

PROMOTING AND MEASURING
ELEMENTARY SCHOOL
CHILDREN'S UNDERSTANDING
OF SCIENCE

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Dipl.-Psych. Julia Schiefer
aus Stuttgart - Bad Cannstatt

Tübingen

2016

Tag der mündlichen Prüfung: 27.01.2017
Dekan: Prof. Dr. rer. soc. Josef Schmid
1. Gutachterin: Prof. Dr. Kerstin Oschatz
2. Gutachter: Prof. Dr. Ulrich Trautwein

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisors Prof. Dr. Kerstin Oschatz and Prof. Dr. Ulrich Trautwein from the Hector Research Institute of Education Sciences and Psychology (HIB) for their continuous support and their foresighted guidance throughout this dissertation project. Their different scientific backgrounds and research approaches were both challenging and highly enriching for me. I learned a lot from their respective perspectives, and I greatly appreciate that I had the opportunity to work scientifically with them. Next, I would like to cordially thank Dr. Jessika Golle, who took over a large part of my supervision and addressed all of my concerns about this dissertation. Her feedback was always very valuable for me. I would also like to thank Dr. Maike Tibus for her help and critical suggestions regarding this dissertation project.

This dissertation is part of the great team work that has transpired in the project *Formative Assessment of the Hector Children's Academy Program*. I would like to thank all my team colleagues—in particular Evelin Herbein—for their cooperation, helpfulness, and support in this project. Furthermore, I would like to thank our student research assistants who helped me a lot with the empirical surveys, data entry, as well as data analysis. I would also like to thank the Hector Foundation II, who funded this work in part, as well as the academies of the Hector Children's Academy Program and the elementary school classes that participated in my empirical studies.

Furthermore, I would like to thank my colleagues from the Hector Research Institute as well as the LEAD Graduate School and Research Network. As my dissertation combines perspectives from education science, natural science education, and psychology, I benefitted a lot from the interdisciplinary approach and the inspiring and constructive work atmosphere. The LEAD Graduate School also offered me the opportunities to meet international experts such as Deanna Kuhn and to present my research at many national and international conferences. Special thanks go to Dr. Norman Rose from the HIB and Dr. Johann Jacoby from the LEADing Research Center for their methodological advice regarding this dissertation project.

Finally, I would like to thank my beloved partner, my family, and my friends for their great support and encouragement and for being there for me when I needed them.

ABSTRACT

Science and scientific knowledge are important parts of our culture and play essential roles in our everyday lives (Bybee, 1997; OECD, 2016). To be able to participate in socioscientific discussions, it is essential to have not only knowledge and skills in STEM (Science, Technology, Engineering, and Mathematics) but also an adequate understanding of the *nature of science* (e.g., Driver, Leach, Miller, & Scott, 1996). An understanding of the nature of science (for reasons of better legibility, we refer to this as an *understanding of science* in the following) refers to an understanding of “what science is and how it is done” (McComas, 1998, p. 50). It lays therefore an important foundation for students’ science learning (Lederman, 2007). As particularly important elements, an understanding of science includes epistemic beliefs (individual representations about knowledge and knowing) as well as an understanding of inquiry-based methods (approaches under which scientific knowledge is generated).

Due to the essential relevance of an adequate understanding of science, promoting such an understanding in students is a normative goal of science education, and for many years, educational research and practice has focused on promoting this understanding as early as elementary school (e.g., European Commission, 2007; Mullis & Martin, 2015; OECD, 2016). Interventions offer one effective way of investigating and promoting students’ understanding of science. Previous intervention studies have shown that inquiry-based approaches, in particular, can be beneficial for fostering students’ understanding of science (e.g., Blanchard et al., 2010). However, a number of questions regarding the effective promotion of elementary school children’s understanding of science are still unanswered. They refer, for example, to the question of how fundamental aspects of students’ understanding of science (e.g., epistemic beliefs and a profound understanding of scientific inquiry methods) can be promoted effectively in elementary school children. Furthermore, open questions exist with regard to how to adequately measure students’ understanding of science. Instruments are required to describe children’s competencies and to measure intervention effects. However, existing instruments for elementary school children cover limited aspects of students’ understanding of science and show somewhat limited reliability and validity.

The three empirical studies that were conducted in this dissertation addressed central questions concerning the measurement and promotion of elementary school children's understanding of science. Specifically, the dissertation focused on (a) the development of a new paper-and-pencil test for assessing elementary school children's understanding of science and (b) the investigation of the effectiveness of a recently developed intervention for third and fourth graders.

With cross-sectional data from 878 third and fourth graders, Study 1 examined the reliability and validity of a new instrument that was developed to measure the understanding of the so-called *scientific inquiry cycle (SIC)* as a central component of the understanding of science. Confirmatory factor analyses confirmed a one-dimensional structure of the test, and the instrument was found to have an acceptable reliability. As expected, the SIC was found to be positively related to cognitive abilities such as fluid intelligence and text comprehension, experimentation strategies, as well as epistemic beliefs.

Studies 2 and 3 investigated the effectiveness of a 10-week extracurricular intervention with regard to the promotion of elementary school children's understanding of science by means of two randomized controlled studies. The intervention was developed by researchers from the university as part of an enrichment program for gifted children (*Hector Children's Academy Program, HCAP*). It focused on the targeted promotion of children's understanding of science and included inquiry-based learning approaches, the ability to work scientifically according to the SIC, as well as reflections on epistemic issues. The results of Study 2 ($N = 65$)—in which the intervention was conducted by the program developers under controlled conditions—revealed that the intervention affects children's epistemic beliefs and epistemic curiosity positively. On the basis of the positive results of Study 2, Study 3 ($N = 117$) investigated the effectiveness of the intervention when it was implemented under real-world conditions by 10 course instructors from the HCAP. In this context, the SIC test was applied to examine the intervention effects. Positive effects were found on children's understanding of science (understanding of the SIC and experimentation strategies) and need for cognition. Intervention effects on epistemic beliefs and epistemic curiosity could not be replicated. Analyses of implementation fidelity revealed that, overall, the course instructors kept to the program and put the intervention—with some limitations—into practice successfully.

The findings of the three studies are summarized and discussed within the broader research context. Implications for future research and educational practice are derived.

ZUSAMMENFASSUNG

Naturwissenschaften und wissenschaftliche Erkenntnisse sind ein wichtiger Bestandteil unserer Kultur und spielen eine entscheidende Rolle in unserem täglichen Leben (Bybee, 1997; OECD, 2016). Um sich an Diskussionen zu gesellschaftswissenschaftlichen Fragen beteiligen zu können, sind nicht nur Wissen und Kenntnisse in den MINT-Fächern (Mathematik, Informatik, Naturwissenschaften und Technik), sondern ein angemessenes Wissenschaftsverständnis entscheidend (z.B. Driver, Leach, Millar, & Scott, 1996). Dieses beinhaltet ein Verständnis dafür, „was Naturwissenschaften sind und wie sie betrieben werden“ (McComas, 1998, p. 50) und bildet eine wichtige Voraussetzung für das Lernen naturwissenschaftlicher Inhalte (Lederman, 2007). Besonders entscheidende Elemente des Wissenschaftsverständnisses sind sowohl epistemische Überzeugungen (individuelle Vorstellungen über die Natur des Wissens und des Wissenserwerbs), als auch ein grundlegendes Verständnis für die naturwissenschaftlich-forschenden Methoden, mittels derer naturwissenschaftliches Wissen generiert wird.

Die Förderung eines angemessenen Wissenschaftsverständnisses von Schülerinnen und Schülern ist seit einigen Jahren ein normatives Bildungsziel im Bereich des naturwissenschaftlichen Lernens und im Fokus der empirischen Bildungsforschung und Bildungspraxis, bereits schon bei Kindern in der Grundschule (z.B. European Commission, 2007; Mullis & Martin, 2015; OECD, 2016). Interventionen bieten eine effektive Möglichkeit, um das Wissenschaftsverständnis von Schülerinnen und Schülern zu untersuchen und zu fördern. Bisherige Interventionen konnten den Nutzen von forschend-entdeckenden Ansätzen für die Förderung des Wissenschaftsverständnisses zeigen (z.B. Blanchard et al., 2010). Jedoch sind viele Fragen im Hinblick auf eine effektive Förderung des Wissenschaftsverständnisses noch nicht beantwortet. Diese beziehen sich zum Beispiel darauf, wie grundlegende Aspekte des Wissenschaftsverständnisses—z.B. epistemische Überzeugungen oder ein fundiertes Verständnis für naturwissenschaftliche Methoden—bei Grundschulkindern gezielt gefördert werden können. Zudem bestehen offene Fragen im Hinblick auf eine angemessene Erfassung des Wissenschaftsverständnisses von jüngeren Schülerinnen und Schülern. Instrumente werden benötigt, um die Kompetenzen von Kindern zu erfassen sowie um Interventionseffekte adäquat messen zu können. Bestehende Instrumente erfassen jedoch nur begrenzte Aspekte des Wissenschaftsverständnisses oder weisen teilweise eine unzureichende Zuverlässigkeit und Gültigkeit auf.

Die drei empirischen Studien, welche im Rahmen dieser Dissertation durchgeführt wurden, adressieren zentrale Fragen im Hinblick auf die Erfassung und Förderung des Wissenschaftsverständnisses von Grundschulkindern. Die Dissertation untersucht insbesondere (a) die Entwicklung eines neuen Papier-und-Bleistift Tests zur Erfassung des Wissenschaftsverständnisses bei Grundschulkindern sowie (b) die Effektivität einer neu entwickelten Intervention für Kinder der dritten und vierten Klasse.

Studie 1 untersuchte mittels Querschnittsdaten von 878 Kindern der dritten und vierten Klasse die Zuverlässigkeit und Gültigkeit eines neu entwickelten Instruments. Dieses wurde konzipiert um das Verständnis für den sogenannten „Wissenschaftszirkel“ als zentrales Element naturwissenschaftlich-forschender Methodik zu erfassen. Konfirmatorische Faktorenanalysen bestätigten die vermutete eindimensionale Struktur des Tests. Es zeigte sich eine angemessene Testzuverlässigkeit. Wie erwartet, fanden sich positive Zusammenhänge des Tests zu kognitiven Fähigkeiten wie fluider Intelligenz und Textverständnis, Experimentierstrategien sowie epistemischen Überzeugungen.

In den Studien 2 und 3 wurde die Wirksamkeit einer entwickelten 10-wöchigen Intervention im Hinblick auf die Förderung des Wissenschaftsverständnisses von Grundschulkindern mittels kontrolliert randomisierter Studien untersucht. Die Intervention wurde von Wissenschaftlerinnen entwickelt und war Teil eines außerunterrichtlichen Förderprogramms für besonders begabte und hochbegabte Kinder (*Hector-Kinderakademien, HKA*). Sie zielte auf die Förderung des Wissenschaftsverständnisses der Kinder und beinhaltete Elemente untersuchend-forschendes Lernens, wissenschaftlicher Arbeitsweisen nach dem Wissenschaftszirkel sowie Reflexionen über epistemische Fragen. Die Ergebnisse von Studie 2 ($N = 65$)—bei der die Intervention von den Entwicklerinnen des Programms unter kontrollierten Bedingungen durchgeführt wurde—zeigten positive Interventionseffekte auf die epistemischen Überzeugungen sowie die Wissbegierde der Kinder. Aufbauend auf diese positiven Ergebnisse wurde in Studie 3 ($N = 117$) die Wirksamkeit der Intervention unter Praxisbedingungen getestet, indem sie von zehn Kursleiterinnen und Kursleitern der Hector-Kinderakademien durchgeführt wurde. In diesem Zusammenhang wurde auch das neu entwickelte Instrument aus Studie 1 eingesetzt. Positive Interventionseffekte zeigten sich auf das Wissenschaftsverständnis der Kinder (ihr Verständnis für den Wissenschaftszirkel sowie für Experimentierstrategien) sowie ihre Freude am Denken. Die Interventionseffekte auf die epistemischen Überzeugungen und die Wissbegierde konnten nicht repliziert werden. Die Untersuchungen zur Manualtreue

zeigten, dass die Kursleiterinnen und Kursleiter die Elemente der Intervention nach Anleitung durchgeführt haben und diese—mit einigen Einschränkungen—erfolgreich in der Praxis umsetzen konnten.

Die Ergebnisse der drei Studien werden abschließend zusammengefasst und hinsichtlich ihrer Bedeutung für die Forschungslandschaft diskutiert. Im Anschluss daran werden Implikationen für weiterführende Untersuchungen sowie die Bildungspraxis abgeleitet.

