Science learning in informal contexts

Behavioral studies on children's knowledge of natural phenomena and family learning in the museum

Franse, R.K.

Publication date
2021

Document Version
Other version

License
Other

3

Families' Manipulations and Conversations at an Open-ended Exhibit in a Science Museum: Individual Characteristics and the Influence of Minimal Guidance Strategies

This chapter is based on: Franse, R.K., Van Schijndel, T.J.P., Plankman, T. I., & Raijmakers, M.E.J. (2020). *Families' Manipulations and Conversations at an Open-ended Exhibit in a Science Museum: Individual Characteristics and the Influence of Minimal Guidance Strategies.* Manuscript under review.

ABSTRACT

A frequently mentioned consequence of open-ended investigation, in formal and informal science learning contexts, is that it reduces learning effectiveness. In a formal school context verbal scaffolds have been shown to significantly contribute to the effectiveness of learning through investigation. The current in-depth study examines the impact of verbal scaffolds and individual characteristics on family learning at an open-ended museum exhibit. The families' manipulations and conversations were observed at an object motion exhibit. Results show that all families (N = 104), with and without guidance, investigated in a meaningful way by performing control-of-variables strategy experiments, investigating a range of variables, and formulating hypotheses and causal explanations. However, the results also show that the process of learning scientific concepts could be improved. Minimal interventions of museum educators positively affected the families' learning process by reducing the number of scientifically incorrect remarks. Interestingly, in addition to discussing the phenomenon, especially the families with highly educated parents discussed the topic of reliability of their experiments, which is an under-investigated aspect of learning through investigation. Only the children's cognitive abilities modestly impacted the families' performance consistently, which implies that not only the families with highly educated parents, but all participating families experienced moments of high-quality open-ended investigation.

Key words: Family learning; Conversation; Museum educator; Control-of-variables strategy; Visitor studies.

3.1 INTRODUCTION

Modern science museums are venues where both children and adults engage with different aspects of science and technology in an informal way (National Research Council, 2009). Aimed at offering visitors meaningful experiences, science museums are recognized as places for learning (Barriault & Pearson, 2010), where visiting families have the intention and expectation to learn 'something' (Falk & Storksdieck, 2010). Interactive exhibits, fostering investigation (inquiry) of natural or physical phenomena, have always taken a central place in science museums (Gutwill & Dancstep, 2017). Interactive exhibits encourage visitors to discover and explore (McLean, 1993; Rennie & McClafferty, 1996), and elicit social interaction between child and parent (Blud, 1990). Discovery exhibits (Hein, 1998) are interactive exhibits that explain and offer scientific content leading the visitor to canonical correct scientific ideas (Gutwill, 2008). More open-ended interactive exhibits invite visitors to design and conduct experiments that might reveal scientific phenomena (e.g., the Investigate! exhibition, Museum of Science, Boston). APE (Active Prolonged Engagement) exhibits in particular invite visitors to do so (Humphrey & Gutwill, 2005). These exhibits have multiple options, give families the opportunity to actively create experiments themselves, and therefore support the investigation process.

From a constructivist point of view (Gopnik, 1996; Inhelder & Piaget, 1958), open-ended interactive exhibits offer families the opportunity to learn through investigation or to actively construct knowledge by investigating their own questions. These exhibits bear similarities to exhibits matching real-world phenomena (Gilbert & Stocklmayer, 2001). Discovery exhibits, in contrast, tend to place visitors in a role of passive recipient of information. These exhibits bear similarities to exhibits providing simple demonstrations of phenomena (Gilbert & Stocklmayer, 2001). Educational effect studies (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004), show that open-ended exploration activities often lead to prolonged investigation, but also to situations in which learners experience difficulties with observing and interpreting observations. One could argue that open-ended investigation occurs at the expense of quality, and therefore hampers knowledge and skill acquisition (Gutwill, 2008). A similar, but opposite example of this trade-off between visitors' active involvement and learning is reported in exhibit design. Aiming to develop exhibits that supported both the investigation processes and the acquisition of scientific content, Gutwill (2008) studied visitors' interaction with counterintuitive exhibits. Gutwill concluded that observing an unexpected effect made visitors curious, and motivated them to better understand how the phenomenon worked, but it did not encourage them to engage in further investigation.

The above examples raise the question whether there are ways to achieve both goals at open-ended exhibits. That is, preserving the open-ended character of an exhibit and therefore leaving enough room for families' own initiatives during an investigation and, at the same time, offer sufficient support to make this investigation an opportunity for learning. In a formal school

context, guidance has been shown to significantly contribute to the effectiveness of learning through investigation (Alfieri et al., 2011; Dobber, Zwart, Tanis & Van Oers, 2017; Lazonder & Harmsen, 2016). Through meta-analyses, Alfieri et al. demonstrated that guided inquiry renders better learning outcomes than other forms of instruction, and Lazonder and Harmsen added that guidance also positively impacts the actions learners perform and the quality of products they create during inquiry. Dobber et al. used a systematic review to provide insight in the specific guidance strategies K-12 teachers use, and distinguished strategies in terms of their directions (from teacher to student) and perspectives of regulation (meta-cognitive, conceptual and social regulation). For museums, research on the effectiveness of these guidance strategies is informative. However, the informal context differs fundamentally from the formal context. For example, learner autonomy and motivation, individual differences between learners, of assessment, social context, learning goals, and learning materials are all different (Callanan, Cervantes, & Loomis, 2011; Falk & Storksdieck, 2005; Hooper-Greenhill & Moussouri, 2000). These differences can affect the suitability of a specific guidance strategy (e.g., adapting interventions to the child's prior knowledge) for the museum context. For example, museum educators meet large amounts of new people every day, while teachers can build on a long term relation with their students which makes it easier to assess the student's prior knowledge. Another aspect to take into account when considering implementing a guidance strategy from the school context to the museum context: learning depends on context. This means that it cannot simply be assumed that an effective guidance strategy in the formal context also impacts learning in the informal context.

In this study, we investigate whether minimal guidance by museum educators can positively contribute to families' learning through investigation at an open-ended exhibit. This way, we aim to generate knowledge on ways to achieve both goals, skill acquisition and concept learning, at open-ended museum exhibits. Learning through investigating in a museum has many facets. Hence, before discussing museum guidance strategies in more detail, we will first introduce the theoretical perspective we have chosen in the current study, and discuss previous literature on learning through investigation at interactive exhibits in a science museum.

3.1.1 Studying Learning through Investigation in Science Museums

In previous work on learning through investigation in a museum, two main perspectives have been distinguished: the cognitive developmental perspective (Callanan & Valle 2008; Gopnik, 1996; Piaget, 1936; Wellman and Gelman, 1992), and the socio-cultural perspective (Bronfenbrenner, 1977; Callanan, Castañeda, Luce, & Martin, 2017; Dritsas, Borun & Johnson, 1998; Rogoff, 2003; Vygotsky, 1978). Several studies take a dual perspective on learning, some have a stronger emphasis on the individual child, while others are more focussed on family learning (e.g., Benjamin, Haden, & Wilkerson, 2010; Gutwill & Allen, 2010). In this study we also choose to combine these perspectives on learning. We aim to take into account both learning in natural situations where all family members contribute to a shared learning situation,

as well as that learning is also an individual effort whereby information from the same shared learning situation can be perceived and processed differently. Additionally, the combined perspective recognizes parents' dual role as a facilitator of the child's learning process and as a learner along with their child.

Examples of how families' learning through investigation is measured: manipulations, conversation, and post-activity learning outcomes. To illustrate the versatility of learning through investigation, we organize the literature by different types of indicators: manipulation, conversation, and post-activity learning outcomes.

In an open-ended exhibit multiple variables can be manipulated to investigate a phenomenon or principle (e.g., moving pictures, shadow size, bracing or gears). A first indicator of learning through investigation is the extent to which visitors' manipulations are informative, i.e. the extent to which they provide relevant and reliable information that allows them to learn about the phenomenon (e.g., Crowley et al., 2001a; Van Schijndel, Franse, & Raijmakers, 2010; Willard et al., 2019). For example, Crowley et al. (2001a) investigated whether children performed informative experiments at a zoetrope exhibit by spinning the zoetrope while looking through a slot. Alternatively, in a similar study by Van Schijndel et al. (2010) high-quality, possibly informative investigations by children were described as: "A child manipulates an object in an active and attentive manner …. [and] … applies repetition and variation to his or her actions" (p. 799). In contrast to measuring the manipulations of individual children, in other studies families' actions were quantified. Pattison et al. (2018), for example, investigated families' mathematical reasoning by coding their' behaviors related to the exploration of mathematical relationships and achievement of mathematical goals at different exhibits. Yet another way to measure families' behavior, besides observing in situ, is by assessing if the created product reflects understanding of a phenomenon or principle. For example, in a learning by design activity, Benjamin et al., (2010) quantified the number of materials used and the triangles and struts created, as a reflection of families' understanding of the engineering principle of bracing.

A second indicator of learning through investigation is the verbal interaction between family members. Different aspects of families' conversations have been studied. A first aspect is the reasoning *content*, quantified as naming properties of the scientific phenomenon at hand (e.g., Leinhardt, Crowley, & Knutson, 2002; Palmquist & Crowley, 2007; Pattison et al., 2018) or sources of content (e.g., Tunnicliffe, 2000; Zimmerman, Reeves, & Bell, 2010). For example, Zimmerman et al., focussing on family learning, investigated the types of epistemic resources (e.g., own knowledge or observations of scientific phenomena) families used during investigation in a museum context. A second aspect concerns the reasoning *skills,* quantified as elements of scientific reasoning, such as formulating hypotheses or interpreting results (Gutwill & Allen, 2010, Kisiel, Rowe, Vartabedian, & Kopszak, 2012), or types of explanations, such as describing evidence or giving explanations (Callanan & Jipson, 2001; Crowley et al., 2001a; Szechter & Carey, 2009; Van Schijndel & Raijmakers, 2016). For example, Szechter and Carey, focussing on learning of the individual child, investigated families' explanations while they

visited an observatory exhibition: the researchers coded six different kinds of explanatory talk for child and parent separately, such as describing evidence and giving explanations. A second example is the study of Gutwill and Allen in which families' reasoning was investigated in terms of the two key scientific reasoning skills: Proposing Actions (PA) and Interpreting Results (IR). PA involved asking questions or making plans at the beginning of an investigation, and IR involved making observations, interpretations, or explanations during or after an investigation. The most informative way of analysing conversations strongly depends on the type of exhibit being discussed.

A third indicator of learning through investigation is post-activity learning outcomes (i.e., what someone remembered, learned or understood), which can be considered the traditional way of measuring learning. In the museum context, these outcomes have been measured by assessing families' knowledge on the function of an exhibit (e.g., procedural or conceptual; Fender & Crowley, 2007), conceptual knowledge (Van Schijndel & Raijmakers, 2016), or content-related talk (Haden et al., 2014). For example, Van Schijndel and Raijmakers used a pretest-posttest design to study preschoolers' change in conceptual knowledge on shadow size as a result of the guided visit to a shadow exhibition. In contrast to Van Schijndel and Raijmakers, Haden et al, focused on family learning, investigated post-activity learning outcomes: the researchers assessed the families' content-related talk after exploring a building construction exhibit.

In the current study, for the assessment of learning through investigation we focus on a) the extent to which the family's manipulations are informative and b) the reasoning content and reasoning skills as evidenced by the family's conversation. The latter approach is based upon the socio-cultural theoretical perspective, as we perform in-situ assessments of families' investigations taking the family as a unit of analysis.

Insights in Learning through Investigation in Science Museums. The research in the museum context described above has rendered various insights into families' in situ learning during investigation. It has shown that family talk and exploratory behavior are correlated (Crowley et al., 2001a) and that families with trained verbal inquiry-skills (e.g., proposing actions, interpreting results) investigate longer and deeper, and families' investigations are more aligned with each other (Gutwill and Allen; 2010). Research into family talk also showed that parents keep the conversation going and parent-child dyads, compared to children with their peers, significantly more often describe evidence or give explanations (Crowley et al., 2001a). Research has shown more positive effects of parents' presence: children spend more time investigating and the quality of children's investigation improves (Crowley et al., 2001a), especially through parents' manipulations (Willard et al., 2019) and explanations (Van Schijndel & Raijmakers, 2016). Additionally, research showed that parents' explanations during investigation improved children's conceptual understanding (Fender & Crowley, 2007) and investigation strategies (Van Schijndel & Raijmakers, 2016). Research into exhibit design and families' learning through investigation showed that the quality of children's investigation differs between exhibits (Van Schijndel et al., 2010): at exhibits that invite families to design

and conduct experiments (APE-exhibits) children more often describe evidence and give directions (Szechter and Carey; 2009). Guidance by museum educators can offer families additional opportunities to learn through investigation (Pattison et al., 2018).

