a bigger crater can be made using a lighter ball, and how a smaller crater can be made by using a heavier ball.



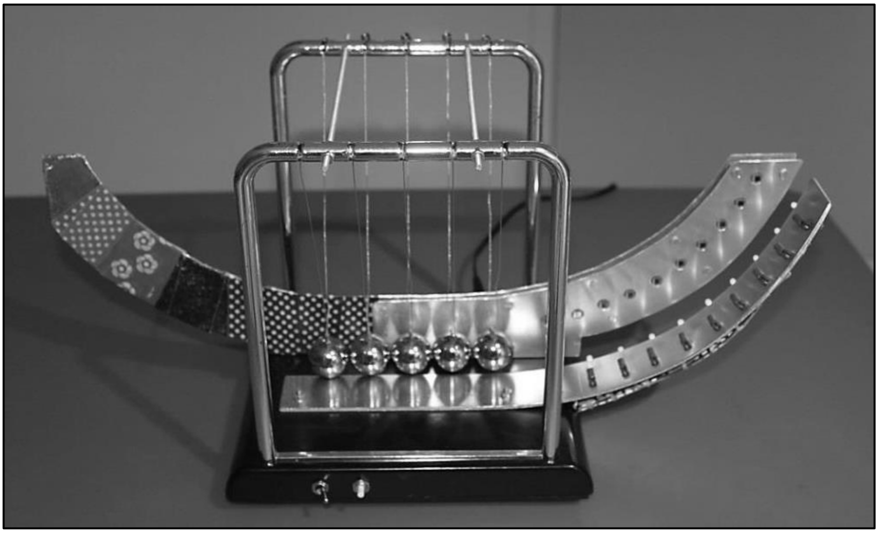
Gravity sequence: Task 10

Name: (electric) Cradle (Dutch name: electric cradle)

Origin: Own design, based on administration of tasks so far. Electric variant of the cradle made by the research instrument service of the faculty of behavioral and social sciences in Groningen.

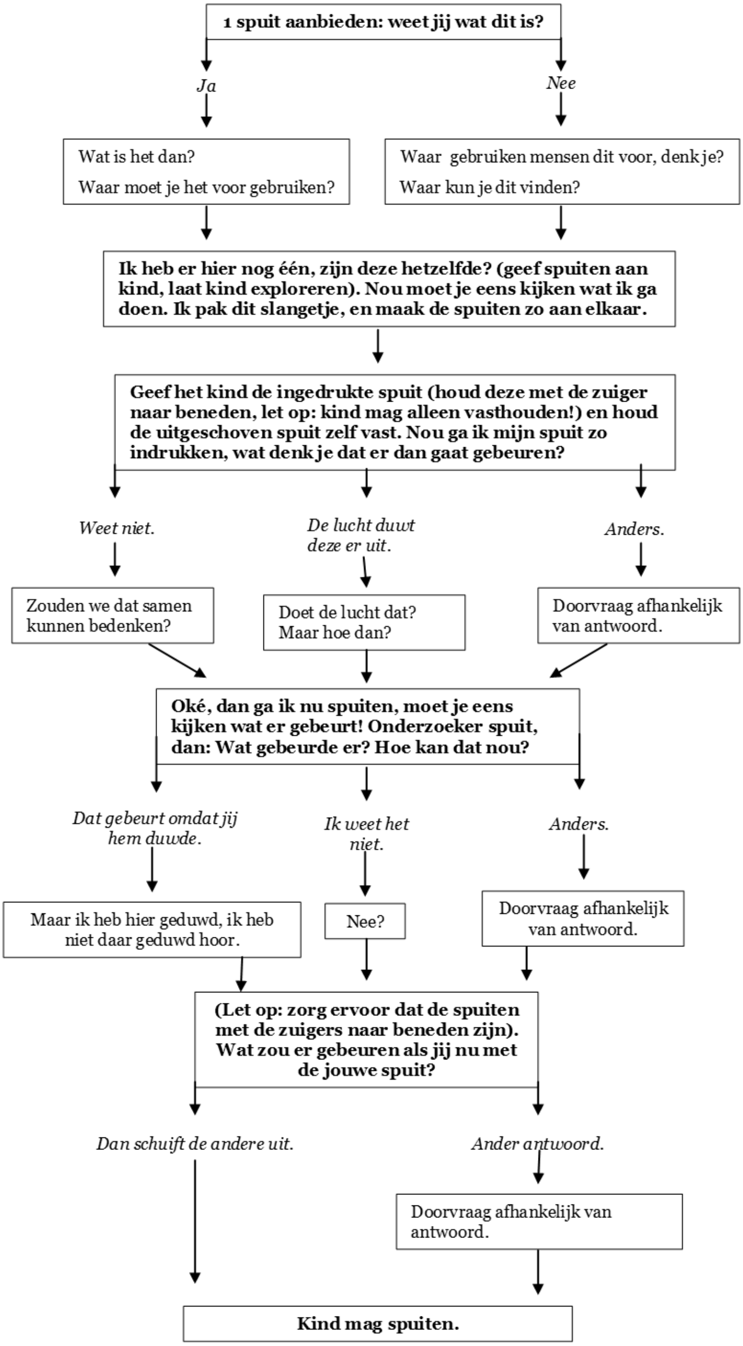
Materials used: Newton's cradle (office supply store), electric wiring and accompanying components, sensors, small light-emitting diodes (hardware store), colored tape (office supply store).

Short description: First a normal cradle (without electricity) is presented. The child is asked what the object is for and if he/she can lift and release one of the balls. Then the child is asked to explain the effect. Subsequently, more balls are released, and effects are predicted and explained. Releasing 3 balls results in the central ball swinging without any apparent interruption. The child is asked to explain this phenomenon. Subsequently, the electric cradle is presented and the child is asked how it works and how the lights go on. Questions such as "Can you make 8 lights go on?" and "How many lights will go on when you release the ball from the green part?" "What if you release 2 from the green part?" are asked, which leads to an exploration of the effect of releasing balls from different distances.



Appendix B: Example of protocols (in Dutch)

Protocol task 2 air pressure sequence: Air squirt



1 spuit aanbieden: weet jij wat dit is?

Ja

Wat is het dan?
Waar moet je het voor gebruiken?

Nee

Waar gebruiken mensen dit voor, denk je?
Waar kun je dit vinden?

Ik heb er hier nog één, zijn deze hetzelfde? (geef spuiten aan kind, laat kind exploreren). Nou moet je eens kijken wat ik ga doen. Ik pak dit slangetje, en maak de spuiten zo aan elkaar.