TABLE OF CONTENTS

1. INTRODUCTION AND THEORETICAL FRAMEWORK	1
1.1. Theoretical Conceptualization of the Understanding of Science	7
1.1.1. Epistemic beliefs.....	10
1.1.2. Inquiry-based methods	12
1.1.3. Elementary school children’s understanding of science.....	14
1.1.4. Relations of the understanding of science to other constructs.....	17
1.2. Empirical Measurement of the Understanding of Science	21
1.2.1. Quality criteria for instruments.....	21
1.2.2. Existing instruments and their boundaries.....	22
1.3. Intervening in Students’ Understanding of Science	28
1.3.1. Interventions and their implementation	28
1.3.2. Intervention approaches to promote the understanding of science.....	30
1.3.3. Development of an extracurricular intervention.....	34
1.4. Research Questions of the Present Dissertation	41
2. STUDY 1: SCIENTIFIC REASONING IN ELEMENTARY SCHOOL CHILDREN: ASSESSMENT OF THE INQUIRY CYCLE	47
3. STUDY 2: FOSTERING EPISTEMIC BELIEFS, EPISTEMIC CURIOSITY, AND INVESTIGATIVE INTERESTS IN ELEMENTARY SCHOOL CHILDREN: A RANDOMIZED STEM INTERVENTION STUDY	79
4. STUDY 3: ELEMENTARY SCHOOL CHILDREN’S UNDERSTANDING OF SCIENCE: THE IMPLEMENTATION OF AN EXTRACURRICULAR SCIENCE INTERVENTION	113
5. GENERAL DISCUSSION	159
5.1. Discussion of General Findings	162
5.1.1. Measurement of the understanding of science	162
5.1.2. Effectiveness and implementation of the intervention	163
5.1.3. Strengths and limitations of the present dissertation	165
5.2. Implications and Future Directions	169
5.2.1. Implications for future research.....	169
5.1.1. Implications for educational policy and practice.....	173
References	176

1

Introduction and Theoretical Framework

1 Introduction and Theoretical Framework

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.”

Sir William Bragg, physicist (1862-1942)

The STEM (Science, Technology, Engineering, and Mathematics) disciplines are an important part of our culture and play essential roles in our everyday lives (Bybee, 1997; Organization for Economic Co-operation and Development [OECD], 2016). Science and scientific knowledge are important in various ways. Science impacts every individual, for instance, with respect to decisions that are based on scientific findings and developments, such as eating genetically modified food, using medicine for reproduction, or medicating ADHD in children (Chopyak & Levesque, 2002; OECD, 2016). STEM subjects also matter to society in terms of economy because there is a need to ensure that there will be enough qualified engineers and natural scientists in the younger generations in order to secure the economic future of society as an industrial location (e.g., Sawyer, 2008; Xue & Larson, 2015).

An understanding of the nature of science is the basis for everyday decision making and an understanding of the importance of science as a central element of our contemporary culture (Driver, Leach, Millar, & Scott, 1996). It also enables critical reflection on scientific knowledge and its boundaries (e.g., Driver et al., 1996; Lederman, 2007). Furthermore, an understanding of the nature of science influences students’ learning of scientific subject matter and their performance in science, which is a prerequisite for careers in STEM fields (e.g., Buehl & Alexander, 2005; Driver et al., 1996; Duschl, Schweingruber, & Shouse, 2007; Kuhn, 2005; Nussbaum, Sinatra, & Poliquin, 2008; OECD, 2016).

An understanding of the nature of science (for reasons of legibility, this will be referred to as an *understanding of science* in the following) refers to an understanding of “what science is and how it is done” (McComas, 1998, p. 50). As key components, epistemic beliefs (i.e., individual representations about knowledge and knowing; Hofer & Pintrich, 1997) and an understanding of inquiry-based methods (e.g., the cyclical process by which scientific knowledge is generated; Kuhn, 2002) are included.

Despite the enormous importance of the understanding of science in modern life, few individuals have a fundamental understanding of how the scientific enterprise operates (McComas, Almazroa, & Clough, 1998). For instance, many misconceptions or myths about science exist. They include assumptions about scientific ideas as absolute and unchanging, about science and its methods providing absolute proof, or about science as a collection of facts, or about scientists as being completely objective in their evaluation of scientific ideas and evidence (McComas, 1998; OECD, 2016). Also the latest results of international large-scale assessment studies, such as *TIMSS* (Trends in International Mathematics and Science Study) and *PISA* (Programme for International Student Assessment) demonstrate the need of promoting students' understanding of science (Martin, Mullis, Foy, & Hooper, 2015; OECD, 2016). For instance, only about 7% of the fourth grades had an understanding of the process of scientific inquiry (Martin et al., 2015).

Because of the relevance of the understanding of science and the existing misconceptions, the understanding of science has been suggested as one critical component of so-called *scientific literacy* (Laugksch, 2000), which stands for “what the general public ought to know about science” (Durant, 1993, p. 129). The promotion of students' understanding of science is one cornerstone of current educational research and practice as early as elementary school (e.g., Bendixen, 2016; Duschl et al., 2007; European Commission [EC], 2007; National Research Council [NRC], 2011; OECD, 2016).

In order to meet this educational goal, questions have been raised about the successful promotion of students' understanding of science (Bundesministerium für Bildung und Forschung [BMBF], 2013; Carnevale, Smith, & Melton, 2011; EC, 2007). Previous research has indicated that inquiry-based learning approaches in particular can foster students' understanding of science (Blanchard et al., 2010; Minner, Levy, & Century, 2009). However, there still is a lack of systematic research on effective ways to promote fundamental aspects of students' understanding of science—such as epistemic beliefs and a profound understanding of scientific inquiry methods—particularly in elementary school children (see Bendixen, 2016; Valla & Williams, 2012).

The present dissertation was aimed at closing this gap and addressing the central question of how young children's understanding of science can be fostered effectively. To reach this goal, an intervention was developed for elementary school children. Interventions provide an important way to foster students' competencies because specific promotion programs and instructional design principles can be systematically compared and

investigated (Humphrey et al., 2016). On this basis, effective programs can be developed and implemented in practice (Lendrum & Wigelsworth, 2013). A 10-week intervention that was developed and evaluated in this dissertation focused on the promotion of fundamental aspects of the understanding of science: (a) adequate conceptions about the nature of knowledge and knowing in science (epistemic beliefs; Hofer & Pintrich, 1997) and (b) inquiry-based methodological competencies. These include, for instance, an understanding of the cyclical and cumulative process that builds the basis for the genesis, construction, and development of scientific knowledge (the so-called *scientific inquiry cycle, SIC*; Kuhn, 2002; Zimmerman 2007). The target group of the intervention were elementary school children who participated in an extracurricular enrichment program (*Hector Children's Academy Program, HCAP*). Because enriched and gifted students have the potential to become future STEM leaders, their promotion in the STEM disciplines has a particular relevance for society and economy (National Science Board [NSB], 2010; Sawyer, 2008).

To be able to adequately investigate the effectiveness of the intervention, a new instrument was developed, scaled, and validated in a first study (Study 1). This was required as only a few paper-and-pencil tests for assessing students' understanding of science existed previously. The existing instruments covered limited aspects of the understanding of science and showed somewhat limited levels of reliability and validity (Mason, 2016). This applies in particular for instruments designed for elementary school children (see Mayer, Sodian, Koerber, & Schwippert, 2014). The new instrument focused on the assessment of children's understanding of the SIC as a central component of the understanding of science (Kuhn, 2002; White, Frederiksen, & Collins, 2009; Zimmerman, 2007).

Subsequently, the newly developed extracurricular intervention was investigated with regard to its effectiveness in promoting children's understanding of science (Studies 2 and 3). In Study 2, the intervention was conducted by the program developers under controlled conditions. On the basis of the positive results of Study 2, Study 3 explored whether the intervention was still effective when implemented under real-world conditions by HCAP course instructors. In this context, the newly developed instrument was applied to examine the effects of the intervention on students' understanding of the SIC.

The dissertation is structured as follows: The introductory Chapter 1 describes the theoretical background of the three empirical studies and aims to embed these studies

within a broader research context. The first section of the introduction (Chapter 1.1.) describes the conceptualization of the understanding of science. In this regard, the development of the understanding of science in elementary school age children and connections to related constructs are discussed. The second section (Chapter 1.2.) focuses on the measurement of the understanding of science. Requirements for testing instruments as well as existing measurement approaches and their boundaries are described. Chapter 1.3. focuses on intervention approaches with regard to students' understanding of science. In this regard, recommended approaches for a successful implementation of interventions are described. Afterwards, the newly developed intervention is presented and embedded in the context of gifted education. The introductory chapter concludes by deriving the research questions that are addressed in the three empirical studies of the present dissertation. These studies are presented in Chapters 2, 3, and 4. In the final Chapter 5, the findings of the three empirical studies are discussed and integrated into the broader research context. The dissertation closes with implications of the current results for future research and educational practice.

1.1. Theoretical Conceptualization of the Understanding of Science

To date, there is no universal view or standard conceptualization of the broad construct: *the understanding of the nature of science* (for reasons of legibility, this is subsequently referred to as an *understanding of science*; Deng, Chen, Tsai, & Chai, 2011). However, a certain consensus on the understanding of science has been recognized among science educators, where Lederman's (1992) operational definition has widely been used. According to his definition, the understanding of science refers to the *epistemology of science* as an understanding of the nature and the development of scientific knowledge as distinct from the scientific process and its contents (Lederman, 1992; Lederman, Wade, & Bell, 1998; Lederman & Zeidler, 1987). The term epistemology of science was derived from the Greek terms *episteme* and *logos* and can be translated as *theory of knowledge* in science (see Greene, Sandoval, & Bråten, 2016). Such individual theories include assumptions about the independence of thought, creativity, tentativeness, an empirical base, subjectivity, testability, and the cultural and social embeddedness of scientific knowledge (Duschl, 1990; Lederman, 1992; Matthews, 1994). An adequate understanding of science includes sophisticated epistemic beliefs (individual representations about knowledge and knowing; see Hofer & Pintrich, 1997; Mason & Bromme, 2010) and an understanding of inquiry-based methods (which build the basis for the genesis, construction, and development of scientific knowledge) (Deng et al., 2011; Höttecke, 2001; Lederman, 2007). These components are relevant for critically reflecting on and judging scientific knowledge (Deng et al., 2011; Driver et al., 1996; Höttecke, 2001; Lederman, 2007).

The understanding of science as a broad construct is grounded in many research disciplines, and all these disciplines contribute to the understanding of the construct (McComas, 1998). These disciplines include the philosophy of science, history of science, sociology of science, and psychology of science (their interplay is shown in Figure 1). The philosophy of science category makes the largest contribution. It provides assumptions about "what science is and how it is done" (McComas, 1998, p. 50). Therefore, it contributes to the epistemology of scientific knowledge, which focuses on the "area of philosophy concerned with the nature and justification of human knowledge" (Hofer & Pintrich, 1997, p. 88). In focusing on how scientific knowledge is developed, the philosophy of science emphasizes the importance of empirical evidence, especially the role of

observation and experimental evidence. It thereby adds to the meaning of creative processes, logical arguments, and skepticism in science. Due to how scientific knowledge is developed, the philosophy of science contributes to the understanding that scientific knowledge has inherent limitations, that it changes over time, and that the changes are usually gradual. Accordingly, scientific revolutions can offer an additional agent of change. The sociology of science category comprises authors' statements about who scientists are and how they work (McComas, 1998). This includes, for example, aspects of scientists' ethical decision making and the clear and open reporting of new knowledge (e.g., peer review, replication of procedures, accurate record keeping). The psychology category contributes an understanding of the characteristics of scientists (e.g., that they should be creative, intellectually honest, and open to new ideas). It also refers to the inherent biases that exist when scientists make observations. Last, according to McComas (1998), the elements from the history of science refer to science as a social tradition. Science has global implications and plays an essential role in the development of technology. This includes the proposal that scientific ideas are often affected by social and historical contexts.