Research into the epistemic resources that families use during investigation showed that families often use their own observations of scientific phenomena and their own knowledge (Zimmerman et al., 2010). It was found that children's pre-knowledge enhances the richness of families' content related talk (Haden et al., 2014) and that families' pre-knowledge enhances the quality of their investigations (Benjamin et al., 2010; Franse, Van Schijndel, & Raijmakers, 2020). Previous research has demonstrated that not only prior knowledge, but also other individual characteristics affect families' investigations, such as gender (Benjamin et al., 2010; Crowley, Callanan, Tenenbaum, & Allen, 2001b; Luce, Callanan, & Smilovic, 2013), science interest (Szechter & Carey, 2009; Tare, French, Frazier, Diamond & Evans, 2011) and age (Geerdts, Van De Walle & LoBue, 2015; Marcus, Haden, & Uttal, 2018). In the current study, to examine in a detailed manner factors contributing to families' investigation, we include several parent and child characteristics. This approach is based upon the cognitive developmental theoretical perspective, as we assess individual characteristics.

3.1.2 Guidance by Museum Educators

Museum educators (also named explainers, facilitators, Uyen Tran & King, 2007) have to deal with a large heterogeneous group of people they hardly know, and with a museum offer that is very diverse in content and degree of open-endedness. Visitors differ from one another not only in age, prior knowledge, and interests, but also in their agendas. To contribute positively to visitors' experiences, museum educators must be sensitive to the needs of visitors, which in some cases might mean that the best strategy is to not interfere (National Research Council, 2009). Autonomy plays an important role in the motivation to learn (Ames, 1992; Linnenbrink, 2007; Pekrun & Linnenbrink-Garcia, 2010; Ryan & Deci, 2000), in particular in an informal context (e.g., Barton & Tan, 2010, Falk & Dierking, 2000). Adults within a family group often take on a gatekeeper role, indicating whether the family needs guidance (Pattison & Dierking, 2012, 2013).

Although science museums are set up in such a way that visitors can find their own way, museum educators do play an important role in supporting visitor learning (Bevan & Xanthoudaki, 2008; Hein, 1998; Hooper-Greenhill, 2007; National Research Council, 2009). Given this important role, it is remarkable how little research has been published on museum educator guidance strategies for stimulating families' investigation during unstructured museum visits (Falk 2004; Uyen Tran & King, 2007; Pattison et al., 2017; cf., Ash, 2004b; Piqueras & Achiam, 2019). In particular, little quantitative research on the effectiveness of museum educator guidance strategies has been published (Pattison et al., 2018).

Pattison et al. (2017, 2018) studied, with a design-based research approach, how expert educators facilitated families' learning at interactive math exhibits. A responsive (i.e., observe,

support, reflect) facilitation model (REVEAL) was developed by the researchers, to flexibly apply different facilitation strategies dependent on families' individual characteristics and behavior. Pattison et al. (2018) found that facilitation has a positive impact on families' engagement time, mathematical reasoning, and satisfaction. As the findings relate to math, further research is needed to determine if the facilitation model and strategies are also effective in other exhibits and contexts.

Also insights into parental guidance strategies in the museum context may provide inspiration for the further development of educators' guidance strategies. Based on earlier research of parent-child interactions during every day scientific thinking (e.g., Callanan & Jipson, 2001; Crowley et al 2001a; Fender & Crowley 2007), Van Schijndel & Raijmakers (2016) analyzed parents' guidance strategies when visiting a shadow exhibition together with their preschooler. Especially describing evidence (e.g., talk about exhibit features or observations), was shown to be positively related to the level of children's exploratory behavior. Follow-up research with experimental design is needed to be able to make causal claims about the relation between describing evidence and exploration. A second relevant question is whether this guidance strategy can be generalized to museum educators who guide heterogeneous family groups during an unstructured museum visit.

3.1.3 Present Study

Research questions. The present study examines the conditions that contribute to families' learning through investigation at an open-ended exhibit. We present an in-depth study about families' situated learning with and without guidance that addresses three research questions:

- What is the quality of families' learning through investigation at an open-ended exhibit in terms of manipulations and conversations? (RQ-1)
- Do individual characteristics (i.e. family members' individual cognitive abilities and science engagement, and parents' beliefs about how children learn) relate to the quality of families' learning through investigation? (RQ-2)
- Do minimal forms of guidance by museum educators increase the quality of families' learning through investigation? (RQ-3a). And are both guidance by Giving Explanations and by Describing Evidence more effective than without guidance? (RQ-3b).

Operationalization. We used a randomized controlled trial with three conditions: one control condition without guidance and two experimental conditions in which families were guided by a museum educator. Keeping in mind the goals of facilitating open-ended investigation *and* offering opportunities for learning, we chose as guidance strategies: 1) the educator describes evidence, as a co-investigator would do (Szechter & Carey, 2009), and 2) the educator gives explanations, as an expert would do. Compared to other strategies (e.g., asking questions), describing evidence can be classified as less-invasive, since families do not

need to respond to the educator's guidance. Giving explanations, on the other hand, is a strategy that educators typically use (e.g., Pattison et al., 2017). By explaining, the educator takes on a knowledgeable role, and consequently positions the visitor in the passive role of information receiver, which might evoke feelings of incompetence (Gutwill, 2008). On the other hand, explaining provides the visitor with scientifically sound information.

At the open-ended exhibit that was selected for the present study, visitors could compare the acceleration of two cylinders rolling from an inclined plane. By investigating, families could find out which physical properties do (i.e., mass distribution), and which do not (i.e., mass and size) predict acceleration differences. Many people, including adults, hold the misconception that an object's acceleration can be explained by mass (Galili, 2001; Hast & Howe, 2012; Inhelder & Piaget, 1958). At the exhibit, Control of Variables Strategy experiments (CVS; Klahr & Nigam, 2004) could be performed by comparing the acceleration of two cylinders that differed in one physical property, but were otherwise equal. The inclined plane was surrounded by text panels explaining the scientific method, and exposing questions about reliable investigation. The exhibit as a whole provided two lines of investigation that families could pursue: the specific experiment (*Rolling cylinders*), and the more general line of investigation of 'how investigation works' (*Reliable outcomes*).

Family teams of one parent and two children between eight and twelve participated in the study. This age range was chosen because: 1) 8- to 12-year-olds can already verbally express themselves about investigation-related topics, and 2) the development of 8- to 12-year-olds' ability to align observed evidence with prior conceptions is debated in the literature (e.g., Chinn & Malhotra, 2002a).

Learning through investigation was measured by observing families' manipulation and their reasoning as becomes evident in their conversation during investigation. As in the formal inquiry-based learning literature, the use of CVS is often applied as a measure for the quality of investigation (Klahr & Nigam, 2004).We measured families' manipulation during investigation by: 1) the proportion of CVS experiments (pE_{cvs}), and 2) the frequency of the three physical properties (mass, size, mass distribution) tested by CVS experiments. Reasoning was measured in two ways. First, the reasoning *skills* were assessed using an adapted version of Gutwill and Allen's (2010) Proposing Actions (PA) and Interpreting Results (IR) distinction. Secondly, the reasoning *content* of the conversations *during* investigation was assessed by noting the frequency of scientifically correct and incorrect remarks about the three physical properties in PA and IR ($PA\text{-}M_c$, $IR\text{-}S_i$, et cetera). People's conceptual understanding of scientifically sound concepts of natural phenomena is a slow process. Although museum experiences with for example counterintuitive exhibits can contribute to this learning process (Strike & Posner, 1992), conceptual change is difficult to achieve and does not tend to be completed until students' high school years (Carey, 2000). It is therefore not likely to happen as a result of a single science museum visit. Nevertheless, people might start using new concepts when talking about phenomena, which does not necessarily imply that a coherent, scientifically sound concept was acquired (Straatemeier, Van der Maas, & Jansen, 2008; Van Schijndel, Van

Es, Franse, Van Bers, & Raijmakers, 2018b). We disregard the educator input during guidance; rephrasing the educator's comments is considered a relevant learning experience.

To answer the second research question, several individual characteristics of the participating family members were assessed. Executive and cognitive functioning is important for learning (Kirschner et al., 2006), but can be challenging when learning through investigation takes place in a museum environment (e.g., museum fatigue, Allen, 2004). Therefore, the children's working memory, as an executive function (Diamond, 2013), and verbal and perceptual reasoning skills, as indicators of cognitive abilities, were measured. For parents, their highest education level was noted. In addition, interest and intrinsic motivation are important preconditions for learning (Vansteenkiste, Simons, Lens, Sheldon & Deci, 2004). Therefore, the children's science enjoyment, and the parents' interest in science were measured. Lastly, adults' epistemic beliefs with respect to learning were assessed as an indication of the extent to which parents value self-discovery in children's learning (Ricco & Rodriguez, 2006).

3.2 METHOD

3.2.1 Participants

112 Families visiting NEMO Science Museum in Amsterdam participated in the study. Eight families were excluded from the analyses due to technical problems (3), self-reported ADHD and autism (1), withdrawing from the study (1), not returning informed consent (1), and being classified as outliers (2). The final sample comprised of 104 families consisting of one adult (P, M_{age} = 43.92 years, SD = 5.44; 59 males, 45 females) and two children (M_{age} = 10.16 years, SD = 1.53; 101 boys, 107 girls). We distinguished the oldest (C_O, M = 11.09, SD = 1.24, 45 boys, 59 girls) and youngest children (C_Y, M = 9.23, SD = 1.21, 56 boys, 48 girls) in the analyses. In most families (80%), the adult was the parent or caretaker of both children. Some parent-child dyads were accompanied by the child's friend (19%), and some children were accompanied by a relative (1%). See 3.1 for more individual characteristics.

3.2.2 Procedure and Study Design

The study was conducted during weekend-days from April to July 2016. Most families (88%) were recruited in the museum, a smaller part (12%) registered before their visit via the museum's website. During museum recruitment, visitors with children (8-12 year-olds) were approached and asked to participate in a scientific study in which they would play at an exhibit while video recordings were made. The criteria for approaching families were their language and group composition. All potentially matching families were asked to participate. Visitors mentioned as the main reason for not participating having alternative plans for their visit. If families agreed to participate, they were welcomed in a research room where they completed written consent forms. Subsequently, the participants each received a small bluetooth

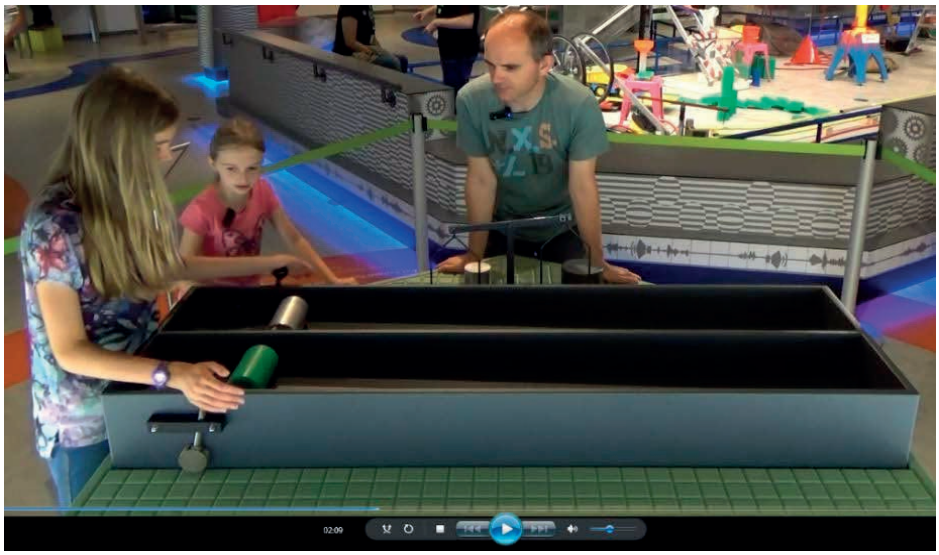
microphone and then were invited to play eight minutes at the exhibit[6]. Families were assigned to a control or guidance condition group without being aware that the study included other conditions. To control for day of the week and time of day, time slots were assigned to experimental conditions at the start of the day. Recruiters were partially aware of the conditions (in approximately 30% of the cases), but had to approach all families meeting the language and composition criteria. The experimental session consisted of three parts: an introduction, investigating at the exhibit, and a posttest. The total procedure took about half an hour.

