Res Sci Educ (2014) 44:215–237
DOI 10.1007/s11165-013-9380-z

Assessment of Preschooler's Scientific Reasoning in Adult–Child Interactions: What Is the Optimal Context?

Heidi B. Meindertsma · Marijn W. G. van Dijk ·
Henderien W. Steenbeek · Paul L. C. van Geert

Abstract In educational settings, continuous assessment of the child's level of understanding is necessary to effectively utilize the principles of scaffolding and to create contexts that can advance the scientific reasoning of the child. In this article, we argue that a child's performance is a dynamic notion that is created by all elements in an interaction, including the task. Therefore, we studied preschoolers' levels of scientific reasoning varying different properties of the assessment context. Young children were interviewed about four scientific tasks using one out of four different protocols (varying in the degree of flexibility and adaptiveness) by an adult. In the first study, different task contents resulted in different performance levels. The second study indicated that the most structured protocol elicited the highest maximum level of reasoning in children and the highest percentage of correct predictions. The third study showed differences between the protocols in the adult's verbal behavior. Adaptation in verbal behavior to different children by the adult did not result in higher scientific understanding by the children, whereas a higher degree of task structure did. Combined, the studies emphasize the importance of context, which has implications for assessment and teaching situations.

Keywords Scientific reasoning · Assessment · Dynamic skill theory · Context

Introduction

Science education in early childhood is currently a popular topic, not only in the scientific community (Eshach and Fried 2005; Gelman and Brenneman 2004) but also in the context of education policy (Duschl et al. 2007; Rocard et al. 2007). One of the reasons for introducing science at a young age is to promote the development of scientific thinking and to encourage positive thinking about science in general (Eshach and Fried 2005) that might lead to an increase in students' interest to study science topics (Mantzicopoulos et al. 2008) which is necessary to ensure enough technical personnel to maintain a high level of wealth in Western societies (Rocard et al. 2007). In several countries, programs have been initiated aimed at stimulating science and technology, for instance the "Curious Minds"

H. B. Meindertsma (✉) · M. W. G. van Dijk · H. W. Steenbeek · P. L. C. van Geert
Faculty of Behavioural and Social Sciences, University of Groningen,
Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
e-mail: h.meindertsma@rug.nl

project in the Netherlands (e.g., Steenbeek et al. 2011) and the "ScienceStart!" curriculum in the USA (e.g., French 2004). Teachers play an important role in promoting children's science learning, especially if they use scaffolding. In order to successfully scaffold in a teacher–child interaction, assessment of a child's initial level by the teacher is necessary (Ruiz-Primo and Furtak 2007; Van de Pol et al. 2010). It has been well documented that context is highly relevant for the performance of people, which is referred to as situated cognition (e.g., Fischer and Bidell 2006; Lave 1988), but the question remains to what extent properties of the assessment context affect the scientific reasoning of young children.

Consider the excerpt of an adult–child dialogue from one of the presented studies about floating and sinking below.

> The interviewer presents the apple: *What will the apple do, do you think?*
> The child responds: *Sink*
> Adult: *Why do you think it will sink?*
> Child: *Because it is heavy.*
> Adult: *Well, let's have a look.*
> The child hands the apple to the adult and the adult places the apple in the water tank. […]
> Child: *Floats*
> Adult: *That apple floats … and how is that possible?*
> [Child is distracted]
> Adult: *How is it possible that the apple floats?*
> Child: *Because it is light, but when I hold it, it is a bit heavy.*

The example shows how a child can reach a higher level of reasoning within the context of the task setting. Here, the child generates a more complex level of reasoning (instead of giving just one argument, that it is heavy, the child increases the number of arguments and combines them in saying that it is light in one context and heavy in another) in the course of one session. This increase does not necessarily mean that from now on the child will always perform at this higher level of understanding. The child is brought to this level of understanding by this particular situation: The child's skill emerges during the interaction. Contextual support can bring this child to a higher level in this situation, but it may well be that this level will not be reached in subsequent tasks. The child does not have one fixed level of performance, but it creates the present level in the situation itself (Fischer and Bidell 2006).

This emergence of a child's skill in real-time interaction with the context has been addressed by many different authors (e.g., Fischer et al. 1993; Fischer 1980; Kloos et al. 2010; Rappolt-Schlichtmann et al. 2007; Thelen and Smith 1994). It also implies that the skill exists only in that specific context at that point in time, a process called soft-assembly (Thelen 1992). The context consists not only of the people around the child but also comprises the task itself. In the case of an adult and a child performing a particular scientific reasoning task together, this reasoning context can be visualized as a triangle (Fig. 1a). This triangle (Steenbeek et al. 2011) shows that it is the interaction process between the three participating components in the situation that forms the competence level of the child.

As the triangle shows, the child is influenced by the task as well as by the adult, but at the same time also actively influences the task and the adult. The context is thus *created* by all participants together, including the child. By way of illustration, if the child manipulates the task (for instance, puts the apple in the water tank), the task itself changes as an influence of the child (the apple is now floating in the water tank). With this change, it influences the child and the adult since the situation is now different (they both see that the apple floats). Next, the adult might alter the task (for instance, by introducing a new object), which also

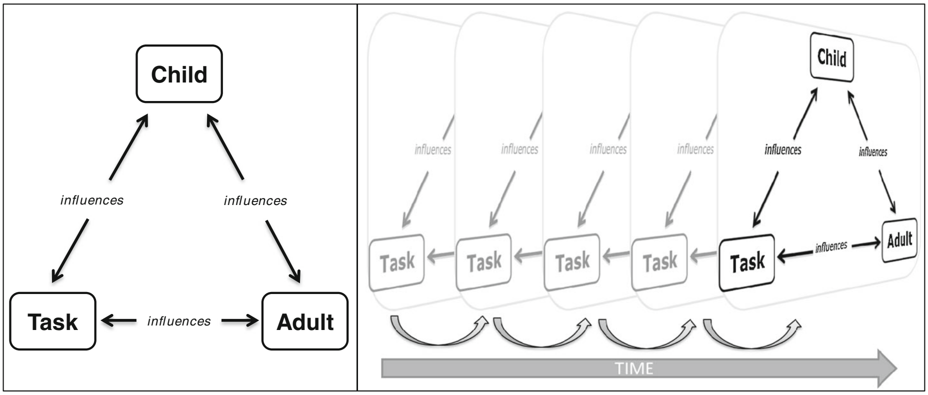

Fig. 1 *Left* talent *triangle* which shows the reciprocal relations between adult, child, and task. *Right* the *triangle* over time: Each change in the *triangle* is a new starting point for the rest of the interaction

affects both task and child. This sequencing of influences when the next step is based on the preceding one(s) is called "iterativeness" (Van Geert and Steenbeek 2005) and is represented in Fig. 1b. It is important to state that the three corners of the talent triangle cannot be seen as isolated variables (Steenbeek et al. 2011) since they are mutually dependent. However, we can focus on the role of task and adult separately and explore its relative contribution.

In our view, both task and adult can act as scaffolds for the child. The notion of scaffolding originates from Wood et al. (1976) who emphasized the role of the adult. They described scaffolding as the adult controlling those parts of the task that are too difficult for the child, so that the child can give attention to those parts that he or she is capable of performing. However, scaffolding is not limited to the support provided by an adult, but can—according to several authors—be divided into three types: (1) social scaffolding (by another person; Hsin and Wu 2011; Salonen et al. 2007), (2) ecological or task scaffolding (Hardy et al. 2006; Kloos et al. 2010; Mascolo 2005), and (3) self-scaffolding (Granott et al. 2002; Mascolo 2005). In this paper, we address only the first two types.

In the current studies, we aim to explore the role of context on the scientific reasoning of young children in an interaction with one adult and one task. Although scientific reasoning can refer to both conceptual understanding and inquiry skills (Zimmerman 2000), we define it here as conceptual understanding. Since the seminal work of Piaget (1930), many studies have focused on the scientific reasoning of children on scientific tasks such as floating and sinking (e.g., Hardy et al. 2006; Havu-Nuutinen 2005; Tytler and Peterson 2004) and the balance scale (e.g., Philips and Tolmie 2007; Pine et al. 2007; Siegler and Chen 1998), but few have actually compared performance over different task settings. However, if the child's understanding is a soft-assembled property, contextual differences in the task and/or adult role can be relevant for the assessment of reasoning skills of young children.