Geef het kind de ingedrukte spuit (houd deze met de zuiger naar beneden, let op: kind mag alleen vasthouden!) en houd de uitgeschoven spuit zelf vast. Nou ga ik mijn spuit zo indrukken, wat denk je dat er dan gaat gebeuren?

Weet niet.

Zouden we dat samen kunnen bedenken?

De lucht duwt deze er uit.

Doet de lucht dat? Maar hoe dan?

Anders.

Doorvraag afhankelijk van antwoord.

Oké, dan ga ik nu spuiten, moet je eens kijken wat er gebeurt! Onderzoeker spuit, dan: Wat gebeurde er? Hoe kan dat nou?

Dat gebeurt omdat jij hem duwde.

Maar ik heb hier geduwd, ik heb niet daar geduwd hoor.

Ik weet het niet.

Nee?

Anders.

Doorvraag afhankelijk van antwoord.

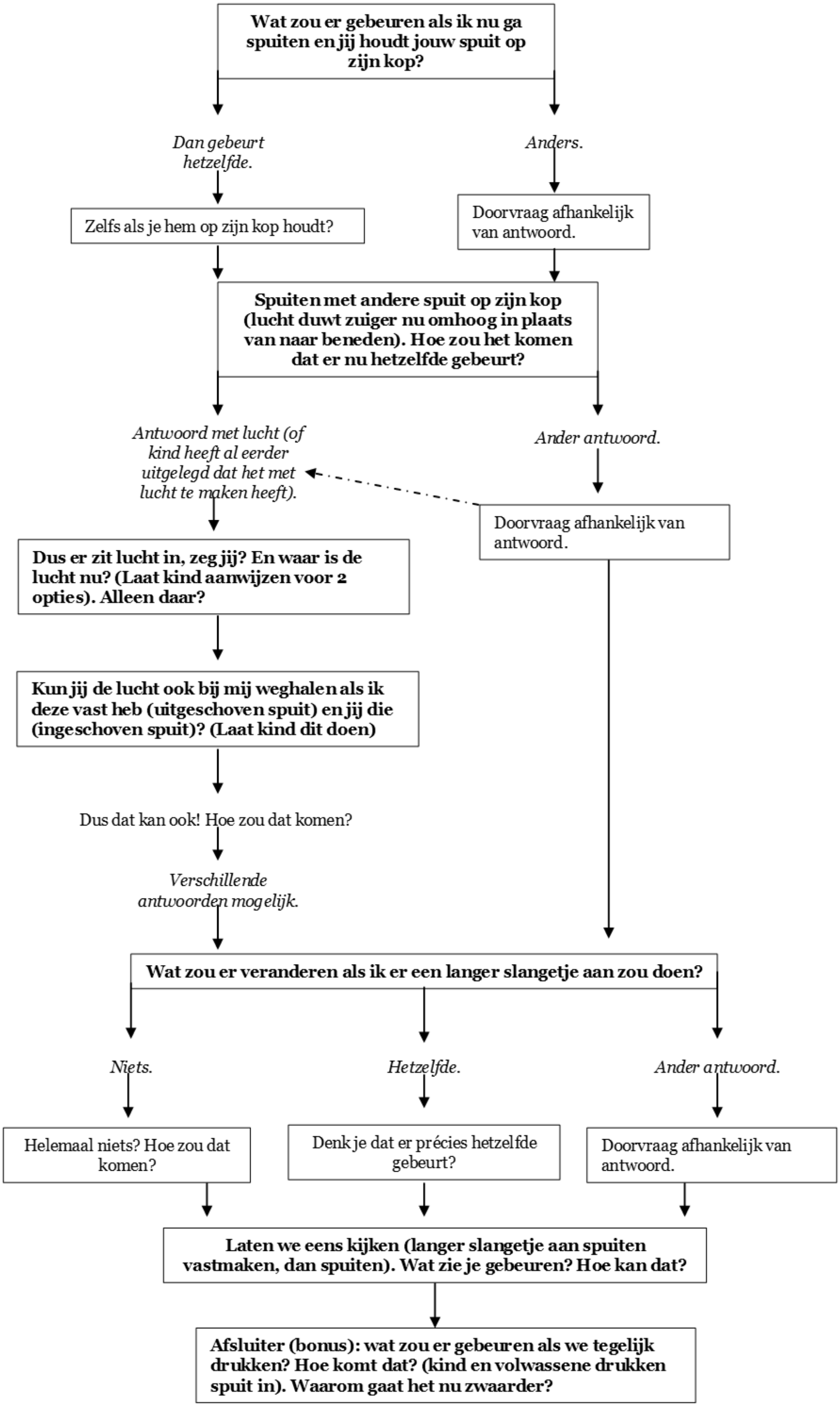
(Let op: zorg ervoor dat de spuiten met de zuigers naar beneden zijn). Wat zou er gebeuren als jij nu met de jouwe spuit?

Dan schuift de andere uit.