Exhibit. The open-ended exhibit provided two lines of investigation that families could pursue: a specific experiment, whereby families could investigate which physical properties can (i.e., mass distribution) and cannot (i.e., mass and size) predict acceleration differences (*Rolling cylinders*). And a more general 'how investigation works' line of investigation, whereby families could examine, by investigating and reading exhibit labels, which preconditions (i.e., equal comparisons, repeated measurements) are important for scientific investigation (*Reliable outcomes*). The exhibit consisted of an inclined plane, and four cylinders that differed in mass, size, and mass distribution (see Figure 1). The exhibit was designed to compare two cylinders per experiment. When doing so, families could perform six different experiments: three CVS (Chen & Klahr, 1999; Van Schijndel, Visser, van Bers, & Raijmakers, 2015) and three non-CVS experiments, in which the cylinders differed in two physical properties (e.g., mass and size, cylinder B and D). Families were free to choose how to approach their investigation (e.g., the combinations and amounts of cylinders, how often to repeat or vary an experiment). The exhibit included a balance scale to compare cylinder mass, and was surrounded by exhibit labels and visuals introducing the empirical cycle (i.e., Investigate, Hypothesize, Experiment, Results, and Conclude).

Guidance. Families were assigned to one of three conditions. In the two experimental conditions, families were accompanied by a museum educator applying either the Describing Evidence or Giving Explanations guidance strategy. In the Describing Evidence condition (ED, $N = 25$) the educators named relevant task aspects and observations (e.g., "the cylinders have the same size, but different weight"), the results of experiments (e.g., "the cylinders arrived at the same time"), or relevant preconditions of investigation (e.g., "the two levers have not been pulled at the same time"). In the Giving Explanations condition (EX, $N = 27$) the educators explained the causal relationship between the cylinders' properties and acceleration (e.g., "weight does not affect speed, how the weight is distributed does."), and the preconditions to investigate in a scientific way (e.g., "to be able to compare which one goes faster, the two cylinders should start at the same time."). In the Control condition (C, $N = 52$) families were not accompanied by a museum educator.

[6] We were primarily interested in the quality of the visitors' investigation, in relation to the individuals' characteristics. Engagement time is very sensitive to a research setting (in comparison to a completely free setting; Bamberger & Tal, 2007; Falk, Koke, Price, & Pattison, 2018). Hence, we were not convinced that engagement time in our setting (the constraints that go together with recording video) would be a measure with strong external validity. Therefore, we decided that the quality of investigation was best compared with (approximately) equal-length, that is, constant, engagement times.

Museum educators instead of researchers performed the test-leader roles, contributing to the ecological validity of the study. Two educators were trained by a researcher (first author), which included video-analysis of families engaging at the exhibit, and the practicing of different guidance strategies at the exhibit. The museum educators and conditions were counterbalanced. Fidelity to the guidance approach was assessed by testing the effect of museum educators for all dependent measures. No significant effects were found. The guidance strategies were derived from the literature (see Introduction), and subsequently adapted to the specific exhibit by the researchers. This resulted in an extensive manual with lists of literal examples of exhibit-related utterances per guidance strategy (see Table 3.1). The process ensured educators had several, relevant intervention phrases at their disposal, and the video-analysis was aimed at training them in the flexible application of the intervention.



Cylinder Property	A	B	C	D
Mass (in kg)	1	**2**	1	1
Size (diameter in cm)	8	8	8	**12**
Mass Distribution	compact	compact	**hollow**	compact

Figure 3.1. On top: a family playing at the exhibit in the study. Below: The four cylinders available at the exhibit. By rolling down cylinders from an inclined plane, families could find out which physical properties do (i.e., mass distribution) and do not (i.e., mass and size) predict acceleration differences. The exhibit was designed to compare two cylinders per experiment. When doing so, families could perform six different experiments: three control of variables (CVS) experiments and three non-CVS experiments, in which the cylinders differed in two physical properties (e.g., mass and size, cylinder B and D). If bold then a property deviates from the standard cylinder (cylinder A).

Table 3.1

Examples of exhibit-related utterances for the Describing evidence and Giving explanation guidance strategies.

	Describing evidence	Giving explanations
Rolling cylinders	● The 'rollers' are the same weight, but not the same size. ● The cylinders differ both in size and weight. ● They are the same weight, but the weight is distributed differently. On one 'roller' all the weight is on the outside, on the other it is equally distributed.	● How fast a 'roller' is down (consequence) only depends on how the weight is distributed (cause), the weight itself does not matter. ● A cylinder with all the weight on the outside (cause) rolls much slower (consequence) than a cylinder with the weight evenly distributed. ● The size of a cylinder (cause) has no influence on how fast the 'roller' is down (consequence).
Reliable outcomes	● The levers were not pulled at the same time. ● The 'rollers' arrived at the same time. ● I saw that this cylinder rolled a bit oblique, the other one rolled down in a straight line.	● Inaccurate investigation can result in different results, therefore it is good to repeat an experiment a number of times. ● In order to be able to compare the cylinders in a fair way, the levers have to be pulled down at the same time

Note. Examples of exhibit-related utterances that were used to train museum educators. Examples are given for both lines of investigation (Rolling cylinders and Reliable outcomes) and both guidance strategies (Describing Evidence and Giving Explanations).

3.2.3 Materials

Posttest. Fifteen individual characteristics were measured through individual questionnaires (adult and children) and tasks (children). This resulted in 16 variables (see Table 3.2): six describing the parents (i.e., gender, age, science interest, two educational levels, and beliefs about learning), and five for each of the two children (i.e., gender, age, science enjoyment, reasoning, and working memory).

Background questions (adult). The adult questionnaire included questions about the adult's gender and age, and the gender and age of both children. The adults' *Interest in Science* was assessed by three 4-point Likert-scale questions (α=.54), relating to the frequency (e.g., 0=never to 3=weekly) with which participants read science-related newspaper articles, listened to or watched science shows on radio and television, and visited science museums. Interest in Science sum scores (range 0-9) are used in analyses. The adults were asked about their *educational level* (Up to Bachelor's degree (L), Bachelor's degree (B), Graduate degree (G)) and this level is used in analyses in two nominal variables (L and B versus G).

How children learn inventory (adult). The adults' beliefs about learning (i.e., learning as active or passive process) was assessed using a 16-statement questionnaire (How Children Learn Inventory; Ricco & Rodriquez, 2006). The adults rated their agreement with the statements on a 5-point-Likert scale (1=strongly disagree to 5= strongly agree). An example

statement is: “When it comes to math and science, children learn by doing, i.e., through hands-on experience and by trying out ideas”. The measure was translated from English into Dutch. Factor analysis resulted in a factor referring to the belief that learning is an active process (LB-A; α= .64, statements 5, 7, 9 and 15), describing a child who learns by experimenting, reasoning and drawing conclusions, 'trial and error', gaining success experiences, and receiving positive adult feedback (Ricco & Rodriquez, 2006; see also 3.7 Supplementary Material). Sum-scores are used in further analyses.

Science enjoyment (children). The children’s questionnaire (α=.63, 7 items) consisted of the Science Enjoyment subscale of the Dutch science and technology attitude instrument for primary school pupils (Walma van der Molen, Wiegerinck, & Rohaan, 2007). The children rated their agreement with statements on a 4-point-Likert scale (1= strongly disagree to 4= strongly agree). An example question is: “I like to learn more about science”. Average scores of the Science Enjoyment subscale are used in further analyses.

Reasoning and working memory (children). The children performed three subtasks of the Dutch version of the *Wechsler Intelligence Scale for Children* (third edition; Kort et al., 2005). Assuming that high quality exploration and conversation is related to cognitive skills and executive functioning (Kirschner et al., 2006), the children’s verbal reasoning (Similarities subtest), perceptual reasoning (Block Design subtest), and working memory (Digit Span Backwards subtest) were tested in the present study. A summed standardized raw score for Reasoning (IQ-R = Zscore Similarities and Zscore Block Design), and a standardized raw score for Working Memory (IQ-WM = Zscore Digit Span Backwards) are used in further analyses.

Coding approach. During the investigation, the families’ manipulation and conversation were recorded on video and audio: two cameras were situated at the exhibit, and all family members (i.e., P, C_O and C_Y) wore wireless microphones. Scoring was based on transcripts of the recordings (in CLAN: MacWhinney & Snow, 1990). Transcripts were first broken down into segments of experiments, whereby an experiment was defined as starting when the cylinders were released and ending when the cylinders hit the bottom edge of the exhibit (see also Figure 1). Each experiment was scored as CVS (1) or not (0) with the *manipulation coding instrument*, and each conversation between two experiments (referred to as ‘speech segment’) was scored as high-level proposing actions (PA=1) or not (PA=0), and as high-level interpreting results (IR=1) or not (IR=0) using the *reasoning skills coding instrument*. Subsequently, the *content* of PA and IR was scored using the *reasoning content coding instrument*. The three coding instruments are further explained below. Inter-rater reliabilities were calculated for the assigned level of PA and IR for each conversation section between two experiments.

Manipulation. The total amount of performed experiments and CVS experiments were determined, and the proportion of CVS experiments (pE_{CVS}) was used in the analyses as a measure for the quality of manipulation (see also 2.2.1).

Reasoning: Skills. The quality of reasoning *skills* were scored by determining the number of experiments in which high-level proposing actions (PA) and high-level interpreting

results (IR) were made. PA and IR (Gutwill & Allen, 2010) are questions, remarks or short conversations made by one or more family members at the beginning or end of an experiment (see also 2.3.2). In the current study, PA include hypotheses (e.g., ”Which cylinder will arrive first?”, “Uhm, this one?”, “Yeah, I think so too, the heaviest one will run fastest.”), or justified actions (e.g., “Wait, we have to start at the same time, otherwise we can’t compare them!”), and IR include causal explanations of the observed result (e.g., ”Ay, mine slowed …”, “… because it bumped into the edge”); see the Supplementary Material (3.7.1 Materials) for the scoring rules. Proportions of experiments preceded by PA (pPA) and proportions of experiments followed by IR (pIR) were used in regression analyses. Inter-observer reliability was found to be ’substantial’ (Landis & Koch, 1977): percentage agreement = 82, and kappa = 0.61 ($p < 0.001$), 95% CI (0.497, 0.722).

Reasoning: Content. The quality of reasoning *content* was scored by classifying each PA and IR with a seven-subscale coding instrument. Five subscales were linked to the *Rolling cylinders* line of investigation of the exhibit, distinguishing scientifically correct (c) and incorrect (i) remarks about causal relations (i.e., PAs and IRs) between acceleration and cylinder properties mass (M), size (S) and mass distribution (D): subscale 1 – Distribution of mass matters for acceleration (PA-D_c, IR-D_c), subscale 2 – Mass does not matter for acceleration (PA-M_c, IR-M_c), subscale 3 – Mass matters for acceleration (PA-M_i, IR-M_i), subscale 4 – Size matters for acceleration (PA-S_i, IR-S_i), subscale 5 – Other (PA-X, IR-X). Two remaining subscales were linked to the *Reliable outcomes* line of investigation of the exhibit: subscale 7 – equal rolling preconditions are important for scientific reasoning (PA-F, IR-F) and subscale 8 – equal inclined planes are important to compare acceleration (PA-I, IR-I). The coding resulted in ten *Rolling cylinders*, and four *Reliable outcomes* variables. However, factor analyses (see 2.3.3) showed that the latter four reasoning *content* variables could be reduced. Therefore, in analyses sum scores of the equal rolling precondition variables PA-F and IR-F (PAIR-F), and sum scores of the equal inclined plane variables PA-I and IR-I (PAIR-I) were used. To conclude, 12 dependent variables were used in the analyses to describe the *content* of families’ reasoning. Inter-observer reliability was found to be ‘substantial’ (Landis & Koch, 1977): percentage agreement = 87, and kappa = 0.80 ($p < 0.001$), 95% CI (0.735, 0.850).

Analysis approach. To study the families’ investigation (RQ-1), and the factors that might impact investigation quality, i.e., individual characteristics (RQ-2) and guiding style (RQ-3), we chose a quantitative analysis approach. For the first research question, factor analyses were performed to determine the dependent variables that were at the core of describing the families’ investigations (see 3.7 Supplementary Material). For the resulting 15 dependent variables (one manipulation, two reasoning *skills*, and twelve reasoning *content*) descriptions of the non-guided control condition were calculated. For the second question, 15 backward regression analyses (starting with a full model and removing predictor variables until the decrease of explained variance became significant) were performed, each with 16 variables (i.e., six parent, five oldest child, and five youngest child independent variables), to find out in what way the individual characteristics affected the quality of manipulation and conversation

(model 1, with an a priori power of .75, medium effect size, alfa = .05). For the last question, for all models 1, a follow-up regression analysis was performed. Optimal model 1 was extended with guiding style (ED and EX versus C). Change statistics of model 2 to model 1 were calculated to examine if guiding style significantly explained more variance than individual characteristics alone. As a robustness check, we checked the full regression models (with all co-variates entered) to test the robustness of the resulting models 1. The resulting covariates that are significant under bonferroni correction in the backward regression models are all significant in the full regression models. Data points in age, working memory (IQ-WM), or science enjoyment were missing for participants of seven families. Therefore, multiple regression analyses were performed with 97 instead of 104 families.