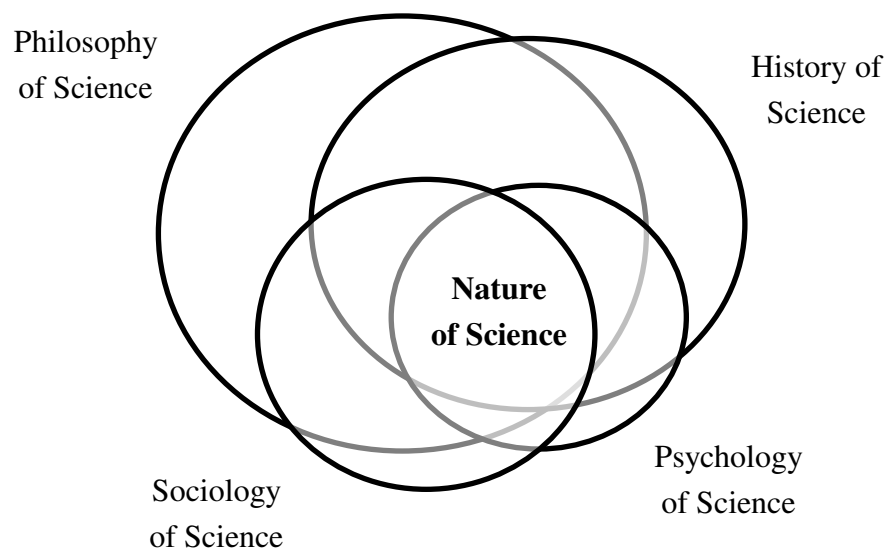


Figure 1. A proposal for the disciplines that add to our understanding of the nature of science, based on a content review of various documents on science education standards. Each discipline's approximate contribution is represented by the relative sizes of the circles (illustration from McComas, 1998, p. 50).

The understanding of science is embedded in science education in Western civilizations (e.g., EC, 2007; Jones, Wheeler, & Centurino, 2015; Kultusministerkonferenz [KMK], 2009; OECD, 2016). Traditionally, science curricula have focused on content knowledge in the natural sciences and on what one needs to know to do science. However, since the importance of the development of an adequate understanding of science was recognized, the perspective on science education has shifted from “*what* we know to *how* we know and *why* we believe” (Duschl, 2008, p. 269). A recent report by the National Research Council (2007) lists four important strands of scientific proficiency for all students. According to the NRC (2007), cited by Duschl, 2008, p. 269), students who understand science (a) know, use, and interpret scientific explanations of the natural world, (b) generate and evaluate scientific evidence and explanations, (c) understand the nature and development of scientific knowledge, and (d) participate productively in scientific practices and discourse. The latest benchmarks of international large-scale studies as PISA or TIMSS confirm the crucial significance of students’ understanding of science, already at elementary school age. According to TIMSS 2015, for instance, fourth graders should at an advanced level be able to “demonstrate basic knowledge and skills related to scientific inquiry, recognizing how a simple experiment should be set up, interpret the results of an investigation, reasoning and drawing conclusions from descriptions and diagrams, and evaluating and supporting an argument” (Martin et al., 2015, p. 67).

In the following chapters, two central elements of the understanding of science—which have a particular relevance in the empirical studies of this dissertation—are discussed more extensively. Chapter 1.1.1. focuses on epistemic beliefs. Chapter 1.1.2. focuses on inquiry-based methods, which build the basis for the genesis, construction, and development of knowledge in science (e.g., Deng et al., 2011). In Chapter 1.1.3., children’s development of the understanding of science is described, and Chapter 1.1.4 completes the introductory conceptual chapter by taking a closer look at how the understanding of science is related to other constructs.

1.1.1. Epistemic beliefs

Epistemic¹ beliefs refer directly to the epistemology of science (see Elby, Macrander, & Hammer, 2016; Lederman, 1992, 2007) and play a crucial role for an adequate understanding of science (Lederman, 1992, 2007). The word *epistemic* is derived from the Greek term *episteme*, which means “knowledge, what is known, or the way of knowing” (Greene et al., 2016, p. 2). The adjective *epistemic* means “of or relating to knowledge” (Kitchener, 2011, p. 92). On the basis of this word origin, Hofer and Pintrich (1997) defined epistemic beliefs as subjective beliefs about the *nature of knowledge* and the *nature of knowing* in science. Beliefs about the nature of knowledge refer to what one believes knowledge is. Beliefs about the nature of knowing are beliefs about the process by which one comes to know in science and the theories and beliefs one holds about knowing (see Elby et al., 2016; Hofer & Pintrich, 1997; Lederman, 2007).

Epistemic beliefs have been described as domain-specific (within a specific discipline), domain-general (independent of a specific discipline), or both (Hammer & Elby, 2002; Muis, Bendixen, & Haerle, 2006; Pintrich, 2002). As there is increasing evidence that epistemic beliefs differ across disciplines (e.g., Buehl, Alexander, & Murphy, 2002; Muis et al., 2006), domain-specific approaches have been the focus of current research and have been recommended (e.g., Greene et al., 2016). This dissertation follows a domain-specific perspective and refers—unless otherwise stated—to epistemic beliefs in the domain of science (see Conley, Pintrich, Vekiri, & Harrison, 2004; Elby, Macrander, & Hammer, 2016).

Independent of questions regarding the domain-specificity of epistemic beliefs, in recent decades, another major line of research has focused on identifying dimensions of epistemic beliefs (Hofer & Pintrich, 1997; Schommer, 1990, 1994). There is currently a debate on this issue (e.g., Chinn, Buckland, & Samarapungavan, 2011). However, this dissertation builds on Conley et al.’s (2004) conceptualization which is for the following reasons particularly suitable for this dissertation: It takes a domain-specific approach and focuses on elementary school children. Conley et al. (2004) built on Hofer and Pintrich’s (1997) definition of epistemic beliefs and distinction between the nature of knowing and the nature of knowledge by differentiating between four dimensions of epistemic beliefs,

¹ The terms *epistemic beliefs* and *epistemological beliefs* have been used interchangeably (see Greene et al., 2016). For the sake of simplicity, only the term *epistemic* will be used the following.

which are assigned to two categories. In the area of nature of knowing, they proposed the dimensions *source of knowledge* and *justification of knowledge*. Under the nature of knowledge, they suggested the dimensions *certainty of knowledge* and *development of knowledge* (see Figure 2).

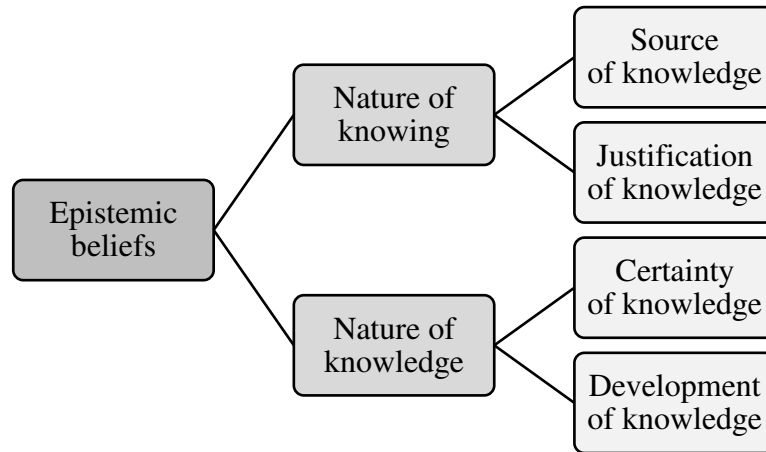


Figure 2. Postulated structure of epistemic beliefs (according to Conley et al., 2004).

The source dimension addresses beliefs about knowledge that resides in external authorities. In less sophisticated stances, knowledge is conceptualized as “external to the self, originating and residing in outside authorities” (Conley et al., 2004, p. 190). More sophisticated stances view knowledge as a product of experimental evidence, thinking, or interacting with others. The justification dimension refers to the role of experiments and how students evaluate claims. Less sophisticated stances include assumptions about absolute or nonreflected judgments. Stances that are more sophisticated include justified judgments and the acceptance of a variety of explanations for scientific phenomena (Conley et al., 2004). The certainty dimension addresses beliefs about whether knowledge is fixed or fluid (see also Hofer & Pintrich, 1997). Less sophisticated stances involve “the belief in a right answer” (Conley et al., 2004, p. 194) or the belief in absolute truths. By contrast, more sophisticated views can be identified by statements such as “there may be more than one answer to complex problems” (Conley et al., 2004, p. 190). Finally, the development dimension is associated with beliefs that recognize science as an evolving subject. Less sophisticated stances regard scientific ideas and theories as unchangeable. Stances that are more sophisticated include statements about how scientific ideas are continuously changing (e.g., due to new discoveries or data; Conley et al., 2004). Sample

items for assessing these dimensions can be found in the description of measurement approaches in Chapter 1.2.2. in Table 2.

Conley et al.'s (2004) conceptualization is based on fundamental work by Hofer and Pintrich (1997), who provided an important foundation for research in the field. Hofer and Pintrich (1997) had previously postulated four dimensions of epistemic beliefs, three of which were adopted by Conley et al. (2004), namely, source of knowledge, justification of knowledge, and certainty of knowledge. Hofer and Pintrich's (1997) fourth dimension was *simplicity of knowledge* (whether knowledge is viewed as the accumulation of facts or as highly interrelated concepts). In line with Elder (2002), Conley et al. (2004) substituted the dimension of development of knowledge for the dimension of simplicity. Elder (2002) investigated elementary school children's understanding of science and identified the development of knowledge as a central aspect in their understanding of science. Because Conley et al. (2004) also focused on elementary school children, they included this aspect as a dimension in their model.

1.1.2. Inquiry-based methods

Inquiry-based methods build the basis for the genesis, construction, and development of knowledge in science (e.g., Deng et al., 2011). As described in Chapter 1.1.1., an understanding of these methods is an important prerequisite for a critical reflection on and judgment of scientific knowledge and therefore a fundamental element of the understanding of science (e.g., Deng et al., 2011; Dogan & Abd-El-Khalik, 2008; Driver et al., 1996; Höttecke, 2001; Lederman, 2007; Ryder & Leach, 2000). Scientific inquiry and the understanding of science have even been described as inseparably intertwined with each other (e.g., Duschl & Osborne, 2002; Grandy & Duschl, 2007; Shipman, 2004).

Inquiry-based methods involve cyclical scientific activities that build on the so-called *scientific inquiry cycle (SIC)*. The SIC includes the following steps: (a) the generation of hypotheses on the basis of a specific research question (derived from theory or the results of previous research), (b) the planning and conducting of experiments, (c) data collection, (d) analysis, (e) evaluation of evidence, and (f) the drawing of inferences (Kuhn, 2002; White & Frederiksen, 1998; White et al., 2009; Zimmerman 2007). The SIC subsumes all individual components of scientific inquiry under a meta-perspective. Those components build the basis of knowledge acquisition and change (Kuhn & Franklin, 2006; Zimmerman, 2007). All of the steps of the SIC can be arranged to build a cycle,

but as inferences from an experiment lead mostly to new research questions or hypotheses and the start of a modified inquiry process, they correspond more closely to a spiral (see Figure 3).

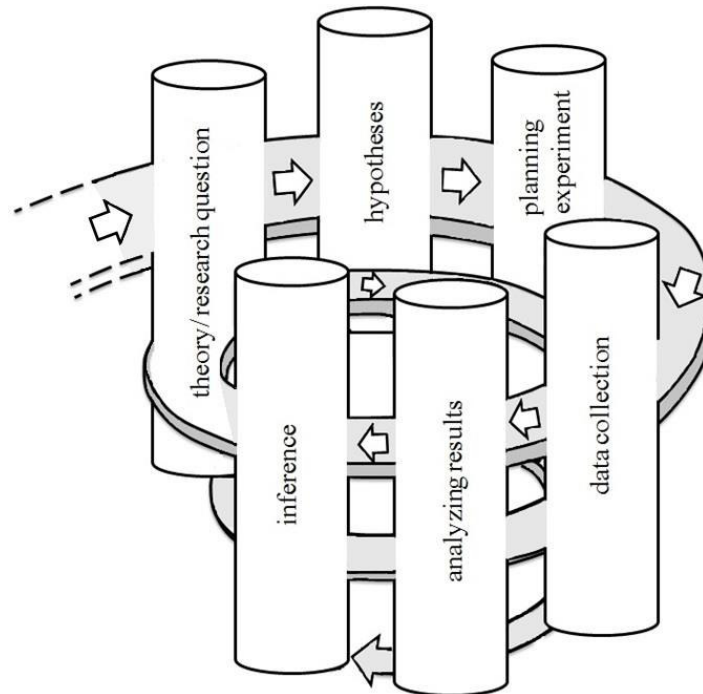


Figure 3. Steps of the scientific inquiry cycle (SIC), authors' own illustration (based on Klahr & Dunbar, 1988; Kuhn, 2002; White & Frederiksen, 1998; White et al., 2009; Zimmerman 2007).

Mature scientific inquiry does not necessarily proceed in the postulated stepwise manner. Furthermore, the exact sequence of the steps differs in the literature, and critical discussions of this matter have ensued (for an overview, see Pedaste et al., 2015). It is of course also possible to start anywhere in the cycle, and scientists do not necessarily proceed through these steps of inquiry in a fixed order (see Pedaste et al., 2015). Nevertheless, the SIC represents the theory-driven deductive approach that has been approved and is applied by scientists in empirical investigations (see Popper, 1935; White et al., 2009). Furthermore, the understanding of these steps is essential for inquiry-based science learning approaches as well as for scientific reasoning and argumentation (Colburn, 2000; Kuhn, 2010; Kuhn & Dean, 2005). Therefore, the SIC is an effective initial model that can enable students to develop the abilities to engage in inquiry and an understanding of its constituent processes (White & Frederiksen, 1998, 2005).