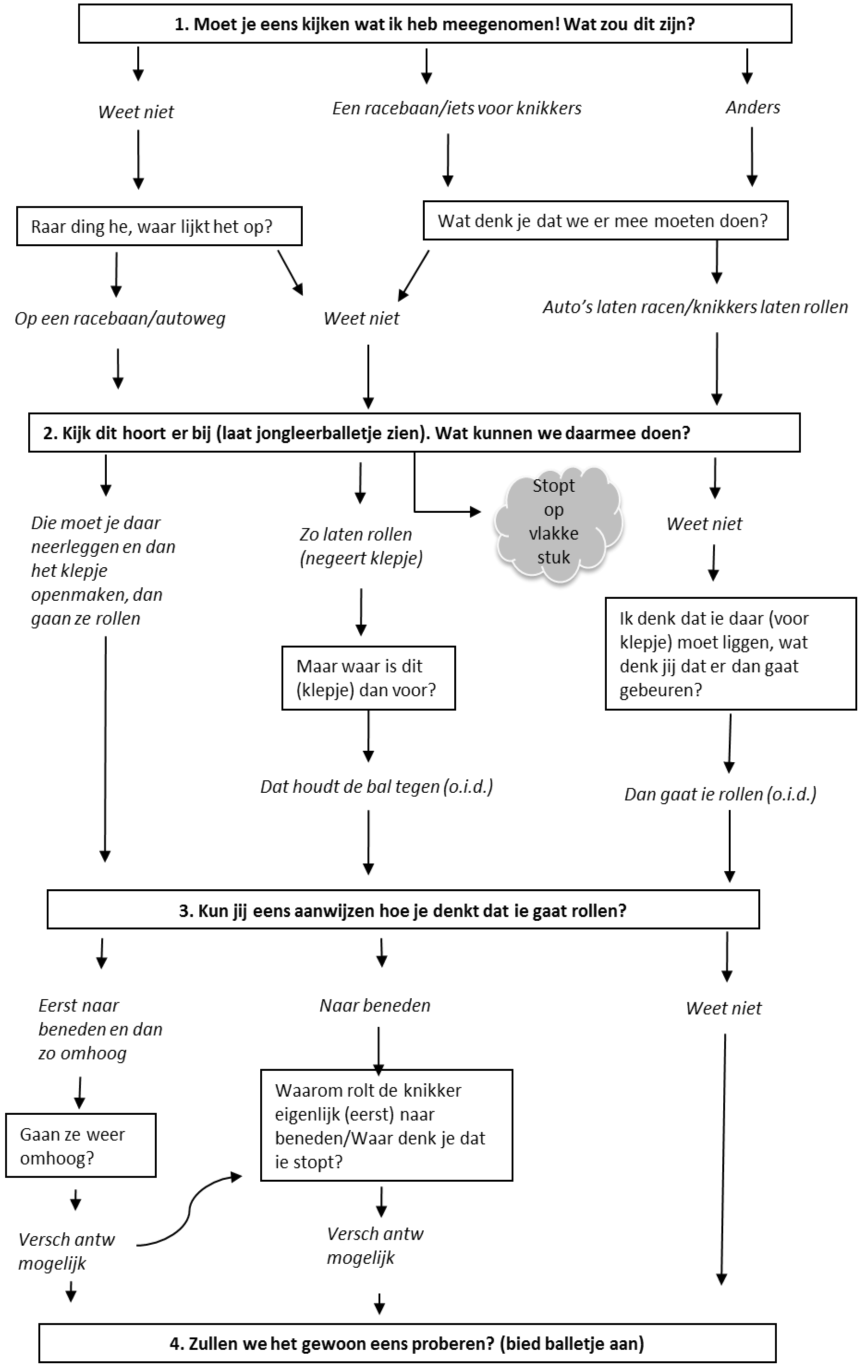
Ander antwoord.

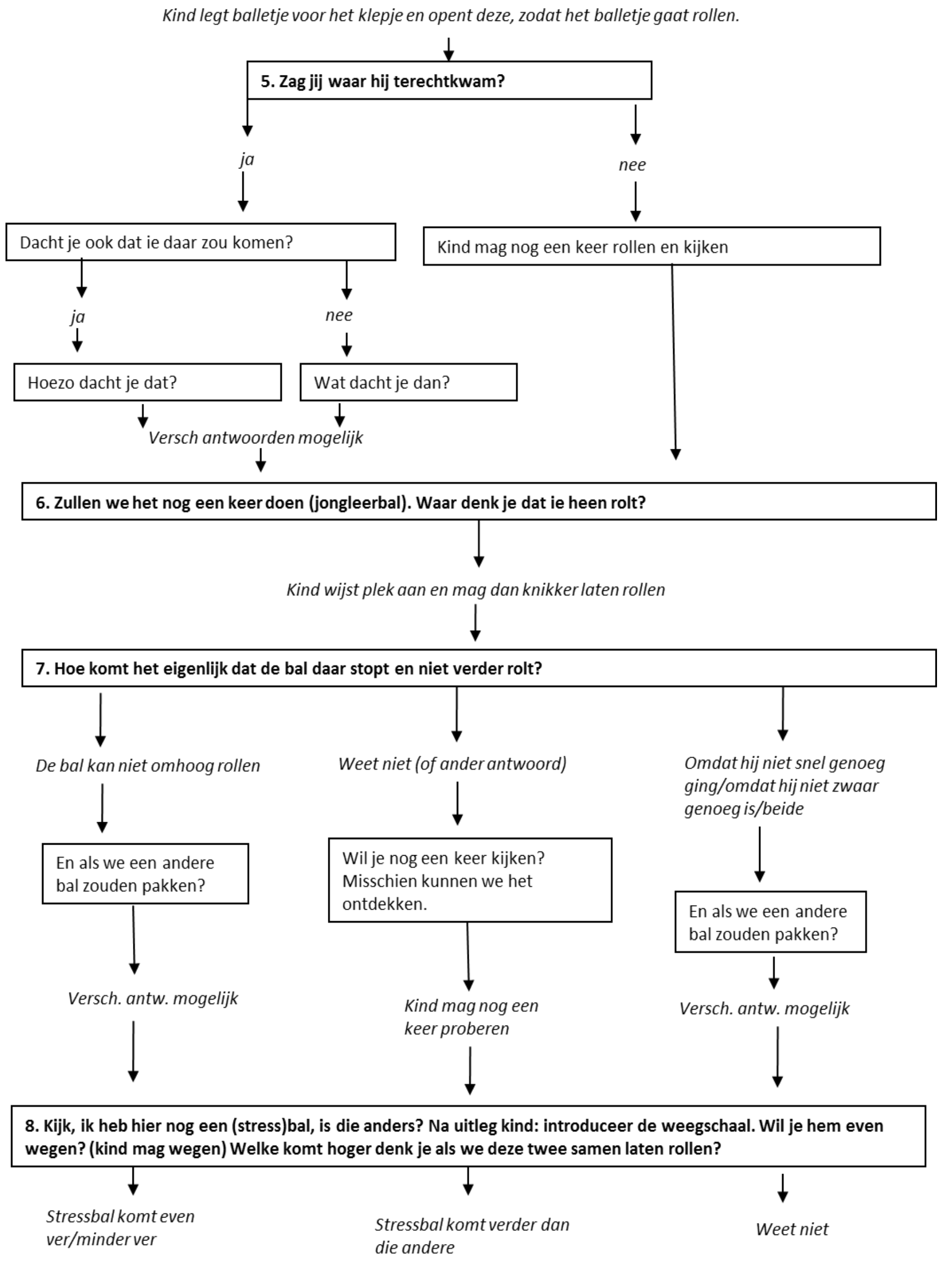
Doorvraag afhankelijk van antwoord.