3.3 RESULTS

3.3.1 Descriptions of Families' Individual Characteristics

Nine of fifteen individual characteristics of the families (n=104) are described below (see also Table 3.2), gender and age are described in 2.1 Participants.

Science enjoyment and interest. Notable, but to be expected in a science museum, is that the children were highly engaged in science (oldest: $M = 3.31$, $SD = 0.39$, and youngest: $M = 3.32$, $SD = 0.45$). The children demonstrated higher science enjoyment (Sc.Enj.; $M = 3.31$, SD=.42) than average primary school pupils ($M = 2.69$; Walma van der Molen et al., 2007), $t(199) = 20.89$, $p < .001$. The adults were also interested in science, but not excessively (Sc.Int.; $M = 4.22$, $SD = 2.09$). Half of the parents reported reading the newspaper science supplement weekly to monthly. Over half of the parents visited science or natural history museums half-yearly to annually, and listened to or watched science programs on radio or television weekly to monthly.

Beliefs about learning: Learning is an active process (LB-A). On average, the adults agreed (sum score 16) to strongly agreed (sum score 20) with the belief that learning is an active process (LB-A; $M = 17.38$, $SD = 2.13$). This view is in line with the museum's hands-on approach, and the set-up of the current study's exhibit.

Educational level. The parents' educational level (72% Bachelor or Graduate, 28% Up to bachelor) was higher than the average level of 35-55 year olds in the Netherlands (41% Bachelor or Graduate, 57% Up to bachelor, 2% unknown; Statistics Netherlands, 2016, StatLine), $t(103) = 7.04$, $p < .001$.

Reasoning and working memory. The children's verbal reasoning, assessed by the Similarities subtask (Norm score $M = 11.43$, $SD = 2.93$), was above the Dutch average ($M = 10$), $t(207) = 7.02$, $p < .001$. The children's perceptual reasoning, assessed by the Block Design subtask (Norm score $M = 11.11$, $SD = 3.19$), was also above the Dutch average ($M = 10$), $t(207) = 5.00$, $p < .001$. The children's working memory, assessed by the Digit Span backwards subtask (Norm score $M = 9.41$, $SD = 4.17$), was lower than the Dutch average ($M = 10$), $t(205)$

= -2.03, $p < .05$). Note that in the analyses instead of norm scores, summed standardized raw scores for reasoning (IQ-R), and standardized raw scores for working memory (IQ-WM) were used. As expected, both reasoning capacity (IQ-R) and working memory (IQ-WM) were higher for the oldest children compared to the youngest children, $t(206) = 6.63$, $p < .001$ and $t(203) = 2.93$, $p < .005$ respectively (Huizinga, Dolan, & van der Molen, 2006, Waber et al., 2007).

Distribution of individual characteristics. To confirm equal distribution of the individual characteristics across the guiding style conditions, MANOVAs (ratio variables) and Chi Square analyses (nominal variables) were conducted for all 16 variables. This showed that only the children's working memory (IQ-WM) differed between conditions ($F(2, 102) = 4.54$, $p < .05$, $F(2, 102) = 3.10$, $p < .05$). Both the oldest ($M = 5.35$, $SD = 1.74$) and youngest children ($M = 4.71$, $SD = 1.51$) in the control condition (i.e., without guidance) had significantly higher IQ-WM than the oldest (M=6.60, SD=1.92) and youngest children ($M = 5.72$, $SD = 1.46$) in the Evidence Description condition ($p_{Tukey\ HSD}$ =.009, $p_{Tukey\ HSD}$ =.038). No differences between conditions were found for the other individual characteristics.

Summary: Profile of participating families. The participants in the study were families of two children and one adult. On average, the oldest and youngest children differed two years in age and enjoyed science. Both the oldest and youngest children scored above average on the reasoning tasks, but scored under the Dutch standard on the working memory task. This is remarkable because reasoning and working memory normally correlate positively. We therefore interpret the results on the Digit Span backwards task with care. The adult had relatively more education than average and was moderately interested in science. On average, the adults saw their child's learning as an active process.

Table 3.2
Covariates, describing individual characteristics of parent, oldest child and youngest child.

			Total	C	ED	EX
Parent (P)						
Gender	Male	*N* (%)	59 (57)	29 (56)	13 (52)	17 (63)
	Female	*N* (%)	45 (43)	23 (44)	12 (48)	10 (37)
Age		*M* (*SD*)	43.92† (5.44)	44.43 (6.04)	43.41 (4.46)	43.37† (5.13)
Science interest		*M* (*SD*)	4.22 (2.09)	4.15 (2.37)	4.16 (1.84)	4.41 (1.76)
Beliefs about learning LB-A		*M* (*SD*)	17.38 (2.13)	17.19 (2.65)	17.52 (1.48)	17.63 (1.45)
Educational level	Up to Bachelor (L)	*N* (%)	29 (28)	17 (33)	5 (20)	7 (26)
	Bachelor (B)	*N* (%)	44 (42)	26 (50)	10 (40)	8 (30)
	Graduate (G)	*N* (%)	31 (30)	9 (17)	10 (40)	12 (44)
Oldest Child (C_O)						
Gender	Male	*N* (%)	45 (43)	24 (45)	10 (40)	11 (41)
	Female	*N* (%)	59 (57)	28 (55)	15 (60)	16 (59)
Age		*M* (*SD*)	11.09 (1.24)	11.13 (1.35)	11.24 (1.06)	10.86 (1.18)
Science enjoyment		*M* (*SD*)	3.31† (0.39)	3.32† (0.42)	3.27 (0.35)	3.33 (0.39)
Digit Span Backwards	Working memory	*M* (*SD*)	5.74† (1.76)	5.35† (1.74)	6.60 (1.92)	5.67 (1.39)
Similarities	Reasoning	*M* (*SD*)	18.08 (4.44)	17.88 (4.35)	17.92 (4.49)	18.59 (4.68)
Block Design	Reasoning	*M* (*SD*)	49.61 (11.09)	48.38 (12.11)	53.16 (10.85)	48.67 (8.66)
Youngest Child (C_Y)						
Gender	Male	*N* (%)	56 (54)	27 (52)	14 (56)	15 (56)
	Female	*N* (%)	48 (46)	25 (48)	11 (44)	12 (44)
Age		*M* (*SD*)	9.23 (1.21)	9.24 (1.26)	9.36 (1.19)	9.12 (1.16)
Science enjoyment		*M* (*SD*)	3.32† (0.45)	3.34† (0.43)	3.29 (0.53)	3.29 (0.43)
Digit Span Backwards	Working memory	*M* (*SD*)	5.03† (1.70)	4.71† (1.51)	5.72 (1.46)	5.00† (2.10)
Similarities	Reasoning	*M* (*SD*)	13.62 (4.26)	13.10 (4.28)	14.00 (4.84)	14.26 (3.62)
Block Design	Reasoning	*M* (*SD*)	42.48 (12.07)	41.67 (9.50)	43.76 (14.61)	42.85 (14.16)

Note. Total number (*N*) and percentages (%) of the family members' gender and the parents' educational level are presented, and average values (*M*) and standard deviations (*SD*) of the other individual characteristics. LB-A = Belief about learning: Learning is an active process. Digit Span Backwards = WISC III-subtest within the Working Memory Index. Similarities = WISC III-subtest within the Verbal Comprehension Index. Similarities, measures the child's abstract, logical thinking and reasoning (Sattler, 1974). Block Design = WISC III-subtest within the Perceptual Reasoning Index. Block Design, measures the child's ability to analyze and synthesize abstract visual stimuli (Sattler, 1974). Wechsler Intelligence Scale for Children (WISC). Total = all participating families (*N* = 104). C= without guidance control condition (*N* = 52). ED =Describing evidence guidance condition (*N* = 25). EX = Giving explanation guidance condition (*N* = 27). † data points were missing for the parents' age, and the children's Digit Span Backwards and science enjoyment. Therefore parent's age N_{TOTAL} = 103 and N_{EX} = 26; youngest children's IQ-WM N_{TOTAL} = 103 and N_C = 51; oldest children's Digit Span Backwards N_{TOTAL} = 102, N_C = 51 and N_{EX} = 26; youngest children's science enjoyment N_{TOTAL} = 100 and N_C = 48; oldest children's science enjoyment N_{TOTAL} = 100 and N_C = 48.

3.3.2 Families' Learning through Investigation (RQ-1)

Families' investigation quality is described by three types of dependent variables: manipulation, reasoning *skills*, and reasoning *content*. Total values, and values per guiding style condition, are reported in Table 3.3. Below we describe the families' investigation quality in the non-guided control condition, which can be seen as a baseline of a regular, unstructured museum visit (RQ-1).

Manipulation. On average, the families carried out 11.00 experiments (*SD* = 3.36). Sixty four percent (*SD* = 22%) of the experiments were CVS experiments (E_{CVS}), which is above the 50% probability level, $t(51) = 4.61$, $p < .001$. Of the experiments 24% were focused on mass, 21% on size and 20% on mass distribution.

Reasoning: Skills. The families' reasoning was focused on scientific investigation. On average, 85% of the experiments were preceded and 92% were followed by investigation remarks. PA's were made for 28% (*SD* = 21%) and IR's were made for 50% of the experiments (*SD* = 18%).

Reasoning: Content. Reasoning *content* was analyzed for both lines of investigation that families could pursue at the exhibit: 84% of the high-level remarks (PA and IR) were about *Rolling cylinders*, and 16% about *Reliable outcomes*. The most frequently made remark was IR-M_i, the incorrect explanation that mass determines rolling acceleration (*M* = 5.12 times per family, *SD* = 3.34). Incorrect hypotheses about mass (PA-M_i) were also made (*M* = 2.35 times per family, *SD* = 2.54). Note that IR (*M* = 5.46, *SD* = 2.19) is higher than PA (*M* = 2.94, *SD* = 1.91). Therefore, relatively speaking, IR-M_i and PA-M_i were mentioned an equal number of times (54% and 46%). The most frequently made correct remark about the relation between cylinder property and acceleration was the explanation that mass distribution causes acceleration differences (IR-D_c; *M* = 1.58 times per family, *SD* = 1.97). With regards to the *Reliable outcomes* line of investigation, families formulated more hypotheses about preconditions for equal rolling (PA-F; *M* = 0.73, *SD* 1.24) than they gave causal explanations (IR-F; *M* = 0.48, *SD* = 0.87). Notably, the families also made remarks about a possible slope difference between the two roller tracks, a difference that was not pre-set in the exhibit design. The families mentioned slope differences equally often in their hypotheses (PA-I; *M* = 0.56, *SD* 0.94) and explanations (IR-I; *M* = 1.06, *SD* = 1.79).

Summary: Quality of families' scientific investigation. The first research question asked what the quality of investigation of families engaging with an open-ended exhibit was (RQ-1). On average, the families performed 11 experiments in eight minutes. They tried out many different combinations of cylinders, and their use of CVS experiments was above chance level. The families were focused on the investigation in their behavior and conversation. Part of these conversations were at a high level. For example, families formulated hypotheses and gave causal explanations for observed results. The *content* of the high-level remarks concerned both lines of investigation that families could pursue at the exhibit (*Rolling cylinders* and *Reliable outcomes*), although most remarks were about causal relations between cylinder

properties and acceleration (i.e., *Rolling cylinders*). Notably, the most frequently made correct remark was related to mass distribution determining acceleration, and the most frequently made incorrect remark related to mass determining acceleration. The families additionally made remarks about a possible slope difference between the two roller tracks that was not pre-set in the exhibit design.

3.3.3 The Impact of Individual Characteristics and Minimal Guidance Strategies (RQ-2, RQ-3)

Multiple regression analyses were performed to test whether families' investigation quality could be explained by individual characteristics (RQ-2) or guiding style (RQ-3). The test statistics of the resulting models, as well as standardized beta coefficients of individual characteristics and guiding style condition are presented in Table 3.4. Effect sizes (Cohen, 1988) of these regression models range from medium to large (i.e., .12 to .32). The regression models for manipulation and reasoning *skills* will be described below, as well as the significant models for reasoning *content*. To prevent capitalization of chance (de Groot, 2014) by a type 1 error, a stricter limit value of significance than regular will be used (i.e., .05/15 = .003). We performed necessary analyses to test assumptions of the regression analyses reported below. There is no concern for multicollinearity: Absolute values of correlations were maximally .6 (age between oldest and youngest child), but for the other predictors maximally around .5. The maximal VIF is 2.7. No clear incidents of heteroscedasticity were observed in the partial plots of predictors and outcome residuals after regressing the predictors separately to all outcome variables.