By systematically altering two parts of the triangle, namely the task and the use of different protocols by the adult, different task contexts are created. Of course, both adult and child are also influenced by their own specific history, personality, etc., and it is virtually impossible to disentangle the contributions of each of these components in the deployment of a particular dyadic reasoning process. Nevertheless, by describing in more detail the emergence of reasoning in conditions where the adult's behavior is guided by different instruction styles and where different tasks are used, we can gain insight in the influence of the context on the scientific reasoning skills of children. We argue that since assessment is an indispensable part of the teacher–learner interaction, we need to know more about the role of the context. In order to describe the relationships between child and task and child and adult in relation to the levels of reasoning produced, we use dynamic skill theory as a conceptual “ruler” (Fischer and Bidell 2006; Fischer 1980).

Dynamic Skill Theory

Dynamic skill theory (Fischer 1980) poses a theoretical framework for understanding the development of reasoning, both at a micro- and macroscopic level (Fischer 1980; Fischer and Bidell 2006). The dynamic part of the theory is present in the description of the self-organizing and adaptive character of skills. Skills are conceived of as arising in interaction with the context and are constantly built up and broadened in other contexts (Fischer and Bidell 2006). Since a skill is built in real time, a minor change in the context can affect the skill as it emerges in the situation (Fischer 1980). This conceptualization is consistent with the idea of the creation of context, as presented in the talent triangle.

Dynamic skill theory proposes a series of “complexity levels of understanding” that can be used to score skills for a wide range of topics among which is scientific reasoning, e.g., about floating/sinking (Rappolt-Schlichtmann et al. 2007) and about building electrical circuits (Schwartz and Fischer 2004). It is argued that the development of children occurs through three qualitatively different tiers. Each tier consists of four different levels that have been corroborated by empirical data (Fischer 1980). However, in descriptive terms, the fourth level is equal to the first level of the next tier, resulting in a description of ten distinct levels in total. Within each tier, each consecutive level is increasingly complex and differentiated and includes all underlying levels. Each level starts with a “single set” level, followed by a “mapping” level (two single sets connected), a “system” level (two mappings connected to a system), and finally a “system of systems” level that also functions as the “single set” level of the next tier.

According to skill theory, development can be portrayed as a dynamic developmental web instead of a static ladder (Fischer and Bidell 2006). The stages a person goes through are not conceived of as fixed steps but as becoming more and more differentiated where one strand can be at a high and highly differentiated level, whereas another strand can be limited to a low, less differentiated state. Furthermore, a child can be at different levels at one moment in time for different domains and even within one domain (Mascolo and Fischer 2010). This implies that there is large variability in the performance of children over and within different skill domains and in different contexts. Within one domain such as science, children might be at different levels for different tasks. For instance, a balance scale task and a floating/sinking task both refer to scientific aspects, but children might perform at different levels at each task.

The Material Context

One way that the task influences the child’s performance is by the affordances of the material. The term “affordance” was introduced by Gibson (1979) to describe which opportunities the environment offers to the person in the eye of the person. An affordance thus refers to both the task (environment) and the person. In order to illustrate this point, Gibson (1979, p. 127) gives an example of the affordances of a chair. If we see a horizontal, flat solid surface at knee height, it affords sitting. We perceive something and know instantly how to act upon it. We do not need a preceding mental representation of a chair as an object-that-is-used-for-sitting in order to show the proper action. However, affordances exist only in relation to a particular person who is characterized by particular “effectivities,” which allow that person to make use of the affordances (Chemero 2003; Shaw et al. 1982). If the horizontal, flat solid surface is actually a wall of 2 m high, most people do not automatically perceive it for sitting on. However, some people will, depending on the situation: In a normal setting, the wall does not afford sitting on, but during a busy concert, many people might see it as a great opportunity to sit (and see the artist on stage). Thus, the affordances of an object depend on the interaction between object, person, and context.

Some affordances have to be learned. Presented with a novel object, children (and adults) do not always know instantly what it affords (Gibson 1988). Exploration is needed to learn about the affordances of a task (Gibson 1988), and exploration is often seen as one of the main components of curiosity (Chak 2010). Some children are more curious than other children, but also the task affects the curiosity and exploration in children (Chak 2010). Both novelty and the difficulty of the task influence the amount of exploration (Henderson and Moore 1980; Ruff 1984).

Another way that the task can influence the performance of a child is the design and the presentation of the task. Kloos et al. (2010), for instance, showed that differences in task presentation influence the performance of children and adults on a floating/sinking task. Changes in the setup of a task can also drastically influence performance as Smith and Thelen (2003) have demonstrated in the context of their object concept experiments: By changing the physical position of infants, the infants no longer made the A-not-B-error.

In our view, a task can act as a scaffold in two separate ways: on the one hand, by the affordances it offers and on the other hand, by the structure of the task setting. Task scaffolding through affordances occurs in the interaction between child and task. These affordances may vary with regard to their familiarity and visibility. The scaffolding through task setup can also be influenced by the adult. As we shall see in the next section, the adult can also have a direct effect on the performance of the child.

The Social Context

Several studies have focused on the adult’s influence on the performance of the child in an adult–child interaction. For instance, Rappolt-Schlichtmann et al. (2007) found that the complexity of answers regarding buoyancy in children was higher if the teacher modeled complex answers than if the teacher only formulated a prediction. Henderson (1984) looked into the effect of context on explorative behavior of children. Aside from the existence of high- and low-explorative children, he found that the adult’s behavior altered the explorative behavior of the child. Hoogsteder et al. (1996) have argued that the interaction patterns should be studied, instead of only concentrating on the teaching role of the adult and consequently on the learning role of the child. By acquiring more insight on the effect of differently structured protocols, we hope to enhance our understanding of the influence of the adult on the child’s performance.

In studies on the reasoning of children, various types of protocols are used without investigating the influence of that specific protocol. For instance, Tytler and Peterson (2003) chose to use a version of the Piagetian clinical interview whereas Havu-Nuutinen (2005) constructed a semi-structured interview using an interactive approach. If there are reasons to believe that a child has a unique and true level of understanding, we should ask ourselves which type of protocol is the most appropriate one for revealing the child’s optimal level. However if, as we have argued, the child’s level of understanding emerges, within particular constraints, on the basis of a specific contextual interaction, then we should ask ourselves which particular kind of context is most appropriate for revealing which particular type of emergent understanding (e.g., a functional or an optimal level).

In educational contexts, the interaction between child and adult is often studied within the framework of social scaffolding. In this context, scaffolding is defined as providing assistance adjusted to the level of the child: In the beginning of a learning process, more assistance is needed and this can slowly be discarded when the child's learning increases (Van Geert and Steenbeek 2005; Wood et al. 1976). This implies that the scaffolding level must be higher than the child's level in order to lift the child's level up to the scaffold's level, which requires a constant adaptation by the adult to the child's performance. Mascolo (2005) speaks of "coactive scaffolding" to emphasize the interactive part of scaffolding. In terms of dynamic skill theory, we argue that scaffolding can be viewed as an attractor for the child to perform at a higher, more optimal level than the functional level of the child.

According to Van de Pol et al. (2010), there are three elements in the adult–child interaction that characterize scaffolding, which are contingency, fading, and transfer of responsibility. Contingency refers to adapting to the child's level, fading means that the scaffold is slowly reduced, and by the transfer of responsibility is meant that the child slowly takes over as learning progresses. In order to successfully implement scaffolding, the teacher should constantly use diagnostic strategies (Van de Pol et al. 2010) to assess the learner's level and the needed scaffolding level. Dynamic skill theory can be used to assess the present level of understanding in the child.

In a particular task setting, adult and child create interaction patterns that can be characterized according to their role distribution (symmetry), their (communicative) instruments, and the pursued aims (Hoogsteder et al. 1996). For instance, when using a test protocol, the role distribution may be more asymmetric than in a playful setting with the adult leading the interaction. One of the communicative instruments might be the degree of "openness" of the adult's verbal behavior. Openness is defined here as the degrees of freedom that the adult leaves for the child, with for example the adult's encouragement giving a child more degrees of freedom in its possible consecutive actions than if the adult were to give an instruction.