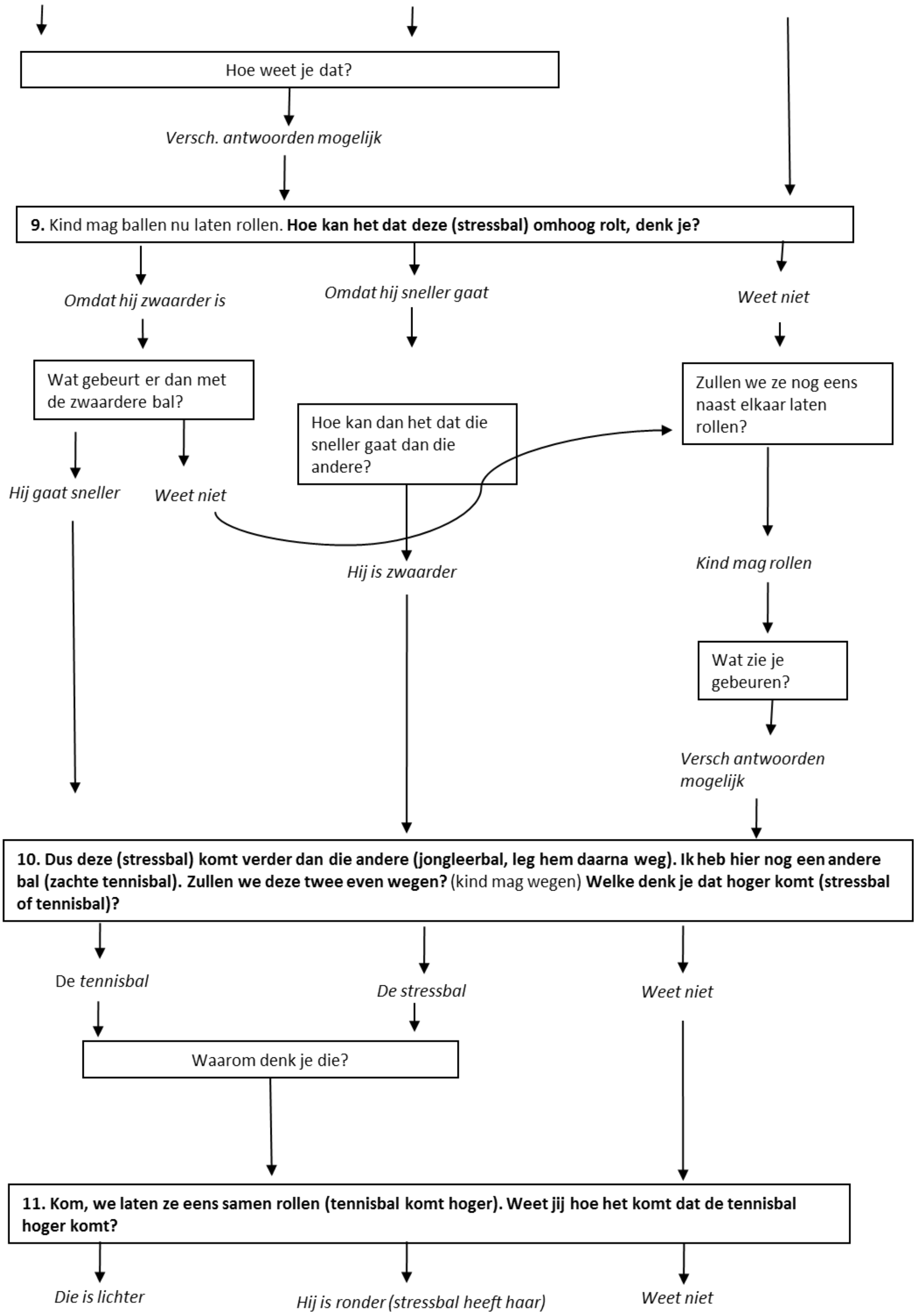
Kind mag spuiten.


Wat zou er gebeuren als ik nu ga spuiten en jij houdt jouw spuit op zijn kop?
Dan gebeurt hetzelfde.
Anders.
Zelfs als je hem op zijn kop houdt?
Doorvraag afhankelijk van antwoord.
Spuiten met andere spuit op zijn kop (lucht duwt zuiger nu omhoog in plaats van naar beneden). Hoe zou het komen dat er nu hetzelfde gebeurt?
Antwoord met lucht (of kind heeft al eerder uitgelegd dat het met lucht te maken heeft).
Ander antwoord.
Doorvraag afhankelijk van antwoord.
Dus er zit lucht in, zeg jij? En waar is de lucht nu? (Laat kind aanwijzen voor 2 opties). Alleen daar?
Kun jij de lucht ook bij mij weghalen als ik deze vast heb (uitgeschoven spuit) en jij die (ingeschoven spuit)? (Laat kind dit doen)
Dus dat kan ook! Hoe zou dat komen?
Verschillende antwoorden mogelijk.
Wat zou er veranderen als ik er een langer slangetje aan zou doen?
Niets.
Hetzelfde.
Ander antwoord.
Helemaal niets? Hoe zou dat komen?
Denk je dat er précies hetzelfde gebeurt?
Doorvraag afhankelijk van antwoord.
Laten we eens kijken (langer slangetje aan spuiten vastmaken, dan spuiten). Wat zie je gebeuren? Hoe kan dat?
Afsluiter (bonus): wat zou er gebeuren als we tegelijk drukken? Hoe komt dat? (kind en volwassene drukken spuit in). Waarom gaat het nu zwaarder?

Protocol task 3 air pressure sequence: Air squirt




Kind legt balletje voor het klepje en opent deze, zodat het balletje gaat rollen.
5. Zag jij waar hij terechtkwam?
ja
nee
Dacht je ook dat ie daar zou komen?
Kind mag nog een keer rollen en kijken
ja
nee
Hoezo dacht je dat?
Wat dacht je dan?
Versch antwoorden mogelijk
6. Zullen we het nog een keer doen (jongleerbal). Waar denk je dat ie heen rolt?
Kind wijst plek aan en mag dan knikker laten rollen
7. Hoe komt het eigenlijk dat de bal daar stopt en niet verder rolt?
De bal kan niet omhoog rollen
Weet niet (of ander antwoord)
Omdat hij niet snel genoeg ging/omdat hij niet zwaar genoeg is/beide
En als we een andere bal zouden pakken?
Wil je nog een keer kijken? Misschien kunnen we het ontdekken.
En als we een andere bal zouden pakken?
Versch. antw. mogelijk
Kind mag nog een keer proberen
Versch. antw. mogelijk
8. Kijk, ik heb hier nog een (stress)bal, is die anders? Na uitleg kind: introduceer de weegschaal. Wil je hem even wegen? (kind mag wegen) Welke komt hoger denk je als we deze twee samen laten rollen?
Stressbal komt even ver/minder ver
Stressbal komt verder dan die andere
Weet niet



Hoe weet je dat?

Versch. antwoorden mogelijk

9. Kind mag ballen nu laten rollen. **Hoe kan het dat deze (stressbal) omhoog rolt, denk je?**

Omdat hij zwaarder is

Wat gebeurt er dan met de zwaardere bal?