Manipulation. ***CVS experiments***. A backward regression analysis, with proportion of CVS experiments (pE_{CVS}) as dependent variable and individual characteristics as predictors, resulted in a significant model with two predictors explaining 17% of E_{CVS} variance, $F(2, 94) = 9.47$; $p < .001$. The model shows that higher proportions of CVS experiments were present in families in which the oldest children had relatively higher reasoning capacity (IQ-R; $\beta = .40$, $t = 4.17$, $p < .001$). The second individual characteristic, the youngest children's working memory (IQ-WM), also contributed to the resulting model, but its unique contribution (i.e., after correction for the predictor variables in the resulting model) was not significant. The proportion of CVS experiments was not explained by guiding style (F_{change} (2,92) = 0.07, p = .929).

Reasoning: Skills. ***Proposing actions.*** The proportion of PA (pPA) was not explained by individual characteristics or guiding style. A backward regression analysis with pPA as dependent variable and individual characteristics as predictors resulted in a non-significant model with two predictors: the parents' educational level and the oldest children's reasoning capacity (IQ-R), $F(2, 94) = 5.92$; $p = .004$, $R^2 = .11$. Change statistics for guiding style as an additional predictor were non-significant (F_{change} (2,92) = 1.13, p = .329).

Interpreting results. The proportion of IR (pIR) was not explained by individual characteristics or guiding style. A backward regression analysis with pIR as dependent variable and individual characteristics as predictors resulted in a non-significant model with youngest children's age as predictor, $F(1, 95) = 3.45$; $p = .066$, $R^2 = .04$. Change statistics for guiding style as an additional predictor were non-significant (F_{change} (2,93) = 1.57, $p = .214$).

Reasoning: Content. ***Line of investigation: Rolling cylinders.*** The amount of *correct mass distribution hypotheses* (PA-D_c) was significantly explained by two individual characteristics, together accounting for 12% of the variance, $F(2, 94) = 6.20$; $p < .003$, $R^2 = .12$. Families where the oldest children had relatively higher reasoning capacity (IQ-R) made more correct mass distribution hypotheses ($\beta = .37$, $t = 3.44$, $p = .001$). The second individual characteristic, the oldest children's age, also contributed to the resulting model, but its unique contribution was not significant (i.e., $p > .003$). Guiding style as an additional predictor did not result in a model explaining significantly more variance, F_{change} (2,92) = 0.60, $p = .551$.

The amount of *correct mass distribution explanations* (IR-D_c) was significantly explained by five individual characteristics, together accounting for 18% of the variance, F(5, 91) = 3.95; $p < .003$, $R^2 = .18$. The five individual characteristics, the oldest and youngest children's age and reasoning capacity (IQ-R) and the oldest children's working memory (IQ-WM), contributed to the resulting model, but their unique contributions were not significant (i.e., $p > .003$). Guiding style as an additional predictor did not result in a model explaining significantly more variance, F_{change} (2,89) = 1.14, $p = .326$.

The amount of *incorrect mass hypotheses* (PA-M_i) was significantly explained by two individual characteristics and by guidance style, together accounting for 17% of the variance, $F_{change}(2,92) = 3.46$, $p < .05$; $F(4, 92) = 4.57$, $p < .003$, $R^2 = .17$. The two individual characteristics, the oldest children's gender and reasoning capacity (IQ-R), and the guiding style contributed to the resulting model, but their unique contributions were not significant (i.e. $p > .003$).

The amount of *incorrect mass explanations* (IR-M_i) was significantly explained by five individual characteristics and guidance style, together accounting for 27% of the variance, F_{change} (2,89) = 9.60, p < .001; $F(7, 89) = 4.69$; $p < .001$, $R^2 = .27$. Compared to no guidance, families in the guidance conditions made fewer incorrect mass explanations ($\beta = -.31$, $t = -3.16$, $p < .003$ and $\beta = -.39$, $t = -3.95$, $p < .001$). The five individual characteristics, the youngest and oldest children's age and reasoning capacity (IQ-R), and the youngest children's science enjoyment, also contributed to the resulting model, but their unique contributions were not significant (i.e. $p > .003$).

The amount of *incorrect size explanations* (IR- S_i) was significantly explained by five individual characteristics, together accounting for 20% of the variance, $F(5, 91) = 4.45$; $p = .001$, $R^2 = .20$. The five individual characteristics, the parents' educational level, the oldest children's reasoning capacity (IQ-R) and working memory (IQ-WM), and the youngest children's reasoning capacity (IQ-R) and science enjoyment, contributed to the resulting model, but their unique contributions were not significant (i.e. $p > .003$). Guiding style as an additional

predictor did not result in a model explaining significant more variance, F_{change} (2,89) = 1.12, p = .330.

Line of investigation: Reliable outcomes. The amount of remarks the families make about the track-quality (PAIR-I) can be explained by five individual characteristics and by guidance style, together accounting for 32% of the variance, F_{change} (2,89) = 5.80, $p < .005$; F(7, 89) = 6.04; $p < .001$, R^2= .32. Compared to families with academically-educated parents, families with parents with low educational level and higher vocational education made fewer remarks about track quality (PAIR-I; β = -.58, t = -4.80, $p < .001$ and β = -.39, t= -3.64, $p <$ .001). Compared to families with no guidance, families with Evidence Description guidance made fewer remarks about track quality (PAIR-I; β = -.33, t = -3.39, p = .001). The three other individual characteristics, the oldest children's age, the youngest children's working memory (IQ-WM), the parents' science interest, and the Giving Explanations guiding style also contributed to the resulting model, but their unique contributions were not significant (i.e., $p >$ .003).

Summary: The impact of individual characteristics and minimal guidance. The second research question asked if individual characteristics impacted families' investigation quality (RQ-2). The way families investigate at an open-ended exhibit is affected by certain child and parent characteristics. The oldest child's reasoning capacity has a positive, medium-sized effect on families' manipulation (increase of CVS experiments), and a small effect on the reasoning *content* (increase of correct remarks regarding mass distribution and acceleration). Marginally, the oldest child's reasoning capacity has a unique contribution (medium to large) to making correct and incorrect remarks about cylinder properties and acceleration. This is also the case for the age of both children, whereby the effect of the oldest and youngest child's age is sometimes opposite, that is, in the context of the other predictors. Furthermore, it is striking that the children's science enjoyment, in addition to the other family characteristics, does not affect the families' investigation quality. Additionally, it is shown that families with academically educated parents more often mention the unintended slope differences of the exhibit. Parents' gender and belief that learning is an active process do not affect families' investigation quality. Furthermore, no effects of parent characteristics on the families' manipulations and reasoning *skills* were found.

The third question asked if minimal guidance strategies by museum educators positively contributed to families' investigation quality (RQ-3a) and if both guidance by Giving Explanations and by Describing Evidence are more effective than without guidance (RQ-3b). Guidance had an effect on the reasoning *content* of the families' conversation during investigation. Accompanied by a museum educator, less often the families concluded incorrectly that acceleration differences are caused by mass differences. This applies to both guidance styles. When accompanied by museum educators who Described Evidence, the families also mention the unintended slope differences of the exhibit less frequently. Guidance does not affect families' manipulations during investigation.

Table 3.3
Outcome variables, describing families' investigation quality.

		Total *M(SD)*	C *M(SD)*	ED *M(SD)*	EX *M(SD)*
Manipulation					
	E	11.12 (3.49)	11.00 (3.36)	11.40 (3.58)	11.07 (3.75)
	pE_{CVS}	0.64 (0.21)	0.64 (0.22)	0.62 (0.24)	0.64 (0.18)
Reasoning: Skills					
	PA	2.90 (1.76)	2.94 (1.91)	2.72 (1.86)	3.00 (1.36)
	pPA	0.28 (0.21)	0.28 (0.21)	0.25 (0.24)	0.28 (0.15)
	IR	5.00 (2.26)	5.46 (2.19)	4.76 (2.65)	4.33 (1.86)
	pIR	0.46 (0.20)	0.50 (0.18)	0.43 (0.22)	0.42 (0.22)
Reasoning: Content					
Line of investigation: Rolling cylinders					
Mass distribution matters for acceleration	PA-Dc	0.63 (1.09)	0.58 (1.09)	0.84 (1.25)	0.56 (0.93)
Mass does not matter for acceleration	PA-Mc	0.06 (0.23)	0.08 (0.27)	0.04 (0.20)	0.04 (0.19)
Mass matters for acceleration	PA-Mi	1.93 (2.09)	2.35 (2.54)	1.56 (1.39)	1.48 (1.50)
Volume matters for acceleration	PA-Si	0.75 (1.26)	0.87 (1.37)	0.48 (0.65)	0.78 (1.45)
Other remarks about cylinders	PA-X	0.24 (0.65)	0.27 (0.77)	0.20 (0.50)	0.22 (0.51)
Mass distribution matters for acceleration	IR-Dc	1.70 (1.86)	1.58 (1.97)	1.60 (1.89)	2.04 (1.60)
Mass does not matter for acceleration	IR-Mc	0.64 (0.98)	0.69 (1.09)	0.52 (0.92)	0.67 (0.83)
Mass matters for acceleration	IR-Mi	3.98 (2.99)	5.12 (3.34)	3.12 (2.01)	2.59 (2.14)
Volume matters for acceleration	IR-Si	1.28 (1.65)	1.50 (1.50)	1.24 (1.96)	0.89 (1.60)
Other remarks about cylinders	IR-X	0.57 (0.95)	0.58 (1.00)	0.56 (0.96)	0.56 (0.89)
Line of investigation: Reliable outcomes					
	PA-F	0.72 (1.10)	0.73 (1.24)	0.72 (1.14)	0.70 (0.78)
	PA-I	0.51 (0.88)	0.56 (0.94)	0.32 (0.80)	0.59 (0.84)
	IR-F	0.48 (0.88)	0.48 (0.87)	0.48 (0.82)	0.48 (0.98)
	IR-I	0.77 (1.55)	1.06 (1.79)	0.20 (0.71)	0.74 (1.51)
Equal rolling conditions	PAIR-F	1.20 (1.62)	1.21 (1.73)	1.20 (1.68)	1.19 (1.39)
Equal inclined plane conditions	PAIR-I	1.28 (2.12)	1.62 (2.39)	0.52 (1.45)	1.33 (1.98)

Note. Average values (*M*) and standard deviations (*SD*), of three outcome variables categories (manipulation, reasoning skills and reasoning content) are presented (n = 104). Column 'TOTAL' contains values of all participating families (N = 104). Column 'C', 'ED' and 'EX' contain values of families in the control (N = 52), describing evidence (N = 25) and giving explanation (N = 27) guidance condition respectively. **Manipulation variables**: E = experiments performed during 8 minutes of play. pE_{CVS}= proportion of CVS experiments. **Reasoning *skills* variables**: PA = high level Proposing Action (i.e. formulating hypotheses). pPA = proportion of PA. IR = high level Interpreting Results (i.e., giving causal explanations). pIR = proportion of IR. **Reasoning *content* variables**: *Rolling cylinders*. Remarks can be scientifically correct (c) or incorrect (i). D_c= scientifically correct remark that mass distribution matters for acceleration . M_c = correct remark that mass does not matter for acceleration ; M_i = incorrect remark that mass matters for acceleration ; S_i = incorrect remark that volume matters for acceleration ; X = other remark about cylinder. *Reliable outcomes.* F = equal rolling conditions. I = equal inclined plane conditions. PAIR-F = sum of F in PA and IR. PAIR-I = sum of I in PA and IR. Note that multiple content categories can be assigned to a PA, consequently the sum of PA-F, PA-I, PA-D_c, PA-M_c, PA-M_i, PA-S_i, PA-X (= 4.84) does not equals PA (= 2.90). Note that multiple content categories can also be assigned to an IR, consequently the sum of IR-F, IR-I, IR-D_c, IR-M_c, IR-M_i, IR-S_i, IR-X (= 9.42) does not equals IR (= 5.00).

Table 3.4
Standardized beta coefficients of regression models of families' investigation quality.