Within the SIC, the prime empirical method is the experiment (e.g., Zimmerman, 2007). Key features of an experiment are control over variables, careful objective measurement, and the establishing of cause and effect relations (NRC, 1996; Zimmerman, 2007). The so-called *control of variables strategy* (CVS; Chen & Klahr, 1999; Zimmerman, 2007) is a basic, domain-general experimentation strategy that comprises the systematic combination of variables. The CVS is relevant for the targeted testing of hypotheses and enables valid inferences to be made from experiments (Simon, 1989; Zimmerman, 2007).

Thinking processes within the SIC are defined as *scientific reasoning* (Kuhn, 2002; Zimmerman, 2007). This process of knowledge acquisition and change encompasses the abilities to generate, test, and revise theories and hypotheses and to reflect on this process (Kuhn & Franklin, 2006; Wilkening & Sodian, 2005; Zimmerman, 2007). Scientific reasoning is considered a cumulative and cyclical process that requires the intentional coordination of theory and evidence (Kuhn, 2002). The understanding of the scientific inquiry cycle can be considered a core element of scientific reasoning.

1.1.3. Elementary school children's understanding of science

The early promotion of young children's understanding of science is the focus of national and international education standards (EC, 2007; NRC, 2011; NSB, 2010; OECD, 2011, 2016). Detailed knowledge about the development of children's understanding of science is a prerequisite for effectively fostering children's abilities and beliefs (i.e., through science interventions or school curricula). A brief overview of the development of the understanding of science is given in the following sections. In line with the presented structure and conceptualization in Chapter 1.1., it focuses on the already defined central elements of the understanding of science: epistemic beliefs and inquiry-based methodological competencies.

Epistemic Beliefs

For many years, elementary school children's epistemic beliefs were not in the focus of cognitive development research (Elder, 2002; Kuhn & Park, 2005). According to Kuhn and Weinstock (2002), the conceptual ambiguity and complexity of this topic are possible reasons for its neglect. Another reason might be the assumption held by researchers who espoused the Piagetian hypothesis that elementary school children are concrete

thinkers and do not possess a level of abstraction that is sufficient for epistemic thinking (Inhelder & Piaget, 1958). But ever since increasing evidence from the cognitive development literature has suggested that young children are already able to develop an understanding of the epistemology of science (e.g., Montgomery, 1992; Wellman, 1990), more research has focused on investigating the epistemic beliefs of children.

The central model for the description of the qualitative development of epistemic beliefs was generated by Kuhn and Weinstock (2002) who defined and described different levels (see Table 1). According to Kuhn and Weinstock (2002), the developmental task that underlies the achievement of a mature epistemic understanding is the coordination of the subjective and objective dimensions of knowing. This progression is reflected in the different levels that people move through as they grow from early childhood to adulthood and progress from a realistic to an evaluativist level. According to Kuhn and Weinstock (2002), epistemic development across the different stages is a progression from “claims as copies to claims as facts, opinions, and finally judgements” (p. 125).

Table 1

Levels of Epistemic Understanding (according to Kuhn & Weinstock, 2002, p. 124)

<i>Level</i>	<i>Assertions (A)</i>	<i>Reality (R)</i>	<i>Knowledge (K)</i>	<i>Critical Thinking (CT)</i>
Realist	A are copies of an external reality	R is directly knowable	K comes from external sources and is certain	CT is unnecessary
Absolutist	A are facts that are correct or incorrect in their representation of reality	R is directly knowable	K comes from external sources and is certain	CT is a vehicle for comparing assertions with reality and determining their truth or falsehood
Multiplist	A are opinions freely chosen by and accountable only to their owners	R is directly knowable	K is generated by human minds and is uncertain	CT is irrelevant
Evaluativist	A are judgments that can be evaluated and compared according to criteria from arguments and evidence	R is not directly knowable	K is generated by human minds and is uncertain	CT is valued as a vehicle that promotes sound assertions and enhances understanding

At the first (*realist*) level, reality is directly knowable, and knowledge comes from external sources and is certain (Kuhn & Weinstock, 2002). According to the *absolutist* view (see Mason, 2016), knowledge is “absolute, certain, non-problematic, right or wrong, and does not need to be justified because it is based on observations from reality or authority” (p. 376). From the *multiplist* view, knowledge is “conceived as ambiguous and idiosyncratic, thus each individual has his or her own views and truths” (Mason, 2016, p. 376). Finally, at the *evaluativist* level, an individual believes that there are “shared norms of inquiry and knowing, and some positions may be reasonably more supported and sustainable than others” (Mason, 2016, p. 376). According to Kuhn (2000), it is only at this level that objective and subjective dimensions are balanced because they are integrated and coordinated.

Transferring this model to children’s development of their epistemic beliefs, Kuhn and Weinstock (2002) suggested that preschoolers could be described as *realists* but already show some epistemic awareness. At the elementary school level, children have mostly been described as *absolutists*, reaching a *multiplist* level between middle and late childhood, a so-called “constructivist theory of mind” (Carpendale & Chandler, 1996). Elementary school age children can recognize, for example, that exposure to different information may lead to different knowledge claims (e.g., Carpendale & Chandler, 1996). Further evidence for the transition from an absolutistic to a multiplist level at the end of elementary school was provided by Elder (2002) who analyzed the epistemic beliefs of fifth graders. Elder (2002) summarized that students at this age had a mixture of naive and sophisticated understandings: On the one hand, children tended to regard scientific knowledge as a developing, changing construct that is created by reasoning and testing. On the other hand, they displayed naive notions of science as a mere activity rather than as directed by aims to explain phenomena in the world.

Inquiry-Based Methodological Competencies

As described in Chapter 1.1.2., the prime empirical method applied in the SIC is the experiment (Zimmerman, 2007). In this section, elementary school children’s conceptions about the role of experiments and their experimentation competence are summarized.

With regard to the role of experiments, children in elementary school frequently possess misconceptions (Höttecke, 2001). They often believe that experimentation is synonymous with the production of effects or finding something out (e.g., Höttecke, 2001). There is evidence that elementary school children do not necessarily associate goal-oriented procedures with experimentation and do not recognize the necessity of repeating experiments or the systematic variation of materials (Meyer & Carlisle, 1996). Furthermore, elementary school children assume that researchers' interpretations of the results of experiments are unbiased and that researchers are not influenced by their expectations or their prior knowledge (e.g., McComas, 1998).

Beyond children's beliefs and assumptions about the function of experiments, their experimentation competence has been the focus of developmental research (Zimmerman, 2007). There is evidence that even preschool children possess simple experimentation competencies and evidence evaluation skills (Mayer et al., 2014; Zimmerman, 2007). They understand, for example, the relation between covariation data and a causal belief (e.g., Koerber, Sodian, Thoermer, & Nett, 2005). In elementary school, children can differentiate hypotheses from evidence and prefer controlled experiments over confounded ones, even though they have trouble spontaneously producing the CVS (Bullock, Sodian, & Koerber, 2009; Bullock & Ziegler, 1999; Sodian, Zaitchik, & Carey, 1991). Research indicates that preadolescent children possess at least a basic conceptual understanding of hypothesis testing and evidence evaluation (Koerber, Mayer, Osterhaus, Schwippert, & Sodian, 2015). The developmental literature has described children's understanding of experimentation (e.g., hypothesis testing, evidence evaluation) as proceeding from naïve conceptions, to partially correct (intermediate) conceptions, and finally to appropriate (mature) conceptions (Koerber et al., 2015; Sodian, Jonen, Thoermer, & Kircher, 2006; Zimmerman, 2007).

1.1.4. Relations of the understanding of science to other constructs

After presenting the conceptualization of the understanding of science and its development, relations to other constructs (personality traits, cognitive abilities, and investigative interests) that might have an impact on students' understanding of science are described in the following. Such relations are relevant for the theoretical conceptualization (to distinguish the construct from related constructs) and for the measurement of the understanding of science (i.e., with regard to its construct validity, which is described in

Chapter 1.2.1.). Such relations are furthermore important in the context of interventions on students' understanding of science. It can thereby be investigated if an intervention affects not only the understanding of science, but also related constructs, which might then be considered as outcome variables (see Chapter 1.3.1.).

Personality Traits: Need for Cognition and Epistemic Curiosity

Engaging in scientific inquiry requires active thinking and reasoning (Kuhn, 2002; Lawson, 2005) and might therefore be closely related to the constructs *need for cognition* (Cacioppo & Petty, 1982; Hofer, 2004) and *epistemic curiosity* (Hofer, 2004; Litman, 2008). Need for cognition is defined as the “tendency of an individual to engage in and enjoy thinking” (Cacioppo & Petty, 1982, p. 116). People with a high level of need for cognition show a pronounced willingness to solve problems through thinking and reflecting. People with a low need for cognition tend to avoid cognitively demanding activities (Oschatz, 2011). Specifically, a need for cognition has been considered an *epistemic motive*, an individual disposition for the willingness to engage in thinking (Oschatz, 2011). Need for cognition has been found to positively affect cognitive behavior such as elaborating on, evaluating, and recalling information (i.e., Peltier & Schibrowsky, 1994) as well as problem solving and decision making (e.g., Nair & Ramnarayan, 2000).

Epistemic curiosity is the desire for knowledge that motivates individuals to learn new ideas, to eliminate information gaps, and to solve intellectual problems (Litman, 2008; Litman & Spielberger, 2003). It has been found to be positively related to epistemic beliefs (Richter & Schmid, 2010), exploratory behavior, and the closure of gaps in knowledge (Litman, Hutchins, & Russon, 2005). There is evidence that need for cognition and epistemic curiosity positively affect problem solving and motivate individuals to learn new things (e.g., Fleischhauer, 2010; Litman, 2008; Litman et al., 2005; Nair & Ramnarayan, 2000; Peltier & Schibrowsky, 1994; Richter & Schmid, 2010). High levels of need for cognition and epistemic curiosity might be important prerequisites for making an effort to examine and solve scientific problems.

Cognitive Abilities

There are contradictory findings regarding the relation of certain aspects of the understanding of science and cognitive abilities. Scientific inquiry requires farsighted

thinking and planning and involves a variety of cognitive and metacognitive abilities (e.g., Kuhn, 2002; Morris et al., 2012; Zimmerman, 2007).

From a theoretical point of view, it can be derived that cognitive as well as meta-cognitive abilities are involved and required for engaging in the SIC (the suitability of the theme in the context of gifted education is described in Chapter 1.3.3.). It can be assumed that deductive as well as inductive reasoning processes are involved in the SIC (see Figure 4). In particular, deductive processes are required in connection with the derivation of hypotheses from theory, and inductive processes are involved in the generalization of findings or the derivation of theories and laws (Lawson, 2005; McComas, 1998).

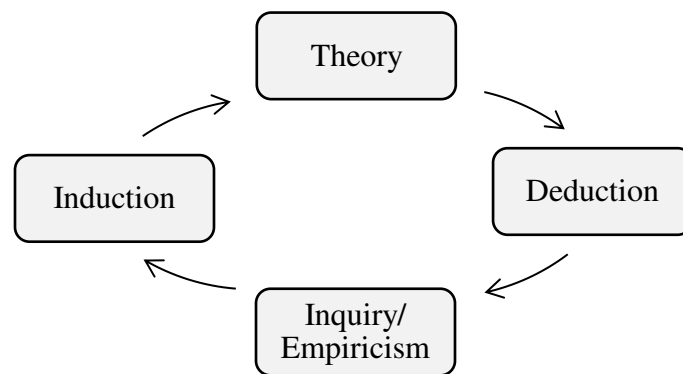


Figure 4. Embedding of deductive and inductive reasoning in the process of scientific inquiry (according to McComas, 1998, p. 59).

Relations between epistemic beliefs and intelligence as well as relations between scientific reasoning and different cognitive abilities have rather rarely been investigated empirically (e.g., Mayer et al., 2014). Results differ in part but point to positive relations between scientific reasoning and measures of general intelligence across different age groups (moderate positive correlations have been found at the elementary school level; Mayer et al., 2014). The scientific reasoning abilities of elementary school children have also been found to be positively related to additional cognitive abilities such as reading skills, problem-solving skills, and spatial abilities (Mayer et al., 2014). Few studies have investigated the relations between epistemic beliefs and cognitive abilities. Empirical results have primarily focused on secondary or university students and have pointed to low to moderate positive correlations (e.g., Trautwein & Lüdtke, 2007).