Hij gaat sneller

Weet niet

Omdat hij sneller gaat

Hoe kan dan het dat die sneller gaat dan die andere?

Hij is zwaarder

Weet niet

Zullen we ze nog eens naast elkaar laten rollen?

Kind mag rollen

Wat zie je gebeuren?

Versch antwoorden mogelijk

10. Dus deze (stressbal) komt verder dan die andere (jongleerbal, leg hem daarna weg). Ik heb hier nog een andere bal (zachte tennisbal). Zullen we deze twee even wegen? (kind mag wegen) **Welke denk je dat hoger komt (stressbal of tennisbal)?**

De *tennisbal*

De stressbal

Waarom denk je die?

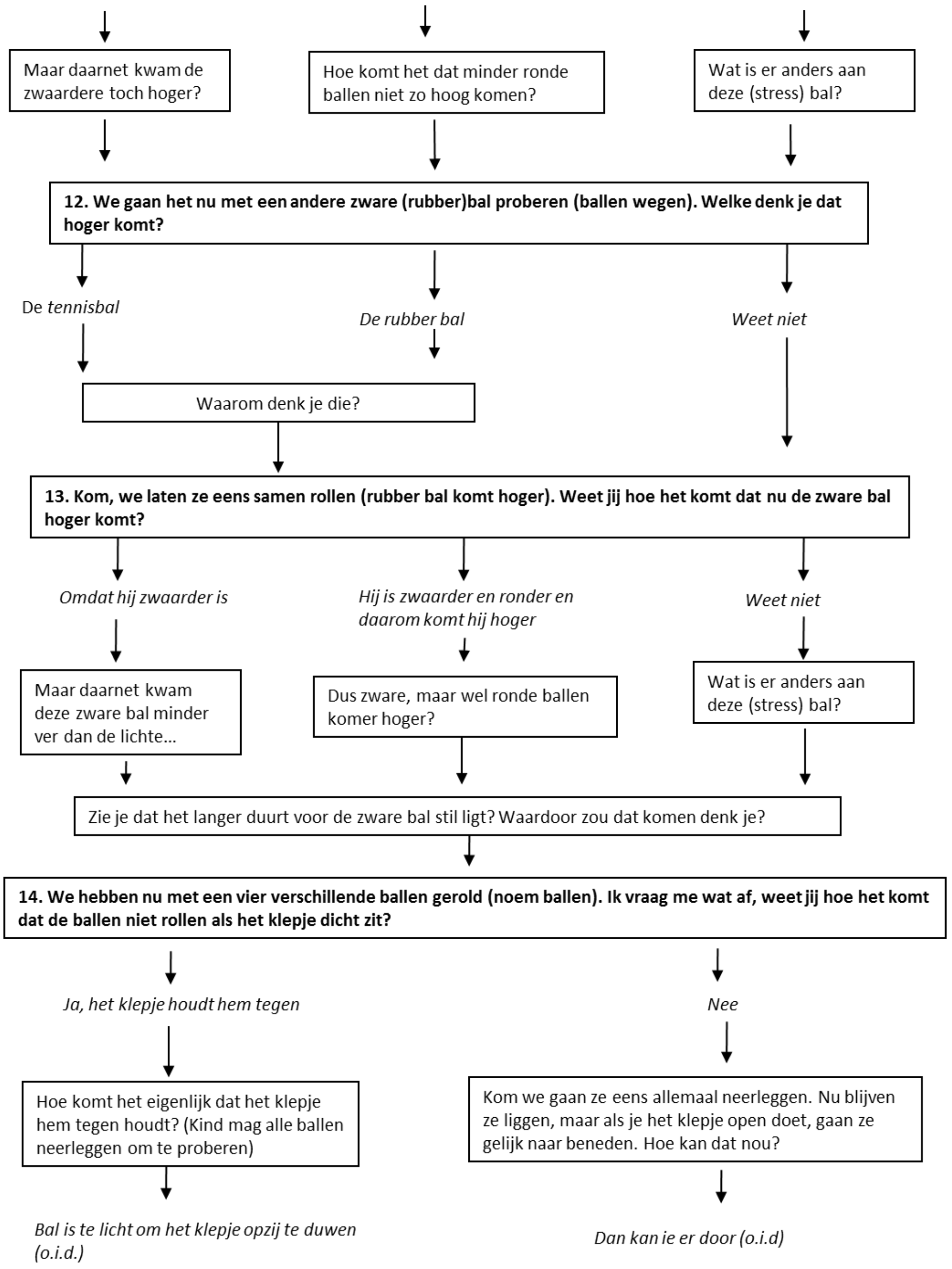
Weet niet

11. Kom, we laten ze eens samen rollen (tennisbal komt hoger). Weet jij hoe het komt dat de tennisbal hoger komt?

Die is lichter

Hij is ronder (stressbal heeft haar)

Weet niet



Maar daarnet kwam de zwaardere toch hoger?

Hoe komt het dat minder ronde ballen niet zo hoog komen?

Wat is er anders aan deze (stress) bal?

12. We gaan het nu met een andere zware (rubber)bal proberen (ballen wegen). Welke denk je dat hoger komt?

De *tennisbal*

De rubber bal

Weet niet

Waarom denk je die?

13. Kom, we laten ze eens samen rollen (rubber bal komt hoger). Weet jij hoe het komt dat nu de zware bal hoger komt?

Omdat hij zwaarder is

Hij is zwaarder en ronder en daarom komt hij hoger

Weet niet

Maar daarnet kwam deze zware bal minder ver dan de lichte…

Dus zware, maar wel ronde ballen komer hoger?

Wat is er anders aan deze (stress) bal?

Zie je dat het langer duurt voor de zware bal stil ligt? Waardoor zou dat komen denk je?

14. We hebben nu met een vier verschillende ballen gerold (noem ballen). Ik vraag me wat af, weet jij hoe het komt dat de ballen niet rollen als het klepje dicht zit?