	Best model	Standardized beta coefficients																	
		Age			Gender			Education		LB-A	IQ-R		IQ-WM		Sc. Int.	Sc. Enj.		Guidance	
		P	C_O	C_y	P	C_O	C_y	L	B	P	C_O	C_y	C_O	C_y	P	C_O	C_y	ED	EX
pEcvs	***F*(2,94) = 9.47;**										**.40**			-.17					
	***p* < .0001, R^2 = .17**										****			ns					
pPA	*F*(2,94) = 5.92;								-.19		.29								
	p = .004, R^2 = .11								ns		**								
pIR	*F*(1,95) = 3.45;			.19															
	p = .066, R^2 = .04			ns															
PA-Dc	***F*(2,94) = 6.20;**		-.24								**.37**								
	***p* < .003, R^2 = .12**		*								***								
PA-Mc	*F*(1,95) = 4.54,								.21*										
	p = .036, R^2 = .05																		
PA-Mi	***F*(4,92) = 4.57,**					.29					.21							-.22	-.22
	***p* < .003, R^2 = .17**					**					*							*	*
PA-Si	*F*(5,91) = 3.54;			.32							.23	-.24		-.21			.22		
	p = .006, R^2 = .16			**							*	ns		*			*		
PA-X	*F*(4,92) = 2.94;		-.18				-.18		-.18		.23								
	p = .024, R^2 = .11		ns				ns		ns		*								
IR-Dc	***F*(5,91) = 3.95;**		-.26	.42							.21	-.23	.26						
	***p* < .003, R^2 = .18**		*	**							ns	ns	*						
IR-Mc	*F*(4,92) = 2.94,				-.19	.19				.18ns	.18								
	p = .025, R^2 = .11				ns	ns					ns								
IR-Mi	***F*(7,89) = 4.69;**		-.30	.31							.31	-.21					.15	**-.31**	**-.39**
	***p* < .0001, R^2 = .27**		*	*							**	ns					ns	***	****
IR-Si	***F*(5,91) = 4.45;**								.23*		.30	-.24	-.31				.19		
	***p* < .003, R^2 = .20**										**	*	**				ns		
IR-X	*F*(4,92) = 3.34;			.22							.28	-.22			.19				
	p = .013, R^2 = .13			ns							*	ns			ns				
PAIR-I	***F*(7,89) = 6.04;**		.23					**-.58**	**-.39**					-.11	-.30			**-.33**	-.15
	***p* < .0001, R^2 = .32**		*					****	****					ns	**			***	ns
PAIR-F	*F*(1,95) = 4.35							-.21											
	p = .040, R^2 = .04							*											

Note. The first column contains the 15 outcome variables that together describe families' investigation quality: one manipulation outcome variable (row 1), two reasoning *skill* variables (row 2-3), twelve reasoning *content* variables of ten about *Line of investigation: Rolling cylinders* (row 4-14) and 2 about *Line of investigation: Reliable outcomes* (row 15-16). The second column contains multiple regression test statistics of the best model predicting families' (N=97) investigation quality. Standardized beta coefficients are given of individual characteristics (column 3-18) and guiding style condition (column 19 and 20). pE_{CVS} = proportion of CVS experiments; pPA = proportion of PA; pIR = proportion of IR; D_c= correct remarks about the relation of mass distribution and acceleration; M_c= correct remarks about the relation of mass and acceleration; M_i = incorrect remarks about the relation of mass on acceleration; S_i = incorrect remarks about the relation of size and acceleration; X= other remarks about the relation of cylinder properties and acceleration. PAIR-I = PA and IR remarks about equal inclined plane conditions; PAIR-F = PA and IR remarks about equal rolling conditions; P=parent, C_O=oldest children, C_Y= youngest children. Education L = Up to Bachelor's degree parents compared to Graduate degree parents; Education B = Bachelor's degree parents compared to Graduate degree parents; LB-A = Belief about learning: Learning is an active process; IQ-R = children's reasoning capacity; IQ-WM = children's working memory capacity; Sc. Int.= parent's interest in science, Sc. Enj.= children's science enjoyment; Guidance ED= Describing Evidence guiding style; Guidance EX= Giving Explanation guiding style. $^{n.s.}p > .05$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .003$, $^{****}p < .0001$. Bold are test statistics that are significant after Bonferroni corrections ($^{***}p < .003$).

3.4 DISCUSSION

The current study used socio-cultural and cognitive developmental perspectives on learning, in studying families' learning at an open-ended museum exhibit. It considers both the learning in natural situations where all family members contribute to a shared learning situation, and the fact that learning is an individual effort in which individual differences play an important role.

3.4.1 Families' Learning through Investigation

Families' learning through investigation was assessed by measuring three aspects: the manipulation of the exhibit: the extent to which families control the variables in their experiments; the reasoning *skills:* the formulation of hypotheses and interpretation of results; and the reasoning *content:* 1) conversation about cylinder properties that might play a causal role in the object's motion down the incline, and 2) conversation about the reliability of the experiments. First, we studied the families' joint investigation, without the guidance of a museum educator.

Manipulation. The families performed above chance (64 %) in constructing CVS experiments, which might be considered in contrast with the evidence showing children's difficulties with understanding and performing CVS experiments in a formal context (Schwichow, Croker, Zimmerman, Höffler & Härtig, 2016). For example, elementary-school children have often been shown to design non-CVS experiments, preventing them from distinguishing effective variables (Bullock, Sodian, & Koerber, 2009; Chen & Klahr, 2008; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). Although CVS is a domain-general strategy that can be applied to investigate various phenomena, it is not 'content free' (Zimmerman, 2007, p. 175). When applying CVS, variables must be recognized and encoded, which is not equally easy for all content areas and depends among others on domain knowledge (Morris, Croker, Masnick & Zimmerman, 2012). This complicates comparing CVS performance between content areas. One explanation for the above chance CVS performance in this study might be that families visiting science museums already possess CVS knowledge (Klahr & Nigam, 2004; Dean & Kuhn, 2007). A result contradicting this explanation is that no effect of parent education on CVS use was found. A second explanation could be the facilitation offered by the exhibit design: the exhibit included one green colored standard cylinder, and using this green cylinder in combination with another cylinder automatically led to a CVS experiment. In addition, the exhibit labels contained an instruction suggesting CVS performance, but this instruction only comprised 15% of the exhibit-label text. Conversation analysis demonstrated that almost all the families (95%) read the exhibit labels, mostly (81%) at the start of their investigation. Nevertheless, reading the exhibit labels did not automatically result in performing a CVS experiment: 62% of the experiments performed directly after the first time reading was a CVS experiment, a percentage similar to the overall CVS performance. Another relevant aspect of the exhibit design, is the exibit's limitation to three cylinder variables (mass, size, mass

distribution), and therefore to six possible experiments. This choice was meant to prevent families from getting bogged down in endless possibilities, but might have increased CVS performance (Chinn & Malhotra, 2002b; Kuhn, Amsel, & O'Loughlin, 1988). A third explanation for the above chance use of CVS in this study could be the adult presence in the family context. Adults investigating by themselves are more focused on informative experiments than children (Klahr, Fay, & Dunbar, 1993), and are better able to design these (Kuhn et al., 1988; Schauble, 1996). Children exploring together with their parents, as in this study, are more likely to perform CVS experiments compared to children exploring alone (Gleason & Schauble, 1999).

To conclude, the current study shows that the families were able to perform CVS experiments for a range of variables when investigating at an open-ended museum exhibit.. The families performed size, mass, and mass distribution experiments with similar frequencies. As concept learning from investigation starts with creating informative situations (Chen & Klahr, 1999; Stender, Schwichow, Zimmerman, & Härtig, 2018), these results show that the museum environment offers valuable learning opportunities.

Reasoning: Skills. All the families playing at the exhibit formulated hypotheses and gave causal explanations, and sixty percent of the families even gave causal explanations for minimally half of the performed experiments. Contrary to concerns that open-ended investigations will not lead to meaningful investigations (Gutwill, 2008), this study demonstrated high quality investigations. These findings are in line with those of Gutwill and Allen (2010), who also reported families formulating hypotheses and interpreting results during investigation at APE-exhibits. An important difference with Gutwill and Allen is that the families in their study were trained in inquiry skills, while the families in this study were not. A second finding concerning reasoning *skills* was that families more often mentioned causal explanations (50%) than hypotheses (28%), which aligns with previous results in field trip (Gutwill & Allen, 2012) and museum (Tenenbaum & Callanan, 2008) contexts. Allen (2002) has suggested that making verbal predictions takes too much effort. Our findings contrast with research showing that families do not always show advanced investigation strategies during unstructured exploration of interactive museum exhibits (Allen, 2002; Diamond, 1986; Randol, 2005). Possibly these contrasting results are explained by the way the visitor experience was framed in this study: by placing the exhibit in a context of scientific investigation through the exhibit labels and title (Hammer, Elby, Scherr, & Redish, 2005). Although most families only glanced at the exhibit labels, this may have been enough to set the stage, especially for the parents. Research into framing shows that a learning context can cue learners' expectations and consequently their cognitive activities and social interactions (Dunbar & Klahr, 1989; Friedman, 1979; Hammer et al., 2005), also in the specific context of family learning in a science museum (Atkins, Velez, Goudy, & Dunbar, 2009; Tscholl & Lindgren, 2016). A second explanation could be the so called 'immediate apprehendability' of the exhibit (Allen, 2004, p. 20). Because the rolling of cylinders refers to a familiar schema (i.e., racing), it helps to structure the activity (Rowe, 2002), and this allowed families to pay time and attention to the experiment

itself. Research showing a similar exhibit, *Downhill Race,* also provoking predictions and explanations (Gutwill, 2002; Perry & Tisdal, 2004) supports this explanation.

Reasoning: Content. In addition to domain general behaviors, such as formulating hypotheses and giving causal explanations, conversations about domain specific concepts can also be seen as indicators for learning from investigation. In this study, the families made content-related remarks about almost all experiments (i.e., 96%). Various topics were discussed, not only concerning the relationship between cylinder properties and acceleration, but also concerning the reliability of the performed experiments. Concerning reliability of the performed experiments, the families mentioned conditions that could cause or reduce experimental error, such as the simultaneous start of cylinders or performing the same experiment several times. Although errors are an inherent aspect of investigation, to the best of our knowledge, families' conversations about the reliability of performed experiments were not previously studied in a museum context. In the scientific reasoning literature, little attention has been paid to children's understanding of experimental error either (Masnick & Klahr, 2003; Schauble, 1996; Zimmerman, 2005). In line with our findings, Masnick and Klahr (2003) found that most 8- and 10-year-olds were able to generate sources for measurement and execution error. However, several studies demonstrated that children have difficulties distinguishing between variance due to error and variance as a consequence of phenomenon behavior (Masnick & Klahr, 2003; Schauble, 1996). Especially when an observed result is not in line with prior concepts, in a formal learning context, children are more likely to explain this in terms of error (Schauble, 1996). Possibly, these results can explain why families discussed the reliability of the performed experiments in the current study.

Most of the conversation was concerned with the relationship between cylinder properties and acceleration. That is, the majority (83%) of families' hypotheses and causal explanations were about cylinder properties. An example of a causal result interpretation is: "Youngest child (C_y): *No, at the same time!* Oldest child (C_o): *No, exactly at the same time!* Parent (P): *Look at that, at the same time.* C_y: *Same, because…, but I think it's because they are the same size* [keeps cylinders on top of each other] *and equally thick, these are really just the same, only this one* [heavy] *is the…, this is a heavier one.* P: *I expected that this iron one* [heavy] *would go down faster, because it is much heavier*. C_o: *Me too*. C_y: *Me too."*.

Talking about cylinder properties, the families mentioned mass more frequently (56%) than suggested by their CVS experiments (33% mass experiments). Their reasoning was often scientifically incorrect. As in the above example, in nearly half of the cases the families incorrectly mentioned a relationship between mass and acceleration. This finding contrasts with the findings of Gutwill and Allen (2010), who observed almost solely scientifically correct interpretations of results. Possibly, the inquiry training in the Gutwill and Allen study resulted in more scientifically correct conversations. However, there were more differences between the two studies. Although both exhibits were open-ended, had multiple options, and could be used by multiple users, the phenomena differed between exhibits. The exhibit in Gutwill and Allen was about connected pendula. The authors do not mention any dominant misconceptions about

the observable misconception, which was a striking characteristic of the exhibit in our study. Hence, another explanation for the relatively high frequency of scientifically incorrect conversations in this study could be that families' prior conceptions impeded them to accurately observe and interpret results. Research has shown that children struggle with accurately observing and interpreting situations when the observations contrast with their prior ideas (Chinn & Malhotra, 2002a, as cited in Morris et al. 2012). Children (Galili, 2001; Hast & Howe, 2017) and even physics students (Gunstone & White, 1981) have been shown to possess the misconception that mass explains inclined object motion. Children and adults have been shown to rely especially on their prior conceptions when experimental results are ambiguous, that is, when experimental outcomes are unclear (Schauble, 1996). In the current study, families encountered a substantial amount of ambiguous data, mostly in mass and size CVS experiments, where small differences due to error changed the conclusion from "cylinders arriving simultaneously" to "one cylinder is faster than the other". If at the current study's exhibit the heavy cylinder would beat the lighter green cylinder, this outcome would agree with the common object motion misconception. If the cylinders would arrive at the same time, the result could be explained by experimental error. A further microanalysis of how ambiguity of experimental outcomes (compared to experiments with only clear experimental outcomes) affects families' manipulations and conversations would be an interesting direction for future museum research aimed at painting a more realistic picture of science in the museum (Chinn & Hmelo-Silver, 2002; Metz, 2004).To conclude, in the current study, the families' concept learning from investigation left room for improvement. However, the open-ended investigation did trigger the families to discuss the topic of reliability of the performed experiments, which is an important, and not frequently investigated aspect of learning through scientific investigation.