Investigative Interests

Vocational interests play an important role in students' achievement in STEM disciplines and can predict later career decisions (Kahn & Scott, 1997; Lapan, Shaughnessy, & Boggs, 1996; Leibham, Alexander, & Johnson, 2013). According to Holland's theory (1997), vocational interests are classified as *realistic*, *investigative*, *artistic*, *social*, *enterprising*, and *conventional* (RIASEC model). Students with a high level of investigative interests prefer activities that involve thought, observation, investigation, exploration, and discovery. They like to solve problems, perform experiments, and conduct research (Holland, 1997). Investigative interests are thus relevant for the development of STEM knowledge and skills (Carnevale et al., 2011).

Empirical evidence has shown positive relations between investigative interests and abilities in math and science (see Ackerman & Heggestad, 1997). Thereby, reciprocal relations between the constructs are theoretically assumed in the following way: On the one hand, it is expected that students' prior achievement influence their interests. Accordingly, students with high achievement in science show high interest in this domain (Ackerman, 1996; Carnevale et al., 2011). On the other hand, students with a high level of investigative interests prefer activities related to science and engage in scientific activities (Holland, 1997). Consequently, they engage more intensely and frequently in such tasks (Ackerman, 1996), which improve students' knowledge and skills, and in the long-term, their science achievement (Carnevale et al., 2011).

Thus, investigative interests might lead to more practical activities that are part of scientific inquiry and might therefore be important for the development and fostering of students' understanding of science. It can be assumed that investigative interests and various scientific activities lead to a deeper understanding of how "the scientific enterprise operates (McComas et al., 1998) and how scientific knowledge develops (Lederman, 1992, 2007).

1.2. Empirical Measurement of the Understanding of Science

1.2.1. Quality criteria for instruments

As described in the introductory chapter, instruments for measuring the understanding of science are required to describe children's competencies and to measure their progress in the context of science learning at school and in extracurricular contexts (i.e., pretest and posttest measures in interventions; Mason, 2016; Zimmerman, 2007). In order to adequately assess children's understanding of science, instruments need to meet a variety of quality criteria. There are clear guidelines regarding the (a) objectivity, (b) reliability, and (c) validity of instruments, to name the most important ones (Cohen, Swerdlik, & Phillips, 1996; Moosbrugger & Kelava, 2008). Objectivity refers to a measure's independence from the people who administer, evaluate, or interpret the test (Moosbrugger & Kelava, 2008). Reliability refers to the degree to which a test is consistent and stable in measuring what it is intended to measure (Moosbrugger & Kelava, 2008).

A high level of validity is most important in the context of test development and therefore described more in depth in the following (see Downing & Haladyna, 2006). Validity refers to how well a test measures what it claims to measure (AERA, APA & NCME, 1999). Different types of validity can be distinguished. More specifically, construct validity can be described as the appropriateness of inferences that are made on the basis of observations or measurements, specifically whether a test measures the intended construct and does not measure other variables (Moosbrugger & Kelava, 2008). Content validity refers to the extent to which a measure represents all of a given construct's facets (Moosbrugger & Kelava, 2008). Criterion validity is the extent to which a measure is related to different outcomes (AERA, APA & NCME, 1999). The validity of a test can be improved by clearly defining and operationalizing the goals and objectives of a measurement instrument or by comparing the measure with measures and data that have already demonstrated good psychometric properties (AERA, APA & NCME, 1999). For a test to demonstrate a high level of validity, a systematic approach must be followed across the entire process of test development (see the model of systematic test development by Downing, 2006). This includes 12 procedures or steps for effective test development: overall plan, content definition, test specifications, item development, test design and assembly, test production, test administration, the scoring of test responses, passing scores,

reporting test results, item banking, and technical reports on the test. These steps should typically be followed in the development of most achievement, ability, or skill tests. According to the author, following these steps tends to maximize the amount of evidence that supports the validity of the intended interpretation of the test score.

1.2.2. Existing instruments and their boundaries

A variety of instruments and approaches have been used to assess different aspects of the understanding of science in children as well as in adults (for a historical overview, see Lederman, Wade, & Bell, 1998). In the following, an overview of existing measurement instruments is provided. Approaches that are appropriate for elementary school children are pointed out. The focus is on instruments that can be used to assess epistemic beliefs and inquiry-based methodological competencies.

Measurement of Epistemic Beliefs

The measurement of epistemic beliefs is complex because of the “nature of the construct itself, its definition, and the different levels at which it can be measured” (Mason, 2016, p. 388). Because there are many definitions, conceptual frameworks, and methodological perspectives on epistemic beliefs, there are different types of measurement. In line with the current review by Mason (2016), the main measurement approaches are summarized and critically reviewed within their corresponding conceptual framework.

Epistemic beliefs as multidimensional sets or systems of beliefs

This approach is based on the definition of epistemic beliefs² in terms of multiple sets of more or less independent beliefs about the nature of knowledge and knowing. As described in Chapter 1.1.1., this line of research is based on work by Hofer (2000; Hofer & Pintrich, 1997) and Schommer (1990; Schommer-Aikins, 2002). The multidimensional perspective on epistemic beliefs has adopted self-report questionnaires that employ Likert-type scales to assess the “degree of agreement with certain statements about

² Mason (2016) uses in his review the term *epistemic cognition*. There is an ongoing debate between different research groups on the use of the terms *epistemic cognition* and *epistemic beliefs*. Epistemic cognition describes the thinking processes that focus on epistemic issues (e.g., Chinn et al., 2011; Greene et al., 2016). However, the terms epistemic cognition and epistemic beliefs can be used interchangeably (for a summary, see Greene et al., 2016). For the sake of simplicity, only the term epistemic beliefs is used in the following.

knowledge and knowing” (Mason, 2016, p. 379). There is no doubt that questionnaires offer advantages because they enable an efficient and standardized measure of epistemic beliefs in group-testing situations or large-scale surveys (Moosbrugger & Kelava, 2008). Questionnaires have primarily been used in studies that have aimed to examine relations between epistemic beliefs and facets of academic achievement, such as reading comprehension, problem solving, text processing, and conceptual change, or academic self-concept and personality variables (Kardash & Howell, 2000; Mason, 2003; Schommer, 1990; Schraw, Dunkle, & Bendixen, 1995; Sinatra, Southerland, McConaughy, & Demastes, 2003; Trautwein & Lüdtke, 2008).

Besides their advantages, questionnaires have also been criticized for a number of reasons. From a psychometric point of view, Mason (2016) pointed out problems such as limited validity and reliability. Instruments might not capture all dimensions of epistemic beliefs adequately, and the theorized underlying factor structures have been difficult to establish definitively. Other criticisms are that it is difficult to map self-reports on to the complexity of the developmental trajectory and that a person’s scores are difficult to interpret (Mason, 2016).

Questionnaires to assess epistemic beliefs as a multidimensional set of beliefs have primarily been developed for secondary school students or adults and have only occasionally been used in studies with elementary school children. On the basis of previous work by Elder (2002) and Hofer and Pintrich (1997), Conley et al. (2004) developed an instrument for fifth graders which showed an acceptable reliability. As described in Chapter 1.1.1., the four dimensions of epistemic beliefs that were measured on a Likert scale are (a) source, (b) certainty, (c) development, and (d) justification of knowledge. The items can be found in Table 2.

Table 2

Items from the Questionnaire by Conley et al. (2004, p. 202f)

<i>Knowledge dimension</i>	<i>Items</i>
Source (-)	<ul style="list-style-type: none"> • Everybody has to believe what scientists say • In science, you have to believe what the science books say about stuff • Whatever the teacher says in science class is true • If you read something in a science book, you can be sure it’s true • Only scientists know for sure what is true in science

Certainty (-)	<ul style="list-style-type: none"> • All questions in science have one right answer • The most important part of doing science is coming up with the right answer • Scientists pretty much know everything about science; there is not much more to know • Scientific knowledge is always true • Once scientists have a result from an experiment, that is the only answer • Scientists always agree about what is true in science
Development (+)	<ul style="list-style-type: none"> • Some ideas in science today are different than what scientists used to think • The ideas in science books sometimes change • There are some questions that even scientists cannot answer • Ideas in science sometimes change • New discoveries can change what scientists think is true • Sometimes scientists change their minds about what is true in science
Justification (+)	<ul style="list-style-type: none"> • Ideas about science experiments come from being curious and thinking about how things work • In science, there can be more than one way for scientists to test their ideas • One important part of science is doing experiments to come up with new ideas about how things work • It is good to try experiments more than once to be sure about your findings • Good ideas in science can come from anybody, not just from scientists • A good way to know if something is true is to do an experiment • Good answers are based on evidence from many different experiments • Ideas in science can come from your own questions and experiments • It is good to have an idea before you start an experiment

Note. Items from the dimensions *source* and *certainty* (-) have to be recoded, as agreement points to less sophisticated epistemic beliefs. On the other hand, agreement with the items from the *development* and *justification* (+) dimensions indicates sophisticated beliefs.

Epistemic beliefs as the developmental progression of cognitive structures

As described in Chapter 1.1.3., developmental psychologists have defined epistemic beliefs in terms of domain-general cognitive structures that characterize a level or stage of cognitive development (e.g., King & Kitchener, 1994; Kuhn, 2000). Kuhn (2000; Kuhn & Weinstock, 2002) labeled developmental progression in terms of relations between objective and subjective positions that move from an absolutist to an evaluativist

point of view regarding knowledge (see levels of epistemic understanding by Kuhn & Weinstock, 2002, described in Chapter 1.1.3.).

Researchers who embrace this developmental perspective have primarily used qualitative measures as interviews to assign respondents to a general epistemic level (e.g., *Reflective Judgment Interview*, King & Kitchener, 1994; *Livia problem*, Kuhn, Pennington, & Leadbeater, 1983). Developmental theorists have furthermore used paper-pencil instruments involving ill-structured scenarios (e.g., Wood, Kitchener, & Jensen, 2002) or fixed-choice questions about contrasting claims (Kuhn, 2000; Kuhn & Weinstock, 2002). Finally, supplemented by interviews, vignettes with text and pictures have been used from the elementary level onwards to assess levels of epistemic development (Mansfield & Clinchy, 2002).

Approaches to assess epistemic beliefs as the developmental progression of cognitive structures can provide on the one hand an exhaustive and authentic description of students' representations and assumptions about knowledge and knowing (Mason, 2016). On the other hand, such methods are very time-consuming and expensive as they require partially complex coding. This can lead to a reduced test objectivity and reliability (Banister, 2011). Furthermore, they can only to a limited extent be applied in group-testing situations.

Epistemic beliefs as situated resources

Researchers who espouse a situative perspective on learning processes have defined the so-called *epistemic resources* (Hammer & Elby, 2002) as fine-grained representations used in a multiplicity of situations. They point to the importance of the context in which learning takes place (Mason, 2016). According to these researchers, epistemic beliefs cannot be measured with traditional quantitative methods but by observations of teaching and learning processes, supplemented by interviews (e.g., diSessa, Elby, & Hammer, 2003). However, like the methods described for epistemic beliefs as the developmental progression, these methods are very complex and time-consuming and are suitable for qualitative research.

Current measures

In reference to Mason (2016), researchers have recently explored new measures or revisited old measures to assess epistemic beliefs. The following alternatives to paper-

and-pencil tests are in the focus of current research: Think-aloud protocols of epistemic beliefs in action (e.g., Mason, Ariasi, & Boldrin 2011), knowledge artifacts and discourse (Sandoval, 2005; Rhu & Sandoval, 2012), cognitive interviews (e.g., Greene & Yu, 2014), and finally, scenario-based instruments (Barzilai & Weinstock, 2015). Those practices were intended to overcome some limitations associated with the tradition of using self-report questionnaires (i.e., a limited assessment of children's developmental stages). However, most of those methods are very complex and further research is required to validate those instruments.

Measurement of Inquiry-Based Methodological Competencies

In the context of the measurement of inquiry-based methodological competencies, different approaches and task formats (qualitative and quantitative) have been developed for children at elementary school age. Most of them have focused on the measurement of scientific reasoning, which can—as stated in Chapter 1.1.2.—be described as the thinking processes within the SIC (Kuhn, 2002; Zimmerman, 2007). The tasks for assessing scientific reasoning focused mostly on single steps and processes within the SIC (in particular experimentation skills or strategies as the CVS). Those tasks have included, for example, interviews, self-directed experimentation tasks, simulation tasks, or story problems (e.g., Bullock & Ziegler, 1999; Carey, Evans, Honda, Jay, & Unger, 1989; Dunbar & Klahr, 1989; Kuhn et al., 1995; Schauble, 1996, quoted from Mayer et al., 2014). Children's performances have thereby been influenced by contextual support (e.g., abstract vs. concrete contexts), task complexity (e.g., single-variable vs. multivariable), response format (e.g., multiple choice vs. production), and prior knowledge in scientific domains (e.g., Bullock & Ziegler, 1999; Chen & Klahr, 1999; Kuhn et al., 1988; Lazonder & Kamp, 2012; Wilhelm & Beishuizen, 2003; Zimmerman, 2007).