Ja, het klepje houdt hem tegen

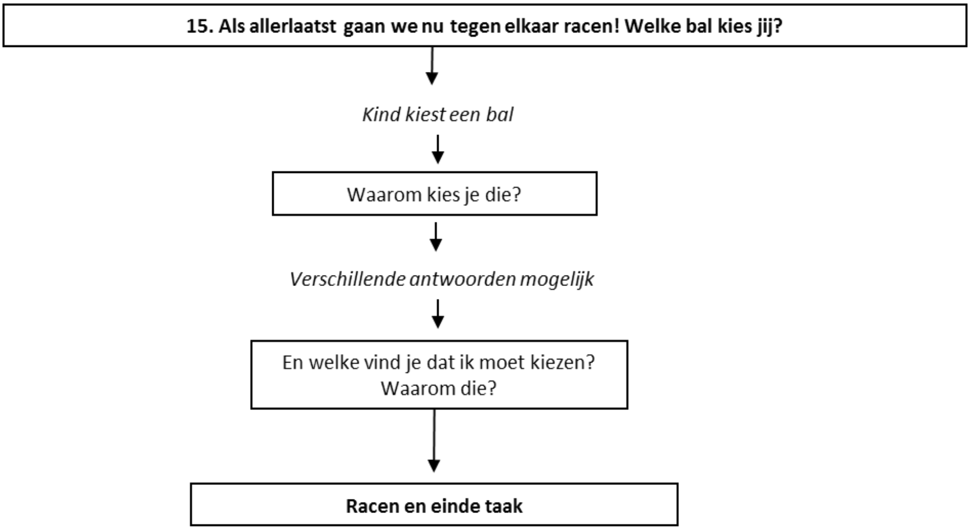
Nee

Hoe komt het eigenlijk dat het klepje hem tegen houdt? (Kind mag alle ballen neerleggen om te proberen)

Kom we gaan ze eens allemaal neerleggen. Nu blijven ze liggen, maar als je het klepje open doet, gaan ze gelijk naar beneden. Hoe kan dat nou?

Bal is te licht om het klepje opzij te duwen (o.i.d.)

Dan kan ie er door (o.i.d)


15. Als allerlaatst gaan we nu tegen elkaar racen! Welke bal kies jij?
Kind kiest een bal
Waarom kies je die?
Verschillende antwoorden mogelijk
En welke vind je dat ik moet kiezen? Waarom die?
Racen en einde taak

Appendix C: Coding of verbal expressions

In order to determine the boy's levels of understanding continuously throughout the task, the verbal expressions were coded in four steps using the computer program MediaCoder (Bos & Steenbeek, 2006). The coding procedure consisted of the following steps.

1. We started with the determination of the exact points in time when utterances of both the boy and researcher started and ended.
2. The second step involved the classification of these verbal utterances. The researcher's utterances were classified into descriptive, predictive, and explanatory questions; expressions of encouragement; follow-up questions; compliments; short explanations; procedural remarks; directing the boy's focus, and remaining utterances that could not be classified. While descriptive questions focus on the current state of the task, predictive questions are directed to future states ("What do you think *will* happen if we push the piston of this syringe?"), and explanatory questions focus on the mechanism of the task ("*How* do you think this works?") The boy's verbal utterances were classified into descriptive, predictive, and explanatory answers; requests; content-related questions, and remaining utterances that could not be classified.
3. As a third step, meaningful units of the boy's coherent expressions were formed (units of analysis). That is, when the boy's task-related answers (descriptions, predictions, or explanations) had a pause in between, but were nonetheless focused on the same topic, these answers were joined together. Each unit ended when the next expression of the boy fell into another category, or when the researcher interrupted the boy (e.g., by asking another question, or by making a procedural remark). An exception was made for expressions of encouragement. If the researcher only encouraged the boy to elaborate, the unit of analysis would not end.
4. In the fourth and final step, the complexity of the boy's answers within a unit of analysis, and the complexity of the researcher's questions were determined. This meant that each unit (for the boy) and each descriptive,

predictive, explanatory, and follow-up question (for the researcher) were rated on a ten level scale, based on the model of dynamic skill theory developed by Fischer (1980). Other utterances, such as compliments or procedural remarks were not rated using the complexity scale, and were set on 0. The complexity levels of the questions and answers ranged from single sensorimotor actions (Level 1) to representational mappings (Level 5); these levels corresponded approximately to the boy's age (see Fischer & Bidell, 2006). At Level 1 (sensorimotor actions), the boy stated single characteristics of the task, such as "This tube is long". At Level 2 (sensorimotor mappings), two elements of the task were coupled, such as "I can push this [piston] into here [the tube of the syringe]". At Level 3 (sensorimotor systems), simple causal mechanisms were stated, such as "If I push this [piston] in, the other one goes upward". At Level 4 (single representations), two causal mechanisms were coupled, or an "invisible" causal mechanism was mentioned, such as "When I push this [piston] in, air causes the other one to move upward". At Level 5 (representational mappings), the boy explained or predicted in terms of two causal relationships including an additional step, e.g., "The piston pushes the air down, which goes through the tube to the other syringe, which piston then gets pushed out by the air". When the boy only answered with "yes" or "no" to a close-ended question, his answers were simply rated as correct or incorrect. More extensive incorrect, irrelevant, and "don't know"-answers were rated as incorrect. To make sure that the "False" category was a legitimate part of the ordinal scale of complexity levels, we checked whether there was any (observable) complex reasoning behind the false answers. This was not the case; they were simple and false, therefore comprising the lowest possible category of complexity in this study.

5. The level assigned to the researcher's questions always comprised the lowest, yet accurate, level on which the question could be answered. For example, Level 1 questions of the researcher focused on single observable characteristics of the task, such as "Is the syringe big?" Questions of the researcher on level 5 were questions that could only be answered if two

causal relationships were coupled, such as: “Why is the air going out of my syringe if you pull the piston of the syringe you are holding?” No differences between close-ended or open-ended questions were made.

6. The questions and units of answers received a code on an ordinal scale from 1 to 5 (ranging from sensorimotor actions to representational mappings). The coding 0 was used to mark the end of each utterance, and for utterances that were not assigned complexity levels. The coding -1 was given to irrelevant, wrong, and “don’t know” answers. The coding 0.5 was given when the boy simply answered a close-ended question right. No wrong answers to close-ended questions occurred.

Appendix D: Technical details of data analyses

In this appendix, we explain the variety of statistical and smoothing techniques we used in more detail. In the case of *n* = 1 studies, Monte Carlo permutation tests are beneficial because they do not require a certain sample size, and no underlying assumptions have to be met (Todman & Dugard, 2001). Taking the sample distribution into account, the Monte Carlo test measures the probability that a difference is caused by chance. This is done by drawing 1000 random samples from the original data, after which one can determine how often the observed or a bigger difference occurs in these random samples (positive cases). This number of positive cases is then divided by the number of drawn samples (1000), which produces a *p*-value comprising the probability that the observed difference occurs in this distribution of 1000 random samples. If the probability that this occurs is small under the null hypothesis that the difference is zero, we can conclude that the observed difference is not merely caused by chance, and that it is a genuine difference. In this chapter, this procedure was used to compare the boy's fluctuations (mean absolute difference between two subsequent complexity levels) during session 1; to compare the frequencies of his complexity levels over the course of the three sessions, and to examine differences in the number of simultaneous in- and decreases and initiations over the course of three sessions. We decided to report all interesting differences, which we defined as all differences with a *p*-value of 0.1 or lower.