This Study

In the current study, we aim to explore the role of the (social and material) context in the assessment of optimal scientific reasoning skills of young children during a dyadic interaction. Although many researchers, such as Piaget (1930) and Tytler and Peterson (2003), have addressed the question of scientific reasoning of children, the present study is one of the first to compare the reasoning of the *same* children in different contexts. Since we define scientific reasoning as the child's conceptual understanding, we focus on children's predictions and their explanatory reasoning. For this purpose, we set up three separate studies that correspond to the following aims. First, with regard to the task–child relation, we aim to describe whether and how different scientific tasks elicit different performance levels (performance in the sense of correct predictions and reasoning skills) in the children. As development is seen as a developmental web (Fischer and Bidell 2006), we expect that children will show different performance levels on different tasks. The question is which of the tasks results in the highest performance in the child.

Our second aim, concerning the adult–child interaction, is to analyze how different social scaffolding contexts, in particular the different level of structure/openness given by an adult, elicit different performance levels. By limiting the freedom of initiatives and replies in the adult, the adult is hindered in the adaptation to the child's level. This means that we expect adults in the highest structured protocol to elicit the lowest level of children's scientific reasoning skills. However, material and social context also interact with each other. A stricter protocol for the adult means that the task setting is also more structured, which might in turn lead to higher performance in children. Therefore, we predict that a compromise in the form of a slightly structured protocol (see "Procedure" section) will result in the highest performance levels for the child.

Finally, we also want to gain more insight in the verbal behavior of the adult using different protocols. Considering the scaffolding rules of contingency, fading, and transfer of responsibility, we expect that adults who adapt their behavior to the behavior of the child will yield higher performance levels in children. Also, if an adult is verbally more open (the child has more degrees of freedom in the reaction to the verbal utterance of the adult, e.g., an open question gives a child several options where an instruction tells the child what to do), a child might have more opportunities for exploration and hence show more insight in the task affordances and perform at a higher level.

General Method

Overview

Three separate studies were conducted in order to answer the research questions formulated above. In the first study, a comparison of children's performance on four different tasks was made to study the influence of the material context. The second study focused on the social context by comparing children's performance when adults used four different protocols. The final study further analyzed the social context by analyzing the verbal behaviors of the adults using these different protocols.

Materials

All children were interviewed about the underlying physical principles of four different tasks, namely a water tank task, a linked syringes task, a balance scale task, and a marble track task. Pictures of the four different tasks are shown in Fig. 2.

Water Tank The task consists of 14 objects and a water tank. The objects are eight daily used items (match, pebble, paperclip, plastic container, coin, pencil, plastic card, eraser) and six round items (marble, clay pebble, golf ball, metal ball, ping-pong ball, apple). Half of the objects will float on the water surface, whereas the other half will sink. The child was asked to predict and explain the floating or sinking of each object.

Linked Syringes In the linked syringes task, two syringes are connected with a single tube. If one of the syringes is pushed inward, the compression of air presses the other syringe outward. The opposite effect is reached by pulling on one of the syringes: The other syringe goes inward. The child was asked to predict and explain the events after the inward or outward movement of the piston.

Balance Scale The balance scale consists of a fixed fulcrum with ten different pins on each beam and 20 cards with an equal weight. Multiple cards can be hung to one pin and also at different distances. The child was asked to predict and explain the movement of the balance scale prior to and after adding one or more cards.

Marble Track The marble track is a wooden track in which a marble first climbs "up the stairs" and then rolls back down. A handle is connected to six eccentric wheels that push six increasingly long blocks up- and downward. The blocks have a slanted top that allows the

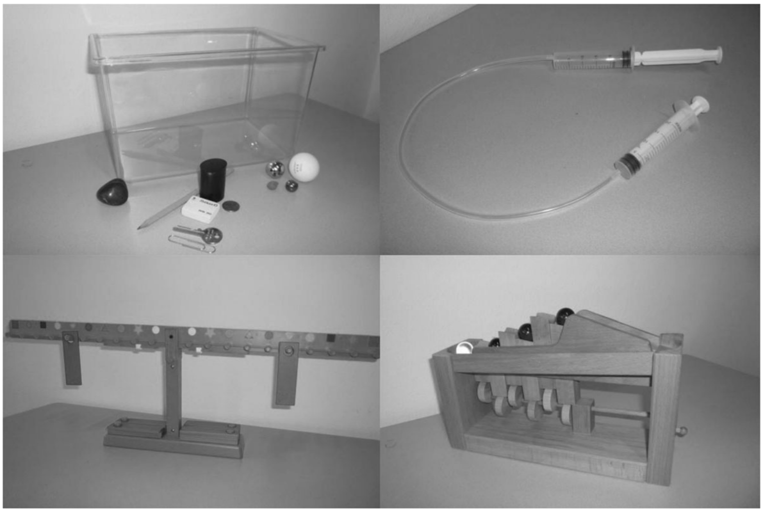

Fig. 2 The four tasks from *top left* to *right bottom*: water tank, linked syringes, balance scale, and marble track

marble to roll down onto the next block. By spinning the handle, the marble moves up these stairs and then rolls back through a steep, cornered slope to the beginning of the stairs. The blocks can be removed in order to let the child reassemble the stairs. The child was asked to predict and explain the working of the marble track.

Procedure

All children in three school classes (from two different elementary schools) received information letters about the research. Children whose parents returned the informed consent (N=35) participated in one of the three studies. Children from the elementary school with two participating classes were randomly selected for studies 1 and 2 or studies 2 and 3 (study 2 involves both samples of study 1 and study 3), whereas the children from the other school were selected for studies 1 and 2.

Each child was interviewed about all four tasks by an adult who was given one of four protocols. Each of the protocols had a general aim of gaining insight in the competence performance of the child. The interviews were not intended for teaching the child something about the task but instead on discovering the reasoning of the child about the task. The protocols were increasingly less structured for the adult: a test protocol, an adaptive protocol, a lesson plan protocol, and a free protocol situation. The protocols are thus rank ordered in accordance to the level of degrees of freedom permitted in the interaction to the adult. Each adult was given one of the protocols, with specifications for each separate task. In all protocols, the child was allowed to touch and/or use the task material. The order of tasks was randomly assigned, and the adults and children were unfamiliar to each other. The tasks were administered in a quiet room in their own elementary schools and recorded with two cameras.

Test Protocol In the test protocol, based on the procedure for administering the WISC-III-NL (Wechsler 2002), there was a standard set of questions. First, the child was asked to predict and explain what would happen. Then, the adult tested this prediction (e.g., place the object in the water tank) and asked the child why this had happened. The adult had to follow the protocol strictly, but was instructed to keep asking until they thought the child had reached his/her highest level of explanation. For each task, there was a fixed order of items and questions.

Adaptive Protocol In the adaptive protocol session, the same fixed order as in the test protocol was used, but the adult was free in the formulation of the questions and the amount of assistance provided. Additionally, a puppet was used to limit the social power distance between the child and the interviewer (Epstein et al. 2008; Nunkoosing 2005). The puppet was introduced as being a bit silly, and during the interview, the puppet questioned the child about his/her explanations until a satisfying answer was given.

Lesson Plan Protocol In the lesson plan protocol session, several educational aims of the task were stated; all focusing on letting the child discover the principles of the task. The adult was encouraged to ask for predictions and explanations. The lesson plan was based on those used in kindergartens and elementary schools in the Netherlands.

Free Protocol The adult receiving a free protocol was free to administer the task in his or her own way.

Although the adults received instructions about how and what to do, these were mostly content related and not necessarily form related. The adults were free in their use of open and closed questions, encouragements, and instructions during the task. Contrary to the test and adaptive protocol, adults in the lesson plan and free protocol condition were free to give feedback if they wanted to.

Coding

All task sessions were videotaped by the first author. The video-recordings were transcribed and then exported to Microsoft Excel where the coding of the verbal utterances of the adults and the children was done by the first and second author. All language of the adult was coded using the scheme presented in Table 1 with regard to the “Openness” of the utterances. The categories refer to the amount of freedom the child receives for responding. For example, giving a single instruction to the child limits the possibilities for the child more than asking an open question.

Of all the verbal utterances of the child, only predictions and explanations were coded. Codes for predictions were “correct,” “incorrect,” or “don’t know.” Explanations were coded according to the hierarchical complexity scale used in skill theory (Fischer and Bidell 2006; Fischer 1980; Meindertsma et al. 2012). The coding used by Rappolt-Schlichtmann et al. (2007) served as starting point. Table 2 summarizes the general coding scheme. In the extended coding book, this general scheme is adjusted for every task, specifying the specific combination of terms that belonged to each level (cf. Meindertsma et al. 2012). The same codes were used for both correct and incorrect explanations, indicating that complexity is considered to be independent of whether the explanations are correct or not. By definition, the lowest possible level for explanations was level 3 (sensorimotor system), since an explanation already contains a combination between what happened and a reason why this happened. Explanations that were given in a series of short utterances with short encouragements of the adult in-between were coded as one explanation. If the adult interrupts the explanation of child by saying anything else than an encouragement, the interrupted explanation was coded as two separate explanations.