So far, hardly any paper-and-pencil tests have been developed to assess children's scientific reasoning. Most recently, a one-dimensional paper-and-pencil test was designed for assessing different components of elementary school children's scientific reasoning using story problems with different response formats (i.e., multiple-choice, forced-choice, multiple-select, open-ended). The items referred to the components goals of science, theories and interpretative frameworks, experimentation strategies, experimental designs, and data interpretation (see Koerber et al., 2015; Mayer, 2011; Mayer et al., 2014). The results indicated that elementary school children in Grades 2 to 4 could be successfully

tested with this instrument, which showed a moderate reliability. The postulated components (e.g., experimentation strategies, data interpretation) formed a unitary construct and could not be separated empirically (Koerber et al., 2015; Mayer et al., 2014). Tasks assessing the understanding of the complete SIC do not yet exist for elementary school children.

Final Appraisal

A variety of approaches have been explored to measure different aspects of the understanding of science. As the understanding of science is a very wide-ranging construct, it is especially challenging to develop reliable and valid instruments for its measurement (Mason, 2016). Regarding the measurement of epistemic beliefs, different quantitative as well as qualitative approaches exist. Qualitative approaches have been in particular used to describe the development or level of children's epistemic beliefs (e.g., Kuhn & Weinstock, 2002). Most of those approaches (e.g., structured interviews, think-aloud protocols) are very time-consuming and complex (i.e., due to required coding), or are not applicable in group testing situations. Therefore, such methods are not suitable for the evaluation of interventions. However, quantitative measurement approaches (i.e., questionnaires) are appropriate for large-scale assessments or group-testing interventions. A variety of questionnaires (with slightly different dimensions) have been developed. For elementary school children, the instrument by Conley et al. (2004) is thereby the only available questionnaire and might due to its acceptable reliability suitable for the evaluation of science interventions.

Regarding the measurement of inquiry-based methodological competencies, also different approaches and task formats have been used. However, hardly any paper-and-pencil tests have been developed to measure elementary school children's scientific reasoning. The recently developed instrument (Koerber et al., 2015; Mayer et al., 2014) focused on different components of scientific reasoning, but was not able to assess the relationships between those components (i.e., by focusing on the understanding of the complete process of the SIC; see Kuhn & Dean, 2005; White et al., 2009; Zimmerman, 2007). It can thereby be assumed that existing instruments have not yet fully covered the theoretical richness of inquiry-based methodological competencies. This strengthens the need for the development of further reliable and valid instruments that can go beyond existing tests and measure central content areas of young children's understanding of the SIC.

1.3. Intervening in Students' Understanding of Science

1.3.1. Interventions and their implementation

As early as elementary school, the development of an adequate understanding of science is a normative goal of science education (e.g., Bildungsplan, 2004; EC, 2007; Mullis & Martin, 2015). As outlined in Chapter 1.1.3., elementary school age children possess a certain understanding of science, which in most cases can be described as rather naïve and absolutistic (e.g., Höttecke, 2001; Kuhn & Weinstock, 2002). Questions regarding an effective promotion of children's understanding of science have therefore been in the focus of educational research and practice for many years (e.g., EC, 2007; NRC, 2011; NSB, 2010; OECD, 2016).

Interventions provide an important approach through which to promote students' competencies (e.g., Lendrum & Wigelsworth, 2013). Interventions have been defined as "purposively implemented change strategies" (Fraser & Galinsky, 2010, p. 459) and are developed to support the behavior, conditions, achievement, or development of a certain target group (e.g., Blase, van Dyke, Fixsen, & Bailey, 2012; Humphrey et al., 2016; Lendrum & Wigelsworth, 2013). In order to test interventions under real-world conditions, it is necessary to implement programs. The implementation is the "process by which an intervention is put into practice" (Lendrum & Humphrey, 2012, p. 635). Thus, interventions and their implementation offer important opportunities for fostering students' understanding of science, and there has been a call for more interventions for elementary school children when they are in their "curiosity golden age" (EC, 2007, p. 12).

Interventions offer the opportunity to foster students' competencies because specific promotion programs and instructional design principles can be systematically compared and investigated. On this basis, effective programs can be developed and implemented (Lendrum & Wigelsworth, 2013). However, in order to develop a successful intervention that can be disseminated in practice, different stages are necessary to ensure that students will benefit from the program (Greenberg, Domitrovich, Graczyk, & Zins, 2005; Humphrey et al., 2016).

First, the instructional goals of an intervention should be defined on the basis of the demands of the target group (e.g., their age, grade level, prior knowledge, ability level). Next, the specific contents and the respective methods and didactic tools should

be chosen on the basis of theoretical and practical considerations (Hulleman & Cordray, 2009). Thereafter, the practicability, acceptability, and appropriateness and utility for the target group should be investigated in a pilot phase (Humphrey et al., 2016). Feedback from the pilot phase can be integrated into the intervention concept.

Afterwards, an efficacy study can be conducted to examine the success of the intervention under optimal and controlled conditions to maximize outcomes (Lendrum & Wigelsworth, 2013). Efficacy studies are typically conducted by the program developers and should demonstrate the internal validity of the program (Lendrum & Wigelsworth, 2013). Regarding the investigation of the efficacy of intervention studies, randomized controlled trials (*RCTs*) are considered the *gold standard* in psychological and educational research (Lendrum & Wigelsworth, 2013; Torgerson & Torgerson, 2013). *RCTs* can determine whether a prescribed intervention is able to produce the desired effects on a specified set of outcomes.

Provided that the efficacy study delivers positive findings, effectiveness studies can be conducted next to provide insight into whether an intervention works when implemented under real-world conditions (e.g., by using the staff and resources that would be normally available; Carroll et al., 2007; Dane & Schneider, 1998; Greenberg et al., 2005). An effectiveness trial tests the effectiveness of an intervention in the presence of contextual factors that might influence the successful adoption, implementation, and sustainability of the intervention (Bonell, Fletcher, Morton, Lorenc, & Moore, 2012; Greenberg, 2010).

Implementer characteristics (e.g., their professional characteristics such as education, skills, and experience, their perceptions and attitudes regarding the intervention, and their psychological characteristics such as stress, burnout, or self-efficacy) are among the most important contextual factors (Humphrey et al., 2016). Thereby, the implementation fidelity, in particular, might affect the success of the implementation and should therefore be in the focus of research (Darling-Hammond, 2000; Hulleman & Cordray, 2009; Humphrey et al., 2016; Rockoff, 2004; Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). Implementation fidelity refers to the degree to which an intervention or program is delivered as intended by the developers (Carroll et al., 2007). To date, there is no standard method of assessing fidelity as it depends strongly on the characteristics of the particular intervention (Abry, Hulleman, & Rimm-Kaufman, 2015). Different approaches for

measuring and considering fidelity have been used in educational research, such as exposure or dosage, quality of delivery, participant responsiveness, and program differentiation (Carroll et al., 2007; Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012). However, the documentation of adherence (e.g., compliance) is a fundamental prerequisite for fidelity and is a suitable and accepted measure in an effectiveness study. Its assessment provides insights into the time-related and organizational practicability of the intervention.

To investigate the effectiveness of an intervention under real-world conditions, an evaluation by means of RCTs is also considered the gold standard (Lendrum & Wigelsworth, 2013; Torgerson & Torgerson, 2013). However, this step has rarely been evaluated (see Fixsen, Blasé, Metz, & van Dyke, 2013), has failed (Spiel, Schober, & Strohmeier, 2016), or has led to reduced effects on the outcomes (e.g., Durlak & DuPre, 2008; Hulleman & Cordray, 2009). The effectiveness of an intervention under real-world conditions is the prerequisite for the scaling up of the program in a larger, more diverse population and in broader training contexts (Gottfredson et al., 2015; Humphrey et al., 2016).

1.3.2. Intervention approaches to promote the understanding of science

A variety of approaches and targeted interventions have been explored to answer the call to promote children's understanding of science (Bendixen, 2016; Cavagnetto, 2010; EC, 2007; Valla & Williams, 2012). In this chapter, the most important general approaches and recommendations with regard to the promoting of students' understanding of science are described. Afterwards, the state of research regarding existing interventions is described, and conclusions are derived.

Approaches for Promoting the Understanding of Science

Ever since the understanding of science has been embedded in science education in Western civilizations, the perspective on students' science education has shifted from *what* they know to *how* and *why* (Duschl, 2008). Recently, educational research and practice has emphasized the importance of *inquiry-based science education (IBSE)* (Blanchard et al., 2010; Colburn, 2000). Beyond the learning of science content and natural science phenomena, IBSE offers students an effective way to comprehend the nature of scientific inquiry, to learn about scientific practice, and to understand how to engage in the inquiry

process (Blanchard et al., 2010; Elder, 2002; Minner, Levy, & Century, 2009). This can occur when students work like scientists themselves (de Jong, 2006). Inquiry-based learning requires students to formulate hypotheses, conduct experiments, and draw conclusions (Klahr & Dunbar, 1988).

In the context of IBSE, a stepwise opening of the inquiry process is recommended (Colburn, 2000). According to Colburn (2000), the following forms of inquiry—ordered from a few to many degrees of freedom—can be distinguished: structured inquiry, guided inquiry, open inquiry, and whole learning cycles within the SIC. The classification of studies investigating the implementation of inquiry learning also depends on the existing levels of degrees of freedom in the research process (Bell, Smetana, & Binns, 2005). At the level of the smallest number of degrees of freedom, experiments are completely guided by the teacher who determines the research theme, the research question, the materials that are used, the design of the experiment, the expected results, the analytic strategy, as well as the conclusions. At the level of the most degrees of freedom, students conduct experiments independently. At this level of so-called open inquiry, the teacher might provide a research theme, but the students specify their own research questions and conduct all of the steps of the inquiry cycle in a self-determined manner. In this context, independent research competence is the educational objective. However, educational research indicates that open inquiry requires a step-by-step implementation, in particular, at the elementary school level (Höttecke, 2010).

In the context of IBSE, the meaning of hands-on science has been recognized (e.g., Flick, 1993; Klahr, Triona, & Williams, 2007). Such practical activities are supposed to lay the foundation for students' scientific processing skills and higher order (abstract) thinking skills (e.g., Aebli, 1980; Piaget, 1966). Furthermore, an explicit reflective approach—in which students' attention is actively directed toward relevant aspects of the epistemology of science via discussions, instruction, or critical scrutiny—has revealed positive effects on students' understanding of science (Akerson & Hanuscin, 2007). Also the use of conflicting information or scientific controversies demonstrated positive effects on students' understanding of science (e.g., Kienhues, Bromme, & Stahl, 2008).

Existing Interventions to Promote the Understanding of Science

Existing interventions concerning the understanding of science can be classified and described on the basis of different criteria. First, interventions differ with regard to

the predefined target group as their (a) age or (b) ability level. (A) Some attempts have been made to foster the scientific interest or science competencies of children as young as preschoolers (e.g., Patrick, Mantzicopoulos, & Samarapungavan, 2009). As described in Chapter 1.3.1., the view that intervening in elementary school is important has increased in recent years (EC, 2007; Bendixen, 2016). However, most interventions still focus on older students at the secondary school or college levels (e.g., Kienhues et al., 2008; Muis, Trevors, & Chevrier, 2016). In comparison with interventions at the secondary school level, interventions at the elementary school level are rather rare (e.g., Bendixen, 2016; Metz, 2011; Ryu & Sandoval, 2012; Valla & Williams, 2012). Those intervention approaches are described below in the third section. (B) Interventions were intended to support students with an average ability level, e.g., whole cohorts within classroom interventions (e.g., Bendixen, 2016; Erdosne Toth, Klahr, & Chen, 2000) or even to foster students with high intellectual abilities within the scope of science enrichment or talent programs (e.g., Stake & Mares, 2001).

Second, interventions differ with regard to their aims: The educational goals of science interventions vary from providing positive experiences in science, increasing school performances, fostering students' motivation or interest, to exposing students to role models or influencing career decisions from a long-term perspective (see Benbow, Lubinski, & Sanjani, 1999; Carnevale et al., 2011; Dorsen, Carlson, & Goodyear, 2006; Hulleman & Harackiewicz, 2009; Tsui, 2007; Valla & Williams, 2012; Veenstra, Padró, & Furst-Bowe, 2012; Wai, Lubinski, Benbow, & Steiger, 2010).