To reveal existing trends regarding the covariation between boy and researcher, we smoothed the raw time series of their complexity levels. For the most optimal picture, we smoothed the data twice using a Loess (local regression) smoothing technique. First, a bandwidth of 10% (e.g., the data of 50 adjacent seconds of the total 498 seconds were used to fit each local polynomial) was used to preserve all interesting local details. To straighten out small irregularities, we smoothed it again with a bandwidth of 10%. The resulting curve shows an estimation of the complexity levels over time, using weighted least squares to fit each local point in the graph, with more weight given to the complexity levels near the local point that is estimated (Jacoby, 2000). A linear trend line was fitted to see if the complexity levels of the boy and researcher would increase or decrease over the course of session 1. Note: For all smoothed Loess curves in this chapter, we used the raw data series as a starting point to stay close to the data found in this study, and to prevent any deformation of the graphs. We did, however, perform alternative analyses to check if *removing* all utterances that were put on zero (utterances that were not assigned a complexity level, such as procedural remarks) would change the graphs or the outcome of our statistical tests to a great extent, which was not the case. The results of these alternative analyses are available from the first author upon request.

The smoothed graphs were normalized using a linear transformation, so that the complexity levels of the boy and researcher were put on the same scale (with the minimum complexity level of each interaction partner set on 0 and the maximum complexity level set on 1). This provided a detailed picture of how increases and decreases in complexity level of the boy and researcher related to one another. To see whether patterns in the interaction would change over time, these normalized smoothed Loess curves were also fitted for the two subsequent visits.

Using the normalized smoothed time series, we repeatedly calculated the covariance between the boy and researcher while shifting the researcher's graph stepwise alongside the graph of the boy. The last column of Table 7 displays how many time points the researcher's graph has to shift to get the most overlap with the boy's graph.

Dankwoord

Groningen, 30 maart 2014

Vanaf mijn dakterras kan ik een groot deel van de stad Groningen zien. Een ideale plek (mits het mooi weer is, en dat is het) om terug te denken aan de afgelopen jaren die ik in deze stad heb doorgebracht en in het bijzonder aan de mensen die mij tijdens het promotietraject hebben gesteund, geholpen en gesterkt.

Eerst Paul en Henderien, de twee mensen zonder wie dit proefschrift er niet was geweest. Bedankt voor de kans om dit project in te duiken. Paul, je bezit zo veel talenten dat ik nog niet heb kunnen ontdekken wat nou je grootste is. Je hebt de afgelopen jaren een fantastische warme afdeling gecreëerd waar je altijd voor iedereen klaarstaat en waar ik me altijd erg thuis heb gevoeld. Ik heb veel geleerd van je mooie en creatieve schrijfstijl. De analysemethoden die je verzint of ontdekt zijn even ingenieus als doeltreffend (ik moest bij een nieuwe analysemethode altijd even flink bijzetten om je behendige Excel-acties te volgen). Ik zal nog heel lang Pauls functies blijven gebruiken. Henderien, je lieve warme persoonlijkheid maakte het altijd heel fijn om met je te werken. Je kon altijd het overzicht zien als ik zelf chaos zag. Meer dan eens heeft jouw gestructureerde manier van denken mij geholpen een artikel te beginnen, te herorganiseren of te eindigen. Onze reis naar Toronto is één van de hoogtepunten van de afgelopen jaren geweest en hierdoor heb ik je persoonlijk nog beter leren kennen. Ik hoop nog lang met je te kunnen samenwerken en nog lang buiten het werk om met je te kunnen praten. Tijdens mijn avonden als “back-up” oppas van Esther kom ik altijd erg tot rust. Bedankt dat ik haar even mocht lenen voor de voorkant.

Ik wil daarnaast de leescommissie bedanken: Kurt Fischer (thank you), Maartje Raijmakers en Alexander Minnaert.

Dan de leuke kinderen die ik voor dit onderzoek mocht bezoeken. De schoolbezoeken eens in de drie à vier maanden waren altijd een feestje. Jullie ideeën over de taakjes en jullie uitspraken maakten het altijd meer dan waard. R., ik vind het nog altijd jammer dat ik niet “het spel met die monsters” voor je heb kunnen meenemen, maar op de “school voor het maken van speelgoed” zijn we nog niet toegekomen aan computerspellen. B., met 4 zwaarden thuis heb je vast nog steeds de allermeeste zwaartekracht ooit. Jouw opmerking hierover is me altijd bijgebleven. S., je broer had gelijk. Ik wist al hoe de taakjes werkten. Wat ik echter niet wist, was of en hoe ze voor jou zouden werken. Ik wil ook jullie ouders en leraren hartelijk bedanken voor de mogelijkheid om jullie te komen bezoeken.

Het lastige aan het volgen van kinderen over een langere periode is dat ze van school kunnen wisselen. Bedankt, alle scholen die niet bekend waren met mij en het project, maar wel bereid waren hun school open te stellen voor mij om kinderen verder te kunnen volgen. In het bijzonder wil ik Gera Brouwer bedanken, die mij niet alleen op haar eigen school toeliet, maar ook met andere scholen in contact bracht. Daarnaast ook dank aan Carla Vink, die mijn bezoeken aan haar school coördineerde, waardoor ik heel efficiënt met verschillende kinderen van haar school aan de taakjes kon werken. In het dorp van mijn ouders mocht ik eerst op de school (die nu helaas gesloten is) de eerste taakjes uittesten, wat ik heel fijn vond. Bij mijn oud-oppaskindjes Dylan en Zoë en hun ouders kon ik terecht om latere taakjes uit te proberen, bedankt!

Een aantal taakjes was er niet geweest zonder Rutger Meissner, wiens creatieve ingenieursbrein ik af en toe mocht lenen. Hartelijk dank voor het meedenken, het maken van de ballenbak en het zoeken naar materialen. Dikwijls werd er dankzij jou een taakje geboren, vaak werden taken door jouw ideeën geperfectioneerd. Van Rooske Franse van Nemo mocht ik het taakje lenen dat ik “de Nemobak” heb gedoopt, hartelijk dank daarvoor. Daarnaast hartelijk dank aan Pieter Zandbergen, voor het herhaaldelijk lenen van camera’s en het terugvinden van data op een harde schijf die op mysterieuze wijze het leven liet. Remco, bedankt voor het “elektriseren” van de cradle, tot op de dag van vandaag het spannendste taakje! Pablo, fijn dat je op het laatst in wilde springen bij het maken van de voorkant van dit proefschrift. Lieve Lucia, bedankt dat ik altijd met praktische vragen bij je terecht kon. Marijn, je bent van grote waarde geweest bij de artikelen waar hoofdstuk 5 en 6 op zijn gebaseerd. Jenny and Dianne, thank you for having me over during the summer of 2010 and for making me part of your research team.