3.4.2. The Impact of Individual Characteristics

The second research question studied the impact of the individual characteristics of the parent, oldest and youngest child on the family's investigation quality.

Children. We found that child characteristics were reflected in the manipulations and the reasoning *content*. With a medium effect size, the oldest children's reasoning capacity positively influenced the quality of the performed experiments (i.e., CVS). This is in line with a previous study demonstrating that children's verbal reasoning and vocabulary could moderately explain their CVS acquisition from self-directed learning in a controlled setting (Wagensveld, Segers, Kleemans &Verhoeven, 2015). This is particularly interesting, because a 3-year longitudinal study has demonstrated that children's CVS knowledge is a predictor for learning science (Bryant, Nunes, Hillier, Gilroy & Barros, 2015). Surprisingly, we found no effect of child characteristics on reasoning *skills*: the families' hypotheses and causal explanations. Within the age range participating in this study (8- to 14-year-olds), an effect of age could be expected, since children younger than ten rarely formulate explicit hypotheses

when performing experiments (Penner & Klahr, 1996), and the capacity for formulating hypotheses and interpreting results increases with age (Dunbar & Klahr, 1989; Klahr et al., 1993; Penner & Klahr, 1996; Schauble, 1996; Zimmerman, 2000, as cited in Haden, 2010). An explanation could be that the variance in reasoning skills was already explained by reasoning capacity, as this increases with age. Another explanation for not finding age effects could be the family context. Possibly the parents and oldest children (average 11 year old) took the lead in formulating hypotheses and interpreting results, and covered up a possible lack of contribution of the youngest children (average 9 year old). Looking at reasoning *content*, it is again the oldest child's reasoning capacity that emerges as a predictor. This child characteristic affected families' scientifically correct conversation about the relation of mass distribution and object motion with a small effect size. Evidence for effects of the oldest child's reasoning capacity on scientifically *in*correct conversations were more limited.

Parents. The effect of parent characteristics was only reflected in reasoning *content* about the reliability of performed experiments. With a large effect size, it was found that higher educated parents, i.e. parents with an academic education, more often mentioned a possible slope difference between the two roller tracks. In contrast, Tenenbaum and Callanan (2008) found that parents' education positively contributed to the reasoning *skills* of parent-child science conversation (e.g., causal result interpretations), but they did not analyze the reasoning *content* of the conversation. However, their study differed in various ways from the current study. For example, it was not aimed at open-ended exhibits and parents had less education.

Overall, the impact of individual parents' and children's characteristics on the families' investigation was smaller than expected. Besides the families' cognitive abilities, no other characteristics significantly affected the quality of the performed experiments (manipulation), and the topics discussed (reasoning *content*). The reasoning *skills* were also not significantly affected by individual characteristics. Thus, moments of high-quality open-ended investigation during the families' unguided visits were not limited to solely highly educated visitors, but applied to all participating families.

3.4.3. The Impact of Minimal Guidance Strategies

The last research question investigated whether minimal interventions by museum educators could positively contribute to the quality of families' investigation at an open-ended museum exhibit. Guidance was always given in response to the families' manipulations and reasoning *content*. With a large effect size, it was found that both guidance strategies, Giving Explanations and Describing Evidence, had a positive effect on the reasoning *content* of the families' conversation, but not on the reasoning *skills*. This finding is partly in line with those of Pattison et al (2018) who demonstrated that facilitation by museum educators positively impacted on families' mathematical reasoning. In the Pattison et al. study the overall mathematical reasoning measure was an average of four dimensions, two of which concerning verbal indicators of mathematical reasoning (talking about mathematical quantities and describing mathematical

relationships among those quantities) and two concerning visitor behaviors and interactions with the exhibits. The effect of facilitation in the current study, concerned families less often incorrectly concluding that acceleration differences were caused by mass differences. This might be expected for a Giving Explanations strategy, but is remarkable for a Describing Evidence strategy. In the latter case, the educator only remarked about the experiments the families performed themselves, without influencing the experiments or questioning the correctness of the families' reasoning. That is, the autonomy of the family was fully respected. In contrast, when applying a Giving Explanations strategy, the educator explained to the families that the slowing down or acceleration of a cylinder is caused by mass distribution differences, and not by mass or size differences. Notably, no clear evidence was found for guidance also stimulating families to make fewer incorrect remarks about size, or more correct remarks about mass (i.e., mass does not matter) or mass distribution. Also in literature it appears that negating a strong misconception is found to be easier than mentioning alternative explanations (Brookman-Byrne, Mareshal, Tolmie, & Dumontheil, 2018; Mareschal, 2016).

With regard to reasoning *content* about the reliability of performed experiments, evidence was found that Describing Evidence, but not Giving Explanations, resulted in a decrease in remarks (i.e., PAIR-I). Possibly, families guided by Describing Evidence were triggered to further discuss alternative cylinder properties mentioned by the educator, while families guided by Giving Explanations dealt with discrepancies between the educators' explanations and observed results by discussing the reliability of performed experiments.

3.4.4 Limitations of the Study

When quantifying families' learning through investigation in the museum context, apart from determining when, what, and how to measure, there are other considerations to take into account. First, families expect certain qualities of museum activities (e.g., they are fun to do). Therefore, participating in a research activity should be rewarding in itself and not evoke a feeling of failure (Bell, Lewenstein, Shouse, & Feder, 2009; Campbell 2008). A second consideration is the dilemma of ecological validity versus experimental control. Comparing two families that visit the science museum at the same time, one engaging in an unstructured visit and the other participating in the study, these families have differences in freedom of choice regarding exhibit choice, time spent and social context that impact the ability to generalize research findings (Bamberger & Tal, 2007; Falk, Koke, Price, & Pattison, 2018). In the current design we tried to optimize the families' free choice within the measurement constraints (especially active consent for videotaping; Pattison, Gutwill, Auster, & Cannady, 2019). The exhibit had the same size and quality as the other museum offer, and was not made especially for this study. The museum was busy and noisy, and the surrounding exhibits were occupied by other visitors. At the exhibit, the families were free to decide how and what to investigate without the presence of the test leader. Any guidance was given by regular museum educators. There were reasons why visitors in the research context may have been less interested in the

content of the exhibit than visitors who started playing on their own. Most parents (81 of 104) indicated that they participated in the study because research is a fun and an interesting activity for their children (and themselves). A second motive was to help make the museum more fun and educational (29 of 104), or the valuing of science learning research (20 of 104). However, it can also be said that the participants were more engaged because they were not disturbed by other visitors, and were asked to play for eight minutes. This commitment possibly had a positive effect on the families' interaction and attention. Pattison and Shagott (2015) showed that active recruitment (as in this study) in comparison to passive recruitment, can result in increased engagement time and exhibit behaviors, and even affects learning. The ability to generalize this study's results to other settings remains an open question. We agree with Pattison et al. (2017) that multiple studies in different settings are needed. This study contributes by showing that high quality investigation at an open-ended exhibit is within reach for museum visitors, that minimal guidance can make a substantial difference, and that several individual factors play a role. We do not intend to conclude that high quality investigations occur all the time during regular museum visits.

Another limitation of this study is its focus on a single exhibit. We purposefully did so, to perform an in-depth study of families' investigations with detailed manipulations and conversations. This is not uncommon in the visitor study literature (e.g., Benjamin et al., 2010; Crowley et al., 2001a; Van Schijndel & Raijmakers, 2016). To facilitate generalization of the results, we selected an open, interactive exhibit that not only addresses a common misconception, but also has many similarities with other Active Prolonged Engagement exhibits (Humphrey & Gutwill, 2005). Despite these similarities, we acknowledge the challenges in the generalization of results across exhibits and contexts. However, since the effects of the applied guidance strategies in this study have previously been demonstrated at other exhibits (Fender & Crowley, 2007; Van Schijndel & Raijmakers, 2016), we think this study makes a valuable contribution to the knowledge base on effective guidance in a museum context.

Choosing to observe families in a naturalistic setting, means accepting that data will be noisy and the study's power will be less, limiting possible findings (Bronfenbrenner, 1979). Nevertheless, we did obtain medium to large effect sizes (Cohen, 1988) for the regression models explaining the families' manipulation and reasoning *content*.

Noise may also have played a role in assessing children's working memory. Although the procedure for administering the digit span task went well, we found deviant results compared to standardized Dutch averages (Kort at al., 2005). Despite absolute values not being representative, individual differences within the study resulted in sufficient variance to use relative values. Yet, it is difficult to say what exactly has been measured by the digit span task: working memory or selective focussed attention (Diamond, 2013). Possibly tiredness caused by the exhibit experience and the demanding nature of the task may have played a role in the below-average scores.

Besides experimental control and ecological validity, we aimed for a pleasant visitor experience. Therefore, to prevent the procedure to last longer than half an hour, and to avoid families from feeling tested or incompetent, we decided not to include a knowledge post-test, in addition to quantifying reasoning *content* in-situ. In future studies, it could be informative to use both *content* measures, to investigate how reasoning *content* and knowledge acquisition are related.

3.4.5 Implications for Museum Practice

Families' learning through investigation in practice. The finding that high quality investigation at an open-ended exhibit during unstructured family visits can occur, does not mean that it will always occur. Firstly, the museum's offer is generally far more extensive than what visitors can do during a visit. Therefore, visitors only pay attention to a selection of the exhibits (Allen & Gutwill, 2009; Diamond 1986; Falk 2006; Falk & Dierking, 1992; 2000). Secondly, visitors are not expected to investigate all exhibits with the same intensity. Scientific investigation requires high levels of concentration, and to keep a balance, moments of low and high intensity will alternate during a visit. However, for museums it is valuable to know that high quality investigation during unstructured family visits can occur and lies within reach of all visitors. This will help to make informed choices when assembling a balanced and varied museum offer, especially when aiming to present a more realistic image of science in the museum (Chinn & Hmelo-Silver, 2002; Metz, 2004; O'Neill & Polman, 2004, as cited in Zimmerman, 2005). One way to present a more realistic image of science, is by addressing the nature of scientific knowledge, and with that the subject of uncertainty in relation to scientific evidence. We have shown that measurement error is a subject of families' conversations during investigation. We can imagine that the design of an exhibit can trigger families to do so to a greater or lesser extent. The open-ended exhibit used in the current study was designed in such a way that measurement error could arise. We presume that talk about reliability is less provoked by exhibits that are less open-ended. For example, an exhibit with only one handle instead of an exhibit with two handles that can be operated separately, as was the case in the exhibit used in the current study. An exhibit with only one handle ensures that both cylinders start rolling down at the same time, and with that reduces measurement error. Or, an exhibit that has only two cylinders (a compact and a hollow) that differ considerably in rolling acceleration, compared to multiple cylinders among which cylinders with the same rolling acceleration. Measurement error has less impact if it is many times smaller than the observed effect, as is the case with two cylinders that differ from each other in rolling acceleration (a compact and a hollow cylinder).

Minimal guidance strategies in practice. The current study showed that both minimal guidance strategies can positively, albeit modestly, contribute to families' scientific investigation at an open-ended museum exhibit. An important difference between the two interventions is that Giving Explanations provides what the outcome of the experiment should

be, while Describing Evidence does not. Possibly the explanations given by the educator do not match with the families' observations in following experiments due to measurement or execution errors. This discrepancy could stimulate families to reason about alternative explanations, but it could also give families a feeling of incompetence (cf., Gutwill, 2008). Moreover, by Giving Explanations the educator takes the role of an expert and the educational aim of the families' investigation could shift from investigating to find out how it works, to investigating how to reproduce the theory (Hein, 1998). In contrast, with Describing Evidence the educator draws attention to cylinder properties that families might not have considered before, and therefore the educational aim of families' investigation remains to find out how object motion works.

In the current study, we trained and asked museum educators to apply one guiding style in isolation, in order to be able to investigate its effectiveness. Although the educators were fully capable of doing so, this is not how we propose to use interventions in museum practice. We imagine educators to have a range of interventions at their disposal, which they can use to respond to families' needs. However, as a museum educator, experiencing the effect of applying a single intervention at a time can yield interesting insights. Following the current research, NEMO Science Museum developed a basic training for educators (Interaction: Visitor – Exhibit – Educator, 2017) in which three intervention strategies were discussed and practiced at multiple exhibits: Asking Questions, Giving Explanations, and Describing Evidence. Although not scientifically studied, over 45 museum educators participated in this training and educators indicated that 1) practicing multiple intervention strategies made them more aware of their own preferred guiding style (e.g., giving explanations), 2) experiencing interaction differences within the visitor-exhibit-intervention triangle was informative, and 3) Describing Evidence is an intervention style they had not commonly used before. It would be interesting to further investigate their guidance in practice and study the effectiveness of the intervention strategies with different types of exhibits.