Third, science interventions have been developed to affect different student outcomes: They have often been intended to foster specific science content knowledge, but also practical skills, scientific processing skills, science concepts, or epistemic change (Andrés, Steffen, & Ben, 2010; Cotabish, Dailey, Robinson, & Hughes, 2013; Muis et al., 2016; Valla & Williams, 2012). There are only a few interventions that have focused on the enhancement of fundamental aspects of the understanding of science as early as elementary school (e.g., Bendixen, 2016; Conley et al., 2004; Metz, 2011; Ryu & Sandoval, 2012; Smith, Maclin, Houghton, & Hennessey, 2000; Sodian et al., 2006). The results of the key studies are summarized chronologically in the following. Smith et al. (2000) tested whether sixth graders could develop more sophisticated epistemic beliefs in a constructivist classroom (by inquiry learning and metacognitive stimulation) compared with a traditional science classroom (by factual learning). Only students in the constructivist

classroom developed an appropriate epistemological stance toward science that focused on the central role of ideas in the knowledge acquisition process and the mental, social, and experimental work that is involved in this process. Conley et al. (2004) investigated whether the epistemic beliefs of fifth-grade students could be enhanced during a 9-week science course. Results showed that students became more sophisticated in their beliefs about the source and certainty of knowledge. However, no reliable changes were found in the development and justification dimensions. Sodian et al. (2006) investigated the effects of a teaching unit about the understanding of science in fourth graders in comparison to regular science lessons. Using the nature-of-science interview (Carey et al., 1989) and one task about the CVS (Bullock & Ziegler, 1999), positive effects on students' understanding of the role of experiments in science as well as the design of controlled experiments. Metz (2011) investigated over two years "practical epistemologies" in the science classroom with first graders. The methods included teaching the goals of scientific inquiry, scaffolding students' ideas, or design of own experimentation. Children displayed afterwards partially sophisticated beliefs that included the uncertainty of results and strategies to improve their designs. Ryu and Sandoval (2012) investigated the improvement of 8-10-year-old children's epistemic understanding from sustained argumentation in a classroom intervention. They found that the students learned how to apply evidentiary criteria in their own written arguments and by evaluating arguments.

Although those intervention studies provide evidence on how children's epistemic beliefs might be successfully fostered, the results should be carefully interpreted. The research design of those studies is limited, in particular there are hardly any experimental or quasi-experimental intervention studies with control groups, which are needed to investigate causal relations (for a review, see also Bendixen, 2016).

Fourth, interventions vary from short-term programs (e.g., Kienhues et al., 2008) to long-term interventions that last for entire school years (e.g., Adey & Shayer, 1993; Metz, 2011; Smith et al., 2000). It can be assumed that the optimal duration of an intervention depends on the intended outcome and the specific program but that sustainable changes in students' understanding require a certain duration (see Barnett, 2011).

To build on this classification, it can be concluded that science interventions focusing on very fundamental aspects of elementary school children's understanding of science are lacking, in particular in students below class level 5 (see Lederman, 2007; McComas, 1998). Furthermore, the existing studies often show methodological shortcomings

such as a lack of proper control groups, a lack of randomization, or the failure to administer a baseline, so that their relevance is partly limited (see Bendixen, 2016; Brody, 2006; Valla & Williams, 2012).

1.3.3. Development of an extracurricular intervention

Because of this identified lack of intervention studies, an intervention for elementary school children was developed within this dissertation as part of an enrichment program for gifted children. This context was chosen because the promotion of students' understanding of science is not only important as part of scientific literacy and general education standards (EC, 2007; Jones et al., 2015; NRC, 1996; OECD, 2016), but also in the context of gifted education (NSB, 2010). In the following, the potential of gifted children is described, the significance of promoting their understanding of science is derived, the importance of interventions with regard to the promotion of the understanding of science for gifted children is pointed out, and finally, the development of an intervention as part of an enrichment program for gifted children is described.

Potential of Gifted Children

According to Subotnik, Olszewski-Kubilius, and Worrell (2011), gifted people have the potential to enrich society in scientific, aesthetic, and practical domains due to their high general cognitive abilities. These authors define giftedness as “the manifestation of performance or production that is clearly at the upper end of the distribution in a talent domain even relative to that of other high-functioning individuals in that domain” (Subotnik et al., 2011, p. 7). Furthermore, they view giftedness as developmental, which means that in the beginning stages, *potential* is the key variable of giftedness. Gifted children possess high cognitive as well as metacognitive abilities (Sternberg, 2005; Waldmann & Weinert, 1990). They demonstrate better skills in the acquisition of knowledge, a better working memory capacity, more efficient information processing strategies, as well as more abstract thinking skills (Waldmann & Weinert, 1990). In later stages, achievement is the measure of giftedness; and in fully developed talents, eminence is the basis on which this label is granted (Subotnik et al., 2011). Regarding the manifestation of giftedness, Subotnik et al. (2011) emphasize the important role of psychosocial variables at every developmental stage. Furthermore, they point out that both cognitive and psychosocial variables are malleable and need to be deliberately cultivated and fostered.

Promoting the Understanding of Science in Gifted Children

Gifted children with a high cognitive potential need to be promoted so that they might show outstanding achievements as adults and might even gain eminence later on. Eminence, which Subotnik et al. (2011) characterize as “contributing in a transcendent way to making societal life better and more beautiful” (p. 7), should be the outcome that gifted education aspires to achieve. From an instrumental perspective, gifted students have high relevance for society as they have the potential to make outstanding contributions to the welfare of all. They have the potential to become future STEM leaders and to “define the leading edge of scientific discovery and technological innovation” (NSB, 2010, p. 5). Young creative thinkers might generate new ideas and find solutions to the major social, economic, and environmental problems that plague the world (Subotnik et al., 2011). Gifted children might be a source of our “future national leaders, scientists, entrepreneurs, and innovators” (Subotnik et al., 2011, p. 11). Against the background of the relevance of mathematical and scientific foundations in modern life, it seems to be especially important to foster gifted students’ understanding of science (OECD, 2016). It can be assumed that gifted students have the capacity to perform at the highest level in STEM domains as adults and thus to support science and economy (NSB, 2010). Furthermore, gifted students—with interests in science subjects as well as in the humanities or arts—might, in particular, take key positions in democratic decision-making processes and act as disseminators of information regarding the responsible use of new scientific findings (see NSB, 2010; Oschatz & Schiefer, in press).

Interventions as a Component of Gifted Education

Interventions for gifted students are particularly relevant as they offer an effective opportunity to foster students in accordance with their individual needs (Stake & Mares, 2001; 2005). Interventions can be one component of students’ enrichment and can offer them an important opportunity to develop their talents (e.g., Heller, Mönks, Subotnik, & Sternberg, 2000; NSB, 2010). In this context, the National Science Board recommended “opportunities for excellence” for talented students (NSB, 2010, p. 2). Interventions can offer such opportunities as they can cover topics that go beyond the school curriculum and provide students the opportunity to work on specific topics in small groups of similarly interested and talented children (e.g., Rinn, 2006). Interventions can strengthen in-

terests, motivation, self-concept, self-esteem, achievement, or the talents of gifted students (Keeley, 2009; Kerr & Robinson Kurpius, 2004; Rinn, 2006). Furthermore, extra-curricular interventions can provide students a stimulating learning environment and avoid some of the in the TIMSS 2015 identified problems. Namely, that schools are often not well equipped for conducting scientific inquiry, that teachers do not often emphasize science investigation in class, and only have few resources for conducting science experiments (Martin, Mullis, Foy, & Hooper, 2015).

Dealing with the questions that surround the understanding of science can be considered suitable for the adequate cognitive activation of gifted children. This requires complex and abstract thinking skills, the cross linking of content areas, deductive as well as inductive reasoning, and the detection of rules and principles (Kuhn, 2002; Lawson, 2005; McComas, 1998). Because gifted children have great potential to achieve high cognitive performance (Subotnik et al., 2011), it can be assumed that gifted children, in particular, have the necessary intellectual capacities to develop an adequate understanding of science at an early age and to benefit from a deeper engagement with the topic.

Development of an Intervention to Promote the Understanding of Science in Gifted Children

Based on the described relevance of the promotion of the understanding of science in children with high cognitive abilities, an intervention was developed by scientists from the *Hector Research Institute of Education Sciences and Psychology* (Oschatz & Schiefer, in press; Oschatz, Schiefer, & Trautwein, 2015). The intervention, a 10-week course titled *Little Researchers—We Work like Scientists*, was part of an extracurricular enrichment program in the German state of Baden-Württemberg, the so-called *Hector Children's Academy Program (HCAP)*. To participate in this program, children have to be nominated by their teacher. No standardized intelligence tests are conducted to select the participants. The intention is that the 10% most talented or best-performing children in an age cohort will be given the opportunity to participate in this state-wide program (Golle, Herbein, Hasselhorn, & Trautwein, in press). Besides school performances, a high level of motivation and interest can be taken into account for the nominations. After admission, children can choose from a variety of courses. The intention is that the courses that are offered will cover topics that go beyond the regular school curriculum and will ensure a high level of cognitive activation, which is a key criterion for teaching quality (Kunter &

Voss, 2011) and is particularly relevant for the teaching of students with high cognitive abilities (Stapf, 2003). The courses offered at the HCAP focus on the STEM disciplines.

Intervention Concept

The Little Researchers course was developed with the goal of promoting a fundamental level of the understanding of science, as also stated in the education plans of many countries (e.g., Bildungsplan, 2004; EC, 2007; OECD, 2016; Wendt et al., 2016). Teaching science content knowledge was not a primary goal of the course. Therefore, the course primarily addresses topics that are already covered in elementary education anyway (e.g., human senses, swimming and sinking). On the basis of these themes, the intention is that the children will be given the opportunity to learn “what science is and how it is done” (McComas, 1998, p. 50).

Overall, the course was developed to foster two central aspects of the understanding of science: the development of adequate conceptions about the nature of knowledge and knowing (epistemic beliefs, as introduced in Chapter 1.1.1.) and the promotion of inquiry-based methodological competencies, which build the basis for the genesis, construction, and development of knowledge in science (as introduced in Chapter 1.1.2.).

To reach this goal, an important framework for the intervention was an inquiry-based learning approach (as introduced in Chapter 1.3.3.). Within IBSE, the intervention was based on the principle of a step-by-step unfolding of the inquiry process (Colburn, 2000). Furthermore, scientific work according to the SIC was another basic design principle of the intervention. The SIC was implemented in a step-by-step fashion and applied in all of the course sessions of the intervention. The sessions were arranged in such a way that the children experienced and applied the cumulative and cyclical process of scientific research. Finally, the transition from hands-on activities to reflection and thinking was the third basic design principle of the intervention (Aebli, 1980; Piaget, 1966). As described in Chapter 1.3.3., an explicit reflexive approach is important for promoting an adequate understanding of science (Akerson & Hanuscin, 2007). On this basis, the children conducted practical research projects and discussed and reflected on their findings afterwards. This process was intended to increase the level of abstraction and reflection in each individual course session as well as across the entire course (see Oschatz & Schiefer, in press).

Description of the individual course modules

The science intervention Little Researchers consisted of four modules, each consisting of one or two course sessions. They are described in the following (see Figure 5 for an overview). Overall, the course consisted of 10 weekly sessions of 90 minutes each.

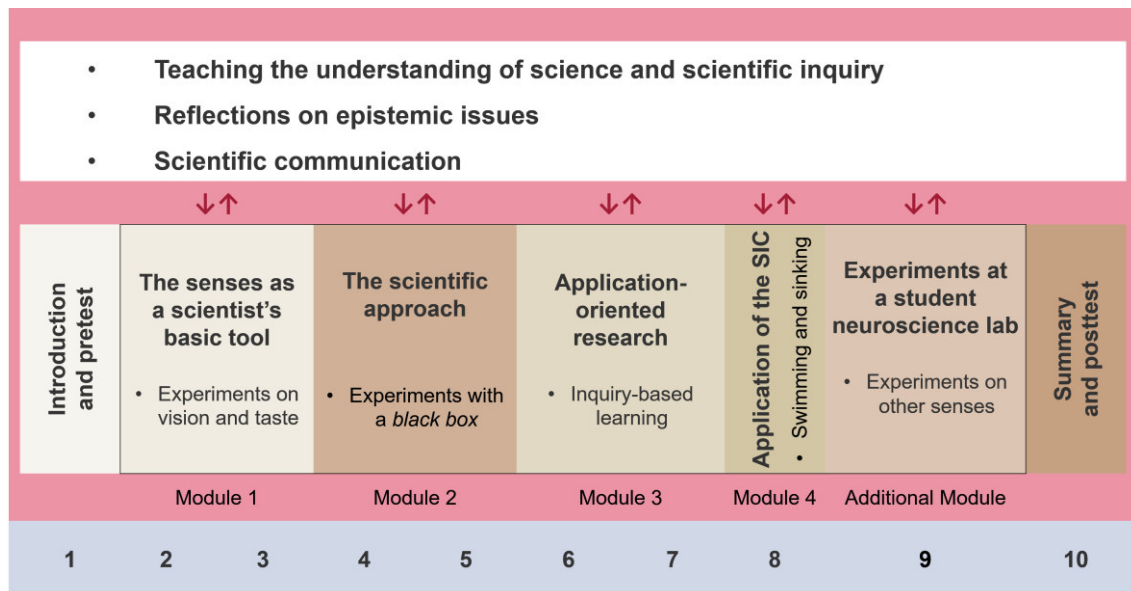


Figure 5. Concept of the intervention “Little Researchers – We Work like Scientists.”