Table 1 Coding scheme for the verbal utterances of the adult

Code	Utterance	Example
7	None	[Adults says nothing]
6	Encouragement	“Yes”
5	Open question	“What will happen next?”
4	Closed question	“Does something happen there?”
3	Giving information	“If I push, yours will go out and if I pull, yours will go in”
2	Instruction	“Pull yours out”
1	Stop	“Stop, stop!”

Inter-observer reliability was determined by comparing two trained coders independently coding four sessions, which resulted in an 80 % agreement on average for the adult categories (range 78–83 %). The mean kappa coefficient was 0.79, which indicates adequately reliable coders. Inter-rater reliability of the children’s categories was also good. For the child categories, eight sessions of study 2 were independently coded by two trained coders, resulting in an average of 87 % agreement on the type of utterance (prediction, explanation, or other; range 91–95 %) with a kappa coefficient of 0.92. The reliability of the correctness of predictions was even higher (94 % agreement on average; range 75–100 %) with an average kappa coefficient of 0.94. The mean kappa coefficient for the explanations was 0.78, with an agreement on average of 82 % (range 77–100 %).

Analyses

Three variables were calculated for the child and used in the analyses in all three studies: the percentage correct predictions, mean complexity level, and maximum complexity level. The other analyses are described in the method of the specific study.

Study 1

The first study was set up to answer the question whether different tasks elicit different levels of performance in the child. In order to do this, children were interviewed about four different tasks by one single adult.

Table 2 Coding scheme for the complexity level of children’s explanations

Code	Complexity level	Content of explanation	Example (water tank task)
6	Representational system	Combination of all relevant representation mappings	“The pencil is light for its size in the water
5	Representational mapping	Two or more parts of the explaining mechanism	“The pencil is light and thin”
4	Single representation	One part of the explaining mechanism	“The pencil is long”
3	Sensory motor system	Child states a relationship between action and result	“Because it is a pencil”

Method

Participants Twenty children (age M=65.4 months; SD=7.9) participated in the first study, 11 boys and 9 girls. The children were attending one of two elementary schools in the north of the Netherlands. The participating adults were all experienced in working with children, either in administering tests or teaching.

Procedure The children were interviewed about the four different tasks during one session with the same adult. The adult received one of four protocols. The protocols were equally distributed so that five children received all tasks in combination with the test protocol, five children all tasks with the adaptive protocol, and so on. The order of the tasks within one session was randomly assigned.

Analyses For the variable percentage correct predictions, missing data (the child did not formulate a prediction) were substituted with a zero, since there was no correct prediction. Missing explanations were not substituted considering there is no complexity level zero. A Monte Carlo analysis (e.g., Good 1999) was performed to test whether the performance on the tasks was different from what would be expected if all data came from one single distribution. For this purpose, the scores of the children on the four tasks were randomly shuffled (10,000 times) over the four tasks. If the simulated data deviated from the empirical data more than 95 % of the 10,000 runs, we can state that it is very unlikely that our empirical findings arise from one single distribution of the observations.

Results and Discussion

As shown in Fig. 3, the tasks resulted in differences in the performance levels. For all three performance measures, the water tank task yielded the highest performance in the children, independent of the protocol that was used. The mean percentage correct predictions for the water tank task was 55.07 (p=0.01), mean complexity level was 3.87 (SD=0.33, p<0.001), and maximum complexity level was on average 4.25 (SD=0.65, p<0.001). The children had most difficulty in predicting what would happen in the linked syringes task (M=30.92, p=0.02), whereas the marble track was more difficult for the variable mean complexity level (M=3.44, SD=0.38, p=0.001). The small differences found between the other tasks were not significant.

However, this higher level of performance on the water tank task does not imply that all children had their best performance at that task. Only eight of the 20 children reached their highest level of correct predictions on the water tank task. Also, just eight children were worst in predicting in the linked syringes task. There was more consistency in the other two variables, mean and maximum level of explanation, with 16 children showing their highest mean level on the water tank task and 17 the highest maximum level. Nevertheless, only in seven of the 20 children, the lowest mean complexity level occurred on the marble track task.

In summary, the results indicate that children showed clear differences in the performance level on the four tasks: The water tank task elicited the highest percentage of correct predictions, the highest mean complexity level, and the highest maximum level of complexity of explanations. This suggests that the affordances of the four tasks are different. It might be speculated, for instance, that the water tank task is quite straightforward in its purpose, whereas the marble track on the other hand is less clear (for instance, the handle is not directly visible from the front view, the marble is often presented separately by the adult, and

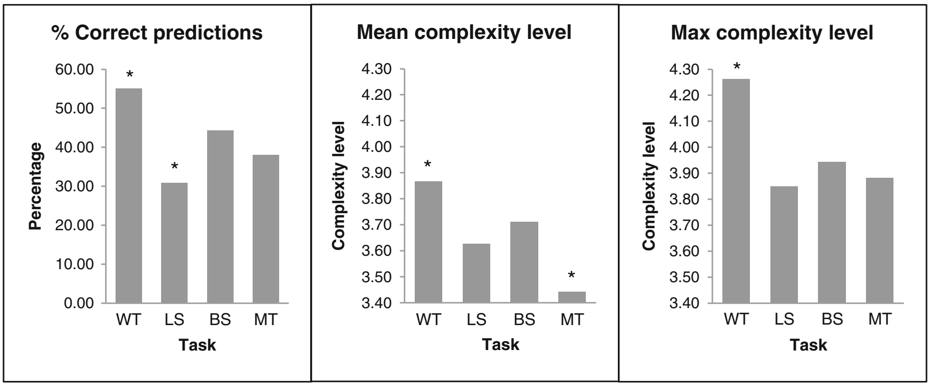

Fig. 3 Percentage correct predictions (*left*), mean complexity level (*middle*), and maximum complexity level (*right*) for the four different tasks. *WT* water tank, *LS* linked syringes, *BS* balance scale, *MT* marble track; *$p<0.05$

only when the handle is used is the staircase really visible). It also implies that familiarity with the affordances is an important factor, such as the notion of the developmental web suggests. Many children already have experience with concepts such as floating and sinking and balance in their everyday life (Callanan and Jipson 2001), for example bath ducks in the bathtub and a seesaw in the playground. The affordances of some of the tasks might already have been learned, whereas other affordances are totally unfamiliar.

On the other hand, novelty elicits explorative behavior (Henderson and Moore 1980), which in turn might make children learn the affordances of new tasks such as linked syringes and the marble track. Even if children might have explored more in these tasks, it did not result in higher performance levels. However, the aim of the interviews was not to teach children or to let them explicitly learn (although they might learn just by predicting and explaining), but to gain insight into the performance levels in different contexts. Familiarity might therefore be more important than novelty.

One other possible reason for the higher performance on the floating or sinking task might be in the structure of the task setting, more precisely the repetition of the problem: Each child was asked to predict and explain the floating or sinking of 14 different objects even in the free protocol situation. Through repetition, the child can build up his/her knowledge during the task.

Study 2

The second study investigates the influence of the social context. In order to do this, an additional group of children was interviewed by four different adults: This time, each was using a different protocol.

Method

Participants An additional group of 15 children (age M=62 months; SD=6.2) received all four protocols. None of these children had already participated in study 1. This time, each protocol was combined with a different task to avoid learning effects. The following combinations were used for the 15 children: (1) test protocol with water tank, (2) adaptive protocol with linked syringes, (3) lesson plan protocol with balance scale, and (4) free protocol with a marble track. Each protocol was given to a different adult, all unfamiliar with the participating children. Also, the results of the 20 children participating in the first study were used in the analysis.