Lieve studenten, bedankt voor het vroege opstaan, het meereizen naar de andere kant van het land en het gezellige appeltaart eten. Ik weet dat het coderen niet altijd even leuk was (dit is een understatement), maar jullie hebben mij er enorm mee geholpen. Lisette, Lotte en Marijke extra bedankt voor het meegaan als ik even geen ondersteuning had. Lisette, fijn dat je de data ook kunt gebruiken voor je onderzoek, het is fijn om met je samen te werken.

Alle lieve collega's en mede-aio's van de afdeling Ontwikkelingpsychologie, bedankt voor de gezellige en leerzame tijd. Heidi en Daan, bedankt voor jullie input tijdens de eerste fase van het onderzoek, ik heb er veel aan gehad. Collega's vanuit het hele land die verbonden zijn aan TalentenKracht, bedankt voor het delen van jullie kennis tijdens bijeenkomsten en congressen. In het bijzonder een bedankje voor de Utrechtse collega's van het allereerste begin, die mij een aantal van hun taakjes als prototype lieten gebruiken. Lieve kamergenoten, Marieke B., Sabine, Tooske, Marieke V. en Annemieke, bedankt voor de kopjes thee en jullie hulp tijdens de dagelijkse werkzaamheden. Naomi, bedankt voor de fijne samenwerking tijdens de cursus Developmental Psychology en voor je leuke gezelschap tijdens de reis naar Florida (dat je me ook beter mee had kunnen nemen naar Austin, is nu wel bewezen, geloof ik….).

Een combinatie van lieve collega's en vriendinnen vond ik in Marieke, Annika, Elisa en Charmaine. Ik vind het heel bijzonder om deel uit te maken van jullie leven. Marieke, ik heb zo ontzettend veel van je geleerd, vooral van je positieve en sociale instelling. Als ik het even niet weet, denk ik vaak: “Wat zou Marieke doen?” Annika, je lieve zachte aard is om jaloers op te zijn, ik probeer er een voorbeeld aan te nemen. Elisa, aan je luisterend oor en fantastisch gevoel voor humor heb ik altijd veel gehad. Charmaine, the glass is always half full when you are around. A little bit of Maltese sunshine in Groningen!

Welmoed, ik mocht af en toe meeliften op jouw plezier en levenslust, dank daarvoor. Je kan mij altijd opbeuren en hard laten lachen. Ik ben daarom heel blij dat je op 8 mei als paranimf naast me staat.

Mijn lieve ouders hebben van hun hooizolder een pension voor gebruikte taakjes en onderdelen gemaakt. Bedankt voor het bieden van een warm onderdak en dan bedoel ik niet alleen voor de taakjes. Bij mijn vader en broertje heb ik ongekende talenten ontdekt toen zij aan de taakjes werkten. Pap, meer dan eens ging je voor mij naar de bouwmarkt. Zelfs wanneer een constructie onmogelijk leek, wist jij een uitweg. Het luchtkanon en de black box waren er niet geweest zonder jouw idee om onderdelen van tuinsproeiers te gebruiken. Ik snap nog steeds niet hoe je op dat idee bent gekomen, maar het werkte!

En dan Ruud, mijn lieve vriendje voor altijd. Zonder jou was dit proefschrift er niet geweest. Je was er altijd voor me en zelfs toen ik dacht dat je voor langere tijd naar Frankrijk zou gaan, heb je het toch klaargespeeld om bijna altijd bij mij te zijn. Dankzij die “move” kan ik nu elke dag lachen om je lange verhalen. Je staat naast me op 8 mei en dan is het jouw beurt om je proefschrift af te schrijven. En daarna? Dan gaan we samen de wereld veroveren*. Just you and I, defying gravity!*

Curriculum Vitae

Steffie van der Steen is geboren op 18 mei, 1986 te Leiderdorp. In 2007 heeft zij haar Bachelor in psychologie (cum laude) afgerond aan de Rijksuniversiteit Groningen en ontving zij een Huygens Talentenbeurs om haar master te doen aan Harvard University. In 2008 rondde zij haar Master in Mind, Brain, and Education af en nam zij de Intellectual Contribution/Faculty Tribute Award in ontvangst. In september 2008 startte zij met haar promotie onderzoek onder begeleiding van prof. dr. Paul van Geert en dr. Henderien Steenbeek. Naast het promotie-onderzoek, gericht op de wetenschappelijke ontwikkeling van kinderen, heeft Steffie meegewerkt aan een onderzoek naar schrijfprocessen (met dr. Jenny Thomson en Dianne Samuelson, Ed.M., Harvard University/University of Sheffield) en aan een onderzoek naar voetbalinzicht (met Ruud den Hartigh, Msc., Rijksuniversiteit Groningen en Université Montpellier 1). Sinds eind 2012 werkt Steffie als universitair docent aan de Open Universiteit (afdeling Psychologie) en als toegevoegd docent aan de Rijksuniversiteit Groningen.

Publicaties

Van Der Steen, S., Steenbeek, H., Van Geert, P. (2012). Using the dynamics of a person-context system to describe children’s understanding of air pressure. In H. Kloos, B. J. Morris, & J. L. Amaral (Eds.), *Current Topics in Children's Learning and Cognition* (pp. 21-44).

Van Der Steen, S., Steenbeek, H., Wielinski, J., & Van Geert, P. (2012). A comparison between young students with and without special needs on Their understanding of scientific concepts. *Education Research International, 2012*. doi: 10.1155/2012/260403

Van Der Steen, S., Steenbeek, H., Van Dijk, M. W. G., & Van Geert, P. (in press). A process approach to children’s understanding of scientific concepts: A longitudinal case study. *Learning and Individual Differences*.

Van Der Steen, S., Steenbeek, H., Van Dijk, M., & Van Geert, P. (submitted). *How to characterize the development of children’s understanding of scientific concepts: A longitudinal microgenetic study*.

Den Hartigh, R.J.R., Van Der Steen, S., De Meij, M., Van Yperen, N.W., Gernigon, C., & Van Geert, P.L.C. (submitted). *Characterizing expert representations during real time action: A skill theory application to soccer.*