3.5 CONCLUSION

A frequently mentioned consequence of open-ended investigation, in formal and informal science learning context, is that it hardly facilitates learning. In this science museum study we conclude, based on the causal talk and informative manipulations, that the families, even without guidance were able to investigate the open-ended exhibit in a meaningful way. Families controlled their experiments and formulated hypotheses and causal explanations. Families' did not only discuss the phenomenon (object motion) during the open-ended investigation, but also had a meta-conversation about performing experiments. Minimal interventions by the museum educators created better opportunities for learning. Both giving explanations and describing the observed evidence (as separate guidance strategies) specifically reduced the families' confirmation of a common misconception during investigation. Individual differences in

children's cognitive abilities and parents' educational background only modestly impacted the families' performance.

3.6 APPENDIX A

Table 3.A1
Descriptions of variables and glossary of acronyms

Abbreviation	Description	Variable
	Manipulation	
E	Number of experiments performed	
E_{CVS}	Number of CVS experiments performed	
pE_{CVS}	Proportion CVS experiments (E_{CVS} divided by E)	Dependent
	Reasoning: Skills	
PA	Number of experiments with high level Proposing Actions (i.e., formulating hypotheses)	
IR	Number of experiments with high level Interpreting Results (i.e., giving causal explanations)	
pPA	Proportion PA (PA divided by E)	Dependent
pIR	Proportion IR (IR divided by E)	Dependent
	Reasoning: Content	
PA-D_c	Total number of D_c mentioned in PAs (scientifically correct PA about the causal relation between mass distribution and acceleration)	Dependent
PA-M_c	Total number of M_c mentioned in PAs (scientifically correct PA about the causal relation between mass and acceleration)	Dependent
PA-M_i	Total number of M_i mentioned in PAs (scientifically incorrect PA about the causal relation between mass and acceleration)	Dependent
PA-S_i	Total number of S_i mentioned in PAs (scientifically incorrect PA about the causal relation between volume and acceleration)	Dependent
PA-X	Total number of X mentioned in PAs (other PA about the causal relation of cylinder characteristics and acceleration)	Dependent
IR-D_c	Total number of D_c mentioned in IRs (scientifically correct IR about the causal relation between mass distribution and acceleration)	Dependent
IR-M_c	Total number of M_c mentioned in IRs (scientifically correct IR about the causal relation between mass and acceleration)	Dependent

IR-M_i	Total number of M_i mentioned in IRs (scientifically incorrect IR about the causal relation between mass and acceleration)	Dependent
IR-S_i	Total number of S_i mentioned in IRs (scientifically incorrect IR about the causal relation between volume and acceleration)	Dependent
IR-X	Total number of X mentioned in IRs (other IR about the causal relation of cylinder characteristics and acceleration)	Dependent
F	Equal rolling conditions	
I	Equal inclined plane conditions	
PAIR-F	Total number F in PA and IR (sum of PA and IR about equal rolling conditions)	Dependent
PAIR-I	Total number I is mentioned in PAs and IRs (sum of PA and IR about inclined plane conditions)	Dependent
	Guidance strategy	
ED	Describing evidence guidance condition (n=25)	
EX	Giving explanation guidance condition (n=27)	
C	Without guidance control condition (n=52)	
ED_C	ED compared to C	Factor
EX_C	EX compared to C	Factor
	Individual characteristics	
Age, P	Parent's age	Co-variate
Gender, P	Parent's gender	Co-variate
Education L	Up to Bachelor's degree parents compared to Graduate degree parents	Co-variate
Education B	Bachelor's degree parents compared to Graduate parents	Co-variate
LB-A	Parent's belief about learning: Learning is an active process	Co-variate
Sc. Int.	Parent's interest in science	Co-variate
Age, C_O	Oldest child's age	Co-variate
Gender, C_O	Oldest child's gender	Co-variate
IQ-R, C_O	Oldest child's reasoning capacity: sum of raw scores on the WISC III-subtests Similarities and Block Design	Co-variate
IQ-WM, C_O	Oldest child's working memory capacity: raw scores on WISC III-subtest Digit Span Backwards.	Co-variate
Sc. Enj., C_O	Oldest child's science enjoyment: sum score on VTB	Co-variate
Age, C_Y	Youngest child's age	Co-variate
Gender, C_Y	Youngest child' gender	Co-variate

IQ-R, C_Y	Youngest child' reasoning capacity: sum of raw scores on two WISC III-subtests: Similarities and Block Design.	Co-variate
IQ-WM, C_Y	Youngest child' working memory capacity: raw scores on WISC III-subtest Digit Span Backwards.	Co-variate
Sc. Enj., C_Y	Youngest child' science enjoyment: sum score on VTB	Co-variate
Education	Parent's highest education level	
WISC	Wechsler Intelligence Scale for Children (WISC)	
Digit Span Backwards	WISC III-subtest within the Working Memory Index	
Similarities	WISC III-subtest within the Verbal Comprehension Index	
Block Design	WISC III-subtest within the Perceptual Reasoning Index	

3.7 SUPPLEMENTARY MATERIAL

3.7.1 Materials

Post-test. Families' individual characteristics have been measured through questionnaires (adult and children) and tasks (children), in a post-test session. Below we report additional information about the internal consistency of the questionnaires that assessed parents' perspective on learning and children's science enjoyment.

How children learn inventory (adult). Adults' beliefs about learning (i.e., learning as an active or passive process) was assessed using a 16-statement questionnaire (How Children Learn Inventory; Ricco & Rodriquez, 2006). Adults rated their agreement with the statements on a 5-point-Likert scale (1 = strongly disagree to 5 = strongly agree). An example statement is: "When it comes to math and science, children learn by doing, i.e., through hands-on experience and by trying out ideas". The measure was translated from English into Dutch. Ricco and Rodriguez (2006) reported a forced two-factor analysis under Varimax rotation resulting in a factor (factor 1) referring to the belief that learning is an active process (α = .63; statements 5, 7, 9, 10), and a factor (factor 2) referring to the belief that learning is a passive process (α = .58; statements 1, 4, 8, 12, 13, 14), each explaining 15% of the variance. For the current study we found non-reliable internal consistency of Ricco and Rodriguez' factor 1 (α = .51; statements 5, 7, 9, 10). A factor analysis with data of the current study resulted in a slightly different factor (factor LB-A) referring to learning is an active process (statement 15 "Children learn math and science through experiencing success and praise or approval from adults" instead of statement 10 "No two children learn math and science in quite the same way") with a higher internal consistency reliability (α = .64). Sum scores of statements 5, 7, 9 and 15 will be used in further analysis.

Science enjoyment (children). Children's science enjoyment was assessed using a subscale (7 statements) of the Dutch science and technology attitude instrument for primary school pupils (Walma van der Molen, ey al., 2007). Children rated their agreement with statements on a 4-point-Likert scale (1 = strongly disagree to 4 = strongly agree). Example questions are: "I like to explore things", "I like to learn more about science". The internal consistency of the subscale in the current study (α = .63) is lower than Van der Molen et al. (α = .88). This could be explained by a difference in variance: compared to Van der Molen et al. study (primary school pupils, M=2.69), children in the current study are more alike (science museum visitors), and have significantly higher average *Science Enjoyment* (M = 3.31, SD = .42), $t(199) = 20.891$, $p < .001$.

Coding approach. The quality of families' conversation during investigation was analysed using two conversation coding instruments (i.e., reasoning *skills* and reasoning *content*). Below the scoring rules of the reasoning *skills* coding instrument are elaborated.

Reasoning: Skills. To assess the reasoning *skills* quality we followed Gutwill and Allen's (2010) approach of classifying difficulty levels of proposing actions and interpreting results. Gutwill and Allen's high-level proposing action is defined as families mentioning both a proposed action and the expected effect of this action. In a similar way, high-level interpreting results were defined as families mentioning both the observed effect as a possible explanation for this effect.

In the current study, a *Rolling cylinders* conversation *prior* to an experiment was classified as PA (PA = 1) if family members, alone or together, formulated a hypothesis. That is, they not only predicted what would happen (e.g., A will arrive later than B) but also elaborated on why this would happen (i.e., because B is heavier than A). A *Rolling cylinders* conversation was classified as non-PA (PA = 0) if family members only predicted, only mentioned a cylinder characteristic, or said nothing. In a similar way a *Rolling cylinders* conversation *afterwards* an experiment was classified as IR (IR = 1) if family members, alone or together, formulated causal explanations. That is, they not only described the result (e.g., A arrived earlier than D) but also explained why this happened (i.e., because D is bigger than A). A *Rolling cylinders* conversation was classified as non-IR (IR = 0) if family members only mentioned the result, only mentioned a cylinder characteristic of said nothing.

In the current study, a *Reliable outcomes* conversation *prior* to an experiment was classified as PA (PA = 1) if family members, alone or together, justified an action or the use of a rule related to measurement or performance errors. That is, they not only mentioned an action or rule (e.g., starting the cylinders simultaneously), but also elaborated the reason for this action or rule (e.g., to be able to compare the cylinders). A *Reliable outcomes* conversation was classified as non-PA (PA = 0) if family members only mentioned the rule, only mentioned the aim (e.g., comparing cylinders), or said nothing. In a similar way, the *Reliable outcomes* conversation *afterwards* an experiment was classified as IR (IR = 1) if family members, alone or together, formulated a causal explanation. That is, they not only described the action or rule (e.g., rolling down in a straight line) but also explained the consequence of this action or rule

in relation to measurement or performance errors (i.e., otherwise we cannot compare the cylinders). A *Reliable outcomes* conversation was classified as non-IR (IR = 0) if family members only mentioned the action or rule, only the aim or motivation, or said nothing.

Analyses approach. Using, the reasoning *content* coding instrument resulted in 14 outcome variables describing the quality of families' reasoning *content* during investigation. Ten describing *Rolling cylinders* (i.e., PA-D_c, PA-M_c, PA-M_i, PA-S_i, PA-X, IR-D_c, IR-M_c, IR-M_i, IR-S_i, IR-X), and four describing *Reliable outcomes* (i.e., PA-F, PA-I, IR-F, IR-I). To identify if this large amount of variables could be reduced Principal components analyses were performed. First the four *Reliable outcomes* variables were analysed. Initial eigen values indicated that the first factor explained 48% of the variance. The second, third and fourth factor had eigen values below one, and explained 23%, 8% and 3% of the variance respectively. The Kaiser-Meyer-Olkin measure of sampling adequacy was .647, which is above the commonly recommended value of .6, also Bartlett's test of sphericity was significant ($\chi 2$ (6) = 53.139, $p <$ .005), the diagonals of the anti-image correlation matrix were around .5 (i.e. r ranges from .620 to .691), and the communalities were all well above .3. The solution for two factors was examined using Oblimin rotations of the factor loading matrix. One factor describes comments about equal inclined plane condition (i.e., PA-I, IR-I), the other about equal rolling conditions (i.e., PA-F, IR-F). Absolute factor loading were all above .713. Adding PA-I and IR-I values, and PA-F and IR-F values, seems useful, since factor loadings are in the same direction. Therefore sum scores (PAIR-I and PAIR-F) have been used in a multiple regression analysis, to see whether reasoning *content* quality could be predicted by person characteristics and guiding style.

Subsequently, it was investigated if the ten *Rolling cylinders* outcome variables could be reduced by Principal components analysis. Both a 4-factor solution (61% explained variance and an eigenvalue of 1.132) and a 5-factor solution (71% explained variance and an eigenvalue of 0.956) resulted in two PA factors that made sense. One factor (factor 1) comprised physically incorrect PAs (e.g., M_i, S_i, X, see also Coding instruments in Method section) and another factor (factor 2) comprises physically correct PA (e.g., D_c, M_c). The absolute factor loads are 0.656 and 0.689 respectively. For a 6-factor (79%) solution, too, two factors could be identified, which represented incorrect and correct hypotheses, however the limit of the absolute factor load for the incorrect hypotheses is 0.466, below 0.5. IR variables showed a less clear picture. Although a 4-factor solution showed a factor with incorrect IR (e.g., M_i and S_i), a second factor have high factor loadings for D_c and X, but M_c was not clearly represented in any factor. A 5-factor solution presents three factors each representing one category (i.e. M_i, S_i and X). Now neither M_c nor D_c were represented. Because the factor analyzes of PA and IR did not produce the same picture, we choose not to reduce these variables. Therefore all twelve *Rolling cylinders* variables have been used in the multiple regression analysis, to see whether reasoning *content* quality can be predicted by person characteristics and guiding style.