Analyses First, the data of the 15 new children were analyzed to determine whether children performed better (or worse) in one of the four protocols. Again, the Monte Carlo testing technique was used to test the differences between the protocol conditions. In this particular condition, the data were randomly shuffled within each child separately. Since each protocol was combined with one task, we also corrected for task difficulty based on the results in study 1. Since the first study showed that the water tank and linked syringes task deviated on the variable percentage correct prediction, we statistically corrected for those tasks by multiplying the scores of the children by respectively 0.75 and 1.35 with a maximum of a hundred percent as score. These corrections were based on the empirically found ratios between the expected mean (when there would not have been differences between the four tasks) and the mean as found in study 1. The mean complexity level variable was also correct with a multiplication of 0.95 for the water tank and 1.05 for the marble track. For the maximum complexity level, only the score on the water tank was corrected by multiplying by 0.95.

Secondly, we performed a Monte Carlo analysis on the data from study 1 to check whether the differences found in the first step were also found over different tasks. The scores were also corrected for task difficulty. For the analysis, the scores of each child were shuffled over the different protocols

Because both samples consist of a small number of cases, we performed a meta-analysis to combine the results from the two analyses. This meta-analysis was executed in the following way. First, each time the simulated value for the $N=15$ group was higher than the observed value, a score of 1 was given. The same was done for the group children from study 1. After this, we calculated how many times of 10,000 runs the sum of these scores was 2 (both simulated values are higher than the observed values). If this occurred in less than 5 % of the runs, the meta-analysis indicates that the protocol indeed results in higher scores. The same procedure was performed if there was an indication that a protocol resulted in a lower score, but then the simulated value needed to be smaller than the observed value to receive a score of 1.

Results and Discussion

For the variable percentage of correct predictions, Fig. 4 shows that there were no differences between the four protocols for the new group. In contrast, the adaptive protocol resulted in the lowest percentage correct predictions in the group of children from study 1 ($M=29.48$, $p=0.02$), whereas the test protocol yielded the highest percentage ($M=51.68$, $p=0.02$). The meta-analysis revealed that, overall, the test protocol was responsible for the highest percentage correct predictions ($p=0.009$). The Monte Carlo analyses for the variable mean complexity level showed that the adaptive protocol was the lowest, but this was only statistically significant in the $N=15$ group ($M=3.50$, SD=0.56, $p=0.03$). The graph in Fig. 4 indicates that overall, the adaptive protocol might be the lowest, which is confirmed by the meta-analysis ($p=0.02$).

For the final variable, the maximum complexity level, the Monte Carlo analyses revealed no differences between the protocols. However, the test protocol seemed to be the highest if we visually inspect Fig. 4. Also, the p values for the test protocol were around 0.15 in both

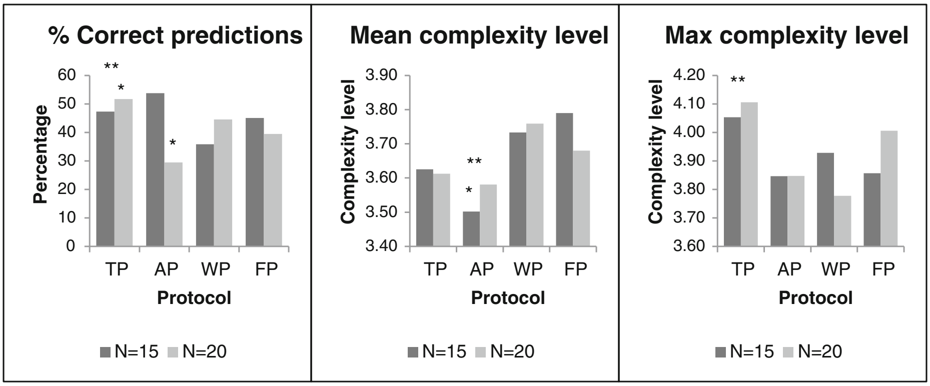

Fig. 4 Percentage correct predictions (*left*), mean complexity level (*middle*), and maximum complexity level (*right*) for the four different protocols with N=15 indicating the added group and the N=20 group indicating the group from study 1. *TP* test protocol, *AP* adaptive protocol; *LP* lesson plan protocol, *NP* no protocol; *p<0.05 in the Monte Carlo analysis, **p<0.05 in the meta-analysis

instances (N=15: M=4.05, SD=0.56, p=0.15; N=20: M=4.11, SD=0.76, p=0.16). Therefore, a meta-analysis was conducted which resulted in a p value of .02, indicating that the test protocol yielded the highest maximum complexity level.

In the "Introduction," we hypothesized that a certain degree of freedom in the task administration for the adult might lead to a better performance of the child. It allows the adult to better adapt his or her behavior to the level of the child, which in turn elicits a higher level of performance in the child. However, the percentage correct predications and the maximum complexity level were the highest in the test protocol condition. This suggests that—contrary to our expectation—a relatively high level of structure results in better maximum performance levels. This finding implies that structuring might be an important component in an assessment situation between child, task, and adult.

In a more structured setting for the adult, it might be that children are more restricted in their exploration and hence learn less about the affordances. Therefore, it is surprising that the test protocol resulted in the highest maximum complexity level in the children. On the other hand, the variable mean complexity level showed a different pattern with the lowest mean level in the adaptive protocol. However, since we did not analyze the amount of exploration in the different protocol conditions, we cannot support or reject this possibility. Ramsey and Fowler (2004) argued that asking questions can help children in their thinking. One important part that makes the test protocol seemingly better than the other protocols might be the repetition of asking questions which might have helped the children to gradually improve their insights (Meindertsma et al., submitted). The verbal behaviors of the adult are the scope of the final study.

Study 3

In the third study, we elaborate on the influence of the social context and in particular on the verbal actions of the adults using the protocols. The aim is to analyze the verbal behavior of the adults using the different protocols. First, the verbal behaviors of the adults per protocol are described, after which the adaptability of the adult's behavior is investigated, and finally, the associations between specific behaviors of the adult and the performance of the children are discussed.

Method

Participants Fifteen children (the same as in study 2; age *M*=62; SD=6.2) participated in this third study with each child interviewed by four different adults with each adult using a different protocol combined with a different task (see "Study 2").

Analyses The verbal behaviors of the adults using different protocols were statistically compared using a repeated measures ANOVA. The mean openness was calculated for the verbal behavior of the adult per child. To answer the question whether adults using different protocols differentially adapt their behavior to the child, the variability of the openness of the adult's verbal behavior was analyzed using a coefficient of variance (CV). This coefficient was calculated for each condition by dividing the standard deviation by the average. First, it was tested whether the CVs of the four conditions were the result of a single distribution. Next, the empirically found pair-wise differences between the CVs of the four conditions (e.g., CV of the adaptive protocol minus CV of the test protocol) were calculated. The ranking of the mean CV of the four protocols was calculated, and the probability that the empirically found ranking of variability among the four conditions were the result of a single underlying data distribution was tested using a Monte Carlo analysis.

Results and Discussion

The verbal behaviors of the adults using different protocols are presented in Fig. 5. The means and standard deviations are given in Table 3, as well as the significant differences. Differences were mostly found in the use of encouragements, giving information and instruction (Table 3). Contrary to our (intuitive) expectations, the most structured protocol resulted in the highest percentage of encouragement. The adult with the test protocol used "encouragement" the most, compared to the adaptive protocol, the lesson plan, and the free protocol. The use of encouragement also differed significantly between the free protocol condition and the lesson plan and the free protocol and adaptive protocol. Encouragement is the most open verbal behavior considering the amount of freedom the child has in his/her own actions. The test protocol did not instruct the adult to use encouragement, but instead stated that the adult should keep questioning the child until the adult believed the child's answer was at the highest possible level. This instruction was not explicitly stated in the other protocols, which might explain the lesser use of encouragement.

The adult in the test protocol also used more open questions than the adults using the adaptive protocol (F (1, 14)=16.02, $p<0.01$), lesson plan protocol (F(1, 14)=7.88, $p=0.01$), and free protocol (F (1, 14)=27.61, $p<0.01$). The smallest percentage of closed questions was asked in the adaptive protocol compared to the test protocol (F (1, 14)=7.95, $p<0.01$) and free protocol (F (1, 14)=7.72, $p<0.01$).

The adult given the test protocol stopped the child less often than in the adaptive protocol (F (1, 14)=10.31, $p=0.01$) and the free protocol (F (1, 14)=5.96, $p=0.03$). There was also a difference between the adaptive protocol and the lesson plan protocol (F (1, 14)=5.14, $p=0.04$).

The frequency of giving information was higher in the less structured protocols: It occurred most frequently in the lesson plan condition compared to the test protocol condition, the adaptive protocol, and the free protocol, with test protocol using it least frequently and less than in the adaptive protocol and the free protocol condition. This giving information might be a form of scaffolding to elicit higher levels of reasoning. However, in the

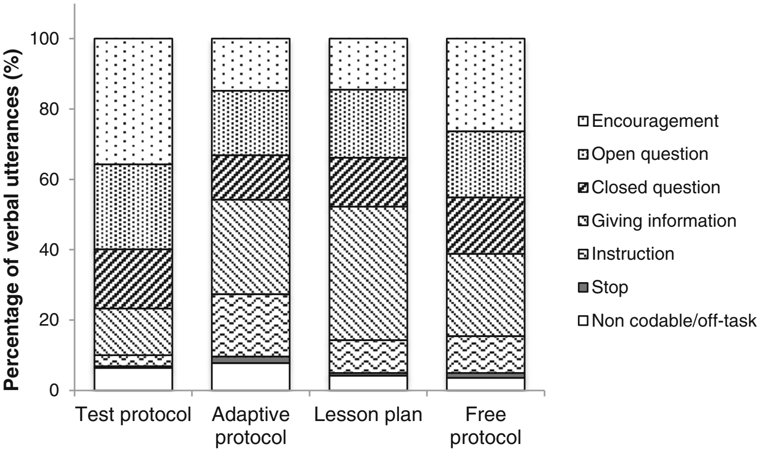

Fig. 5 Percentages of used verbal behaviors by adults in the different protocol conditions

second study, no evidence was found that the less structured protocols (with more degrees of freedom for the adult) resulted in higher levels of scientific reasoning in the child. In order to test whether the adults adapted their behavior to the specific child, the variability of the verbal behaviors was analyzed.

Instructing was least frequent in the test protocol condition compared to the lesson plan, adaptive protocol, and the free protocol. In the adaptive protocol, the percentage of giving instruction was the highest compared to lesson plan and to the free protocol.

The openness over time—for each protocol separately—is displayed in Fig. 6. For each interview session, the mean openness of the adult's verbal behavior is displayed. Visual inspection of Fig. 6 shows that some protocols have more variability in the openness over sessions than others. When the coefficient of variability is calculated, the test protocol is the least variable (CV=0.025), whereas the variability of the adaptive protocol (CV=0.039), lesson plan protocol (CV=0.05), and free protocol (CV=0.12) is higher. Only the variability in the free protocol condition proved to be higher than in the other conditions ($p<0.001$). However, if we compared the ranking of the conditions from the highest variability to the lowest variability (that is: free protocol–lesson plan–adaptive protocol–test protocol) with simulated rankings, the analysis showed that the observed ranking was significantly different than expected ($p=0.04$).

The stability of the behavior across sessions with different children confirmed the expectations (and perhaps even the requirement) that the more structured the protocol, the less variable the behavior of the adults. However, the results of study 2 demonstrated that, on average, the most variable protocol was not yielding the highest performance level. In order to explore which behaviors of the adult influenced the performance of children, correlation coefficients (Spearman's rho) between the verbal behaviors of the adults and the children's performance were calculated.

In Table 4, the statistically significant correlations are presented. The table shows that overall, *only* the percentage of encouragement used by the adult correlated with the maximum complexity level of children. Within the test protocol, positive correlations were also found between mean openness and maximum complexity level. Within the adaptive protocol, positive correlations were found between encouragement and mean and maximum complexity level. Stopping the child verbally by the adult using the adaptive protocol also correlated positively with the percentage correct predictions.

Table 3 Mean percentage and standard deviation per verbal behavior for the four different protocols with the significant *p* values between the different protocols

		M	SD	TP *F* (*p*)[a]	AP *F* (*p*)[a]	LP *F* (*p*)[a]
Encouragement	Test protocol	35.71	4.60			
	Adaptive protocol	14.77	3.75	226.58 (<0.001)		
	Lesson plan	14.46	5.24	234.77 (<0.001)	ns	
	Free protocol	26.36	6.33	29.32 (<0.001)	42.83 (<0.001)	103.79 (<0.001)
Open question	Test protocol	24.17	4.01			
	Adaptive protocol	18.35	4.77	16.02 (0.001)		
	Lesson plan	19.36	5.26	7.88 (0.01)	ns	
	Free protocol	18.78	3.17	27.61 (<0.001)	ns	ns
Closed question	Test protocol	16.83	3.85			
	Adaptive protocol	12.63	4.30	7.95 (0.01)		
	Lesson plan	13.95	5.05	ns	ns	
	Free protocol	16.08	2.50	ns	7.72 (0.02)	ns
Giving information	Test protocol	13.32	2.00			
	Adaptive protocol	26.92	6.27	50.27 (<0.001)		
	Lesson plan	38.00	6.68	164.73 (<0.001)	19.08 (0.001)	
	Free protocol	23.40	2.63	109.63 (<0.001)	ns	49.69 (<0.001)
Instruction	Test protocol	3.11	2.14			
	Adaptive protocol	17.72	3.71	168.15 (<0.001)		
	Lesson plan	9.30	3.83	48.31 (<0.001)	31.15 (<0.001)	
	Free protocol	10.41	4.63	35.02 (<0.001)	31.91 (<0.001)	ns
Stop	Test protocol	0.45	0.52			
	Adaptive protocol	1.86	1.79	ns		
	Lesson plan	0.74	1.07	ns	5.14 (0.04)	
	Free protocol	1.40	1.64	5.96 (0.03)	ns	ns

[a] *df*=1, 14

The use of questions, both open and closed, correlated negatively with some of the variables in the test and adaptive protocol. The frequency of open questions correlated negatively with mean complexity level in the test protocol condition and with all performance levels in the adaptive protocol condition. The amount of closed questions also showed a negative relationship with mean complexity level in the test protocol as well as with percentage correct predictions in both the test and the adaptive protocol.

No significant correlations were found in the least structured conditions: the lesson plan and the free protocol condition. The reported correlations mainly suggest that encouragement in the more structured protocols is positively correlated whereas asking questions is negatively related. The amounts of questions were negatively correlated with performance levels of children whereas encouragement had a positive correlation. The hypothesis of Ramsey and Fowler (2004) that asking questions can promote children's thinking was not supported in our study. The more questions the adults asked in the more structured settings, the lower the performance levels of the children.

In summary, main differences in verbal behavior were found in the use of encouragements, giving information and instruction. Of these three behaviors, only the use of

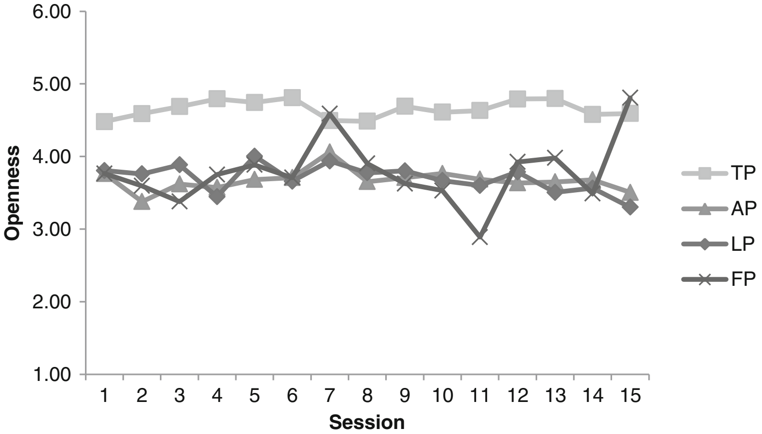

Fig. 6 Average openness scores over time for each of the conditions. The timeline represents the sequence of children, which means that the first session for the test protocol might be with another child than the first session for the adaptive protocol

encouragement correlated highly with some aspects of the performance of the children. Although the adult using the free protocol condition was most variable, suggesting adaptation to the child, this did not result in the highest level of performance in the children.

The description of the results tacitly assumes that the adult's asking is an independent variable, resulting in a particular effect on the dependent variable, which is the child's level of performance. However, as argued in the "Introduction," there is a constant interaction between participants and task elements. In this ongoing interactive cognitive construction process, an adult's question may trigger a particular level of cognitive explanation in the child, but a particular level of cognitive explanation in the child may also trigger a consecutive, particular type of question in the adult. For instance, it is likely that during the interaction process, repeated low levels of explanation in the child may lead to more questioning in the adult, especially under conditions in which the adult is instructed to attempt to obtain the best possible view on a particular child's performance level, which adults might interpret as the highest possible performance level under the given

Table 4 Significant Spearman correlations between adult's verbal behavior (percentage of used behavior) and children's performance

Protocol	Verbal behavior	% Correct predictions	Mean complexity level	Maximum complexity level
Overall	Encouragement			0.32*
Test protocol	Mean openness			0.60*
	Open question		−0.54*	
	Closed question	−0.67**	−0.68**	
Adaptive protocol	Encouragement		0.57*	0.61*
	Open question	−0.60*	−0.58*	−0.58*
	Closed question	−0.63*		
	Stop	0.71*		

*$p<0.05$; **$p<0.01$

circumstances. This tacit interpretation may encourage adults to ask more questions in particular if the child's level tends to be repetitively low.

The transactional relationship between the child and the adult is likely to be true also for the transactional relationship between the child (or the adult) and the materials offered. Children with higher levels of understanding may show a higher level of manual exploration of the materials, dependent on the actual affordances these materials offer, and high levels of manual exploration on their turn may lead to higher levels of verbal understanding in the child, or may lead to more or more specific questions in the adult. This particular interpretation is in line with the notion of the dynamic talent triangle discussed in the theoretical introduction.

General Discussion

The three studies have provided us with a greater insight in the role of the material and social context on the child's level of understanding. The first study demonstrated that even within one domain such as science, preschoolers' performances are variable between tasks. In our study, a floating and sinking task elicited the highest level of performance for all three variables (percentage correct predictions, mean, and maximum complexity level of explanation). This was probably the most well-known task to the children and has a highly repetitive character. In addition, the second study, focusing on the social context, indicated that a test protocol was best in eliciting the highest maximum complexity level as well as the highest percentage correct predictions whereas the adaptive protocol resulted in the lowest mean complexity level. The behavior of the adults in the third study showed differences in the verbal behavior of the adults when using different protocols. There are also differences in the stability of the adult's behavior with different children, but this adaptation does not necessarily result in better performance levels of the child. The frequency of questions was negatively correlated with higher levels of performance, whereas encouragement and more open behavior in general related positively to the performance of children in the more structured conditions.

The aim of this paper was to assess the influence of the material and social context on the child's level of understanding. The results are in line with previous research which emphasizes the influence of the context (e.g., Fischer et al. 1993; Kloos et al. 2010; Smith and Thelen 2003). Our findings strengthen the idea of the presented triangle: Adult, task, and child create their own context, and a change in one of the corners can affect the whole system. In this assessment situation, it may be speculated that the reason why the standard testing protocol results in such high levels is that it challenges the child more explicitly than the other protocols do, also by encouraging the child, which for the child might be an additional prompt to try to increase their level of functioning during the interaction process. Since the adult's degrees of freedom in the interaction are virtually limited to repeated questioning and using encouragements in the test protocol situation, it is easy to see why the interaction process that emerges from it is one in which the child as well as the adult are highly challenge-oriented, resulting in high levels of performance (as long as these higher levels are within reach of the children).

The high degree of structure cannot only be a form of task scaffolding but is also related to social scaffolding. The adult using the test protocol constantly repeats the same order of questions: prediction, explanation, observation, and explanation. The test protocol resulted in the highest maximum complexity level for the explanations. We suspect that this repetition can be regarded as a scaffold for children's performance, although previous research mostly focused on direct support instead of task structuring (Hsin and Wu 2011; van Geert and Steenbeek 2005). Although it is not directly part of the key characteristics of scaffolding (Van de Pol et al. 2010), one could link it to transfer of responsibility. By taking over the responsibility of structuring the task, the child reaches a higher maximum level.

Following another key characteristic of social scaffolding, contingency, adaptation to the preschooler's level was expected to lead to higher levels of performance. The results indicated otherwise: In our study, the least adaptive protocol showed the highest maximum levels of reasoning. Scaffolding is predominantly studied in the field of learning processes. The aim of the interviews here was to assess the highest possible level, which is necessary in the process of scaffolding to determine the needed scaffolding level (Van de Pol et al. 2010). This means that in assessment situations, adaptiveness of the adult may not be the key factor for eliciting higher levels of reasoning. Instead, a child might perform better in more structured conditions.

Even within the same task domain (science), children can perform on different levels, as predicted by dynamic skill theory (Fischer and Bidell 2006). This can be partly due to the children's familiarity with the affordances of the tasks (Gibson 1988). In a way, this familiarity can act as a task scaffold, comparable to the way task structure does. For instance, the floating/sinking task might be a more predictable task for the child (for each object, regardless of the protocol, the adult wants to know if it floats or sinks) than the marble track (questions used by adults varied from "what will happen" to "can you let the marble go the other way").

Limitations and Future Research

It might be argued that the robust difference between protocols is not due to the protocol influencing the performance levels in the children, but is a reflection of the individual differences of the participating adults using the protocol. We agree that it may not be possible to unravel the personal style of an adult from the prescribed protocol. However, we consider the actual behavior of the adult and its influence on the child's performance as the central aspect that matters. Here we have demonstrated that a highly standardized approach with many encouragements and repeated questioning was associated with a higher performance of the child.

This study focused on the verbal behaviors of both adult and child. In the present study, nonverbal behavior has not been measured. However, nonverbal behavior of the child is also important, since Pine et al. (2004) and Goldin-Meadow et al. (1993) showed that in a particular task, children's nonverbal performances in the form of gestures might represent the higher cognitive level than their verbal performances.

Further, we have to take into consideration to what extent these results are generalizable to other ages. The participants were all preschoolers, ranging in age from 4 to 6 years. Since we state that the context is created by all participants, including the child, different child characteristics, such as higher reasoning skills, can lead to different influences of the context. Similarly, different adults might influence the context differently, even when using the more structured protocols.

Future research has to study whether older children, who are capable of reasoning at a higher level (Fischer 1980; Siegler and Alibali 2005), need the same level of task structuring or react the same to the different protocols. Also, further research needs to focus on the temporal unfolding of the cognitive interaction processes in individual adult–child dyads, in addition to analyzing the non-verbal behaviors of the adult and the child from a cognitive level. This could inform us, for instance, whether encouragements by the adult precedes higher performance or vice versa, or both. Finally, inter- and intra-individual variability was not studied here, whereas we know that there can be large differences between and within children (e.g., Van Dijk and Van Geert 2007).

Implications and Conclusion

In many countries, science education is an upcoming field of interest, especially in early childhood. Our results contribute to the knowledge about teaching science to young children. First, teachers are part of the social context, and their behavior is part of the interaction process that leads to a child's performance. In the long-term, there are different developmental pathways children can follow while learning about science and technology. Various protocols can be helpful in this long-term learning process. However, when *assessing* a child's *optimal* level (which is a necessary step if teachers want to make use of scaffolding in the learning process), a more strict protocol that challenges the child and includes a lot of repetition and encouragement can be the best choice. Thus, teachers have to consider whether they are using the right structuring and verbal behavior to ensure assessing a child's optimal reasoning level.

Teachers also have to be aware of the influence of the material context. A child's previous familiarity with the affordances of a task seems to play a large role in a child's performance on the short term. This naturally also affects long-term development. In our view, getting children acquainted with a large range of tasks should be an important goal in science education.

The presented findings also have implications for researchers. First, researchers have to consider the social and material context in their research design just as the teachers need to do in the classroom setting. If they want to assess a child's optimal level, they too can best use a test protocol. Also, in their choice of a suitable task, a child's familiarity with the affordances needs to be kept in mind. Second, the decision for outcome measures has to be taken carefully. In our case, skill theory proved to be a good measure for the comparison of performance in different contexts, but effects differed even for closely related variables such as mean and maximum complexity level. The maximum level in our study might be related to the optimal level of performance as proposed by Fischer (1980) and Fischer and Bidell (2006), whereas the functional level is more closely approximated by the mean level. That is, there is a need for further validation of what is actually measured in studies of learning.

Concluding, these three studies showed that relatively small changes in context can result in changes in a child's performance level which is in accordance with the idea of the scientific reasoning triangle: The three participants (adult, child, and task) together create the context for performance. Task performance has to be considered as a process that unfolds in real time, in a way both facilitated and constrained by the properties of all elements involved, namely the child, the adult, and the task context. Protocols can be interpreted as particular constraints on the social and cognitive affordances provided by the adult, leading to different types of unfolding of the real-time interaction process. This has implications for both researchers and educators (in all fields of education) who have to be constantly aware of the influence of their choices in the creation of the context.

References

Callanan, M. A., & Jipson, J. L. (2001). In K. Crowley, C. D. Schunn, & T. Okada (Eds.), *Explanatory conversations and young children's developing scientific literacy* (pp. 21–49). Mahwah: Lawrence Erlbaum Associates.

Chak, A. (2010). Adult response to children's exploratory behaviours: an exploratory study. *Early Child Development and Care, 180*(5), 633–646. doi:10.1080/03004430802181965.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2), 181–195. doi:10.1207/S15326969ECO1502_5.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). In R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.), *Taking science to school: learning and teaching science in grades K–8*. Washington, DC: National Academies.

Epstein, I., Stevens, B., McKeever, P., Baruchel, S., & Jones, H. (2008). Using puppetry to elicit children's talk for research. *Nursing Inquiry, 15*(1), 49–56. doi:10.1111/j.1440-1800.2008.00395.x.

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

Fischer, K. W. (1980). A theory of cognitive development: the control and construction of hierarchies of skills. *Psychological Review, 87*(6), 477–531. doi:10.1037/0033-295X.87.6.477.

Fischer, K. W., & Bidell, T. R. (2006). In R. M. Lerner & W. Damon (Eds.), *Dynamic development of action and thought* (pp. 313–399). Hoboken: Wiley.

Fischer, K. W., Bullock, D., Rotenberg, E. J., & Raya, P. (1993). The dynamics of competence: How context contributes directly to skill. In R. H. Wozniak & K. W. Fischer (Eds.), *Development in context: acting and thinking in specific environments, The Jean Piaget symposium series* (pp. 93–117). Hillsdale: Lawrence Erlbaum Associates.

French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly, 19*(1), 138–149. doi:10.1016/j.ecresq.2004.01.004.

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

Gibson, J. J. (1979). *The ecological approach to visual perception* (p. 332). Boston: Houghton, Mifflin and Company.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology, 39*(1), 1–42. doi:10.1146/annurev.ps.39.020188.000245.

Good, P.I. (1999). Resampling methods: a practical guide to data analysis. Boston: Birkhauser.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: using the hand to read the mind. *Psychological Review, 100*(2), 279–297. doi:10.1037//0033-295X.100.2.279.

Granott, N., Fischer, K. W., & Parziale, J. (2002). Bridging to the unknown: a transition mechanism in learning and development. In N. Grannot & J. Parziale (Eds.), *Microdevelopment: transition processes in development and learning* (pp. 131–156). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511489709.006.

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "Floating and sinking". *Journal of Educational Psychology, 98*(2), 307–326. doi:10.1037/0022-0663.98.2.307.

Havu-Nuutinen, S. (2005). Examining young children's conceptual change process in floating and sinking from a social constructivist perspective. *International Journal of Science Education, 27*(3), 259–279. doi:10.1080/0950069042000243736.

Henderson, B. (1984). Parents and exploration: the effect of context on individual differences in exploratory behavior. *Child Development, 55*(4), 1237. doi:10.2307/1129993.

Henderson, B., & Moore, S. G. (1980). Children's responses to objects differing in novelty in relation to level of curiosity and adult behavior. *Child Development, 51*(2), 457. doi:10.2307/1129279.

Hoogsteder, M., Maier, R., & Elbers, E. (1996). The architecture of adult–child interaction. Joint problem solving and the structure of cooperation. *Learning and Instruction, 6*(4), 345–358. doi:10.1016/S0959-4752(96)00020-5.

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

Kloos, H., Fisher, A., & Van Orden, G. C. (2010). Situated naïve physics: task constraints decide what children know about density. *Journal of Experimental Psychology. General, 139*(4), 625–637. doi:10.1037/a0020977.

Lave, J. (1988). *Cognition in practice: mind, mathematics, and culture in everyday life (Vol. 3)*. Cambridge: Cambridge University Press.

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2008). Young children's motivational beliefs about learning science. *Early Childhood Research Quarterly, 23*(3), 378–394. doi:10.1016/j.ecresq.2008.04.001.

Mascolo, M. (2005). Change processes in development: the concept of coactive scaffolding. *New Ideas in Psychology, 23*(3), 185–196. doi:10.1016/j.newideapsych.2006.05.002.

Mascolo, M., & Fischer, K. W. (2010). The dynamic development of thinking, feeling and acting over the lifespan. In W. F. Overton (Ed.), *Biology, cognition and methods across the life-span. Volume 1 of the Handbook of life-span development (vol. 45)*. Hoboken: Wiley.

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.

Nunkoosing, K. (2005). The problems with interviews. *Qualitative Health Research, 15*(5), 698–706. doi:10.1177/1049732304273903.

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*(1), 95–117. doi:10.1002/icd.504.

Piaget, J. (1930). *The child's conception of physical causality*. New York: Harcourt, Brace & Company.

Pine, K. J., Lufkin, N., & Messer, D. (2004). More gestures than answers: children learning about balance. *Developmental Psychology, 40*(6), 1059–1067. Retrieved from http://www.apa.org/journals;http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ684676&loginpage=Login.asp&site=ehost-live&scope=site.

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

Ramsey, J., & Fowler, M. (2004). "What do you notice?" Using posters containing questions and general instructions to guide preschoolers' science and mathematics learning. *Early Child Development and Care, 174*(1), 31–45. doi:10.1080/0300443032000153471.

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*(2), 98–108. doi:10.1111/j.1751-228X.2007.00010.x.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science education now: a renewed pedagogy for the future of Europe (EUR 22845). Economy and Society.* Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf.

Ruff, H. A. (1984). Infants' manipulative exploration of objects: effects of age and object characteristics. *Developmental Psychology, 20*(1), 9–20. doi:10.1037/0012-1649.20.1.9.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84. doi:10.1002/tea.20163.

Salonen, P., Lepola, J., & Vauras, M. (2007). Scaffolding interaction in parent–child dyads: multimodal analysis of parental scaffolding with task and non-task oriented children. *European Journal of Psychology of Education, 22*(1), 77–96. doi:10.1007/BF03173690.

Schwartz, M., & Fischer, K. W. (2004). Building general knowledge and skill: cognition and microdevelopment in science learning. In A. Demetriou & A. Raftopoulos (Eds.), *Cognitive developmental change. Theories, models and measurement* (pp. 157–185). New York: Cambridge University Press.

Shaw, R. E., Turvey, M. T., & Mace, W. M. (1982). Ecological psychology: the consequence of a commitment to realism. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes* (2nd ed., Vol. 2, pp. 159–226). Hillsdale: Lawrence Erlbaum Associates.

Siegler, R. S., & Alibali, M. W. (2005). *Children's thinking* (4th ed.). Upper Saddle River: Pearson Prentice Hall.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: a microgenetic analysis. *Cognitive Psychology, 36*(3), 273–310. doi:10.1006/cogp.1998.0686.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*(8), 343–348. doi:10.1016/S1364-6613(03)00156-6.

Steenbeek, H. W., Van Geert, P. L. C., & Van Dijk, M. W. G. (2011). The dynamics of children's science and technology talents: a conceptual framework for early science education. *Netherlands Journal of Psychology, 66*(3), 96–109.

Thelen, E. (1992). Development as a dynamic system. *Current Directions in Psychological Science, 1*(6), 189–193. doi:10.1111/1467-8721.ep10770402.

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

Tytler, R., & Peterson, S. (2003). Tracing young children's scientific reasoning. *Research in Science Education, 33*(4), 433–465. doi:10.1023/B:RISE.0000005250.04426.67.

Tytler, R., & Peterson, S. (2004). From try it and see to strategic exploration: characterizing young children's scientific reasoning. *Journal of Research in Science Teaching, 41*(1), 94–118. doi:10.1002/tea.10126.

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

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*(1), 7–33. doi:10.1002/icd.506.

Van Geert, P. L. C., & Steenbeek, H. W. (2005). The dynamics of scaffolding. *New Ideas in Psychology, 23*(3), 115–128. doi:10.1016/j.newideapsych.2006.05.003.

Wechsler, D. (2002). *Wechsler intelligence scale for children* (3rd ed.). Lisse: Harcourt Test. Nederlandse Versie.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. doi:10.1111/j.1469-7610.1976.tb00381.x.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*(1), 99–149. doi:10.1006/drev.1999.0497.