Module 1: The senses as a scientist's basic tools

The first module focused on introducing scientific inquiry and experimentation. As described in Chapter 1.1.3., elementary school children often possess inaccurate misconceptions about the role of experiments. This role was addressed in this module (see Höttecke, 2001). By conducting experiments on the senses (which can be considered a scientist's most *basic tools*), the children learned about the functioning, the potential, as well as the boundaries of the human senses (e.g., biases, optical illusions) in the context of scientific work. In connection with this, it became understandable to the children that researchers might counter such biases and sources of error, for example, by repeating experiments, by precisely recording and documenting their results, or by exchanging information with other scientists.

Against the background of the instructional design principle of the step-by-step unfolding of the inquiry process, the experiments in the first module were firmly guided and clearly structured. Against the background of the second design principle, the inquiry cycle was introduced in the first module. In accordance with the third design principle—

the transition from doing to reflecting—the goal was that the children would apply the steps of inquiry and gain practical experience in experimentation, precise observation, and the documentation of results. In this context, the goal was that they would learn about the importance of repeating experiments and accurately documenting results.

Module 2: The scientific approach—Experiments with a “black box”

The second module was based on the framework of the first module and deepened as well as broadened its content (e.g., with regard to the forming and testing of hypotheses). Hypotheses define the object of research and guide the research process as well as data collection and interpretation. Hypotheses therefore play a central role in the context of scientific inquiry. In contrast to the research objectives of the first module, in the second module, the children investigated an unknown object, a so-called “Black Box” (Frank, 2005). By applying this method, the students were able to use a concrete model for scientific inquiry. They were presented similar black boxes that could not be opened. When the boxes were moved, they produced specific sounds. The students got to use different tools (e.g., their own senses, magnets, wire, radiographs) and had to figure out what produced the sounds (the “secret of the black box”). While doing so, they had to develop and test different hypotheses. As in “real science,” they were not able to look into their research object. They repeatedly discussed the results of their investigations in simulated “research congresses.” The following aspects of scientific work could be demonstrated and practiced with the black box (see Frank, 2005): (a) perceiving and describing phenomena, (b) formulating and testing hypotheses, (c) recording and documenting observations, (d) interpreting results and comparing results with the initial hypotheses, (e) social exchange and scientific communication.

Module 3: Application-oriented research – Inquiry-based learning

The third module focused on the further unfolding of the inquiry process (see Colburn, 2000) as well as on introducing a central methodological research strategy, the CVS (Chen & Klahr, 1999), which was introduced in Chapter 1.1.2. In this module, the children applied IBSE with an application-oriented research question. The goal was that they would come to understand that research in the natural sciences is intended not only to explain natural phenomena or to generate new theories but also to contribute to the tech-

nological progress of society (McComas, 1998; OECD, 2016). They were given the opportunity to deepen the previously introduced research strategies and conduct their research projects independently. To reach this goal, they worked on a problem from daily life (car safety—development of crash protection). In this task, they were able to test their ideas directly and use evidence to improve their constructions. In the third module, the course concept enabled a step-by-step unfolding of the inquiry process. The participants could test their problem-solving competence and use their creativity to solve the problem.

Module 4: Application of the inquiry cycle—Experiments on swimming and sinking

The fourth module focused on summarizing all course contents and repeatedly applying the steps of the SIC. On the basis of the problem on swimming and sinking (Why do certain things swim and others sink?), the children were able to apply the previously acquired research strategies and research methods (SIC and CVS). They were given the opportunity to use their observations and results directly to generate and distort hypotheses. Thereby, they experienced the social and communicative aspects of science, as the materials were prepared in such a way that the different research groups came up with contradictory results that they had to discuss (see Kienhues et al., 2008).

Additional module: Experiments in a student neuroscience lab

In addition to the described modules, in the course “Little Researchers,” the children were given the opportunity to visit the student Neuroscience lab at the University of Tübingen (CIN). The experiments addressed the human senses (vision), which had already been introduced and were then complemented by experiments on the human sense of touch as well as the electric senses of certain fish. In the laboratory, the children were given the opportunity to meet scientists and visit a “real” environment where research is conducted.

1.4. Research Questions of the Present Dissertation

The starting point for the current research was the importance of and the call for the effective promotion of elementary school students' understanding of science. The understanding of science is supposed to be relevant for students' science learning as well as to prepare them for their later participation in socioscientific issues as responsible citizens in a society that is determined by science and technology (Jones et al., 2015; OECD, 2016). Recently, the early promotion of the understanding of science as early as elementary school was in the focus of research and educational practice (e.g., Bendixen, 2016; EC, 2007) and was specifically addressed within the research presented here. The present dissertation focused on central questions revolving around the (a) measurement and (b) promotion of elementary school children's understanding of science. These questions are relevant as they are at the intersection of educational research, developmental and cognitive psychology, as well as natural science education and are derived in the following.

(A) Instruments are required to contribute to the description of children's development as well to assess the effectiveness of interventions (pre- and posttest comparisons). Existing measurement instruments and their limitations for assessing students' understanding of science were described in Chapter 1.2.2. To date, only a few paper-and-pencil tests exist, although they are required for group-testing situations (e.g., in large-scale studies or intervention studies). Existing instruments show partially limited reliability, scaling, or validity and focus only on specific aspects of the understanding of science. The present dissertation was aimed at extending previous research on measurement instruments and expanding the existing tests so that they could be used with elementary school children. Therefore, we focused on the development of a new instrument that could be used to assess elementary school children's understanding of the whole scientific inquiry cycle (SIC; see Chapter 1.1.2.) in Study 1 of this dissertation. The SIC is a central component of the understanding of science and inquiry-based learning approaches. So far, the steps of the SIC have usually been investigated and assessed independently of each another, although they are interdependent and interrelated (Wilhelm & Beishuizen, 2003). It can be assumed that the individual components (e.g., designing experiments, data interpretation) can be demonstrated or trained but are not sufficient to allow for targeted empirical research and reflection on this research (see Kuhn & Franklin, 2006). This strengthens the central role of the understanding of the SIC as a meta-perspective on

scientific inquiry. In the context of the development of the SIC test, a special emphasis was placed on meeting quality criteria as well as quality standards (Downing, 2006). First, the target construct and content of the test was defined and restricted to exist within the broad field of the understanding of science. Test administration modality (paper-and-pencil) was determined, the psychometric model (item response theory, see Embretson & Reise, 2013) was chosen, and the timeline of the development of the test was planned. This stage of clearly defining and planning a test is essential for ensuring its validity, as stated in the *Standards* of AERA, APA, and NCME (1999): “The validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction” (p. 17). Next, test specifications were made with regard to the choice of testing format, number of items and item format (single-choice items as well as sorting tasks that require active problem solving), test stimuli, item scoring rules, and time limit. Afterwards, items were developed and discussed several times with distinguished experts in the field of the understanding of science and cognitive development. A first version of the instrument was tested in a pilot sample of 10 elementary school children. Furthermore, think-aloud methods were used with three children to ensure the comprehensibility of the items. All test administrators participated in a mandatory training to ensure a competent, efficient, and standardized administration of the SIC test in school classes and in the STEM courses of the HCAP.

(B) In addition to questions regarding the measurement of elementary school children’s understanding of science, questions concerning the effective promotion of their understanding of science were addressed in this dissertation. As described in Chapter 1.3.3., a variety of science interventions have recently been developed. However, there is a lack of programs that have focused on the promotion of very fundamental aspects of the understanding of science such as general science methods or children’s epistemic understanding, especially at the elementary school level. To fill this gap, we developed a new 10-week intervention program for third and fourth graders that was intended to foster fundamental aspects of their understanding of science such as sophisticated epistemic beliefs and inquiry-based methodological competencies. In order to ensure their quality, it is recommended that programs are evaluated with high-quality designs (i.e., RCTs) at all stages between their development and their broad dissemination in practice (e.g., Humphrey et al., 2016 or Lendrum & Humphrey, 2012). To guarantee this, we focused on the

entire process of developing, evaluating, and implementing an intervention and investigated questions about the effectiveness of the program at different stages with high-quality designs. Studies 2 and 3 employed randomized controlled study designs, which are considered the gold standard in educational research (Torgerson & Torgerson, 2013). Study 2 investigated the effectiveness of the intervention under highly controlled conditions, whereas Study 3 investigated the effectiveness of the program when implemented by regular HCAP course instructors. As pointed out in Chapter 1.3.1., this crucial step of real-world implementation has rarely been evaluated and tends to fail or lead to reduced outcomes (Durlak & DuPre, 2008; Fixsen et al., 2013; Hulleman & Cordray, 2009; Spiel et al., 2016). However, the present dissertation enabled the evaluation of a program under real-world conditions. In this regard, questions regarding implementer characteristics as well as implementation fidelity were explored.

The target group of the intervention—third and fourth graders who were nominated to participate in an extracurricular enrichment program—was chosen for two reasons: First, there is a need for the fostering of students' understanding of science at the elementary school level (EC, 2007). Thus, there is a need for more interventions for younger children at this age level (e.g., EC, 2007; Valla & Williams, 2012). According to the high cognitive abilities of gifted children (see Chapter 1.3.4.), gifted children, in particular, can be expected to benefit from the intervention as they possess the cognitive prerequisites for reflecting on epistemic issues. Second, from an instrumental perspective, the promotion of gifted students has a high societal relevance as such students have the potential to show high achievements in the STEM domains as adults and might support science and economy (NSB, 2010).

Study 1 (*Scientific Reasoning in Elementary School Children – Assessment of the Inquiry Cycle*) presents the development, scaling, and construct validation of a new paper-and-pencil test for elementary school children in Grades 3 and 4. The instrument was designed to assess a central element of the understanding of science, the scientific inquiry cycle (see Chapter 1.1.2.). We investigated whether the newly developed items could be used to measure the understanding of the SIC reliably. IRT modeling and confirmatory factor analyses were used to investigate the (latent) item structure, model fit, and test reliability. Study 1 used data from a cross-sectional study with elementary school classes and participants in STEM enrichment courses at the HCAP. Furthermore, we explored the relations between the SIC test and cognitive abilities such as fluid and crystallized

intelligence, text comprehension, and experimentation design skills. In addition, we investigated how the SIC was related to epistemic beliefs in the domain of science.

Study 2 (*Fostering Epistemic Beliefs, Epistemic Curiosity, and Investigative Interests in Elementary School Children: A Randomized STEM Intervention Study*) examined the effectiveness of an extracurricular science intervention for elementary school students in Grades 3 and 4 (as described in Chapter 1.3.3.). In Study 2, the effectiveness of the program was investigated under highly controlled conditions—conducted by the program developers from the university—to reach a high level of internal validity and fidelity (according to the recommendation by Humphrey et al., 2016). The 10-week intervention included inquiry-based approaches to as well as reflections on epistemic issues. We explored whether the intervention affected participants’ epistemic beliefs (as a central element of the understanding of science) as well as their epistemic curiosity and investigative interests.

Study 3 (*Elementary School Children’s Understanding of Science: The Implementation of an Extracurricular Science Intervention*) focused on implementing the science intervention from Study 2 under real-world conditions. Using a larger sample and course instructors from the HCAP (teachers and course instructors with a background in the natural sciences), we investigated whether the intervention would still be effective in promoting participants’ epistemic beliefs. Furthermore, new instruments for assessing central elements of the understanding of science—which were not yet available at the time of the first effectiveness study—were used for the program evaluation. In this context, we investigated whether the newly developed SIC test could be successfully implemented for measuring students’ development. We also assessed characteristics of the course instructors and measured implementation fidelity.

The prior preparatory and conceptual work, goals, research questions, samples, and statistical analyses from the three empirical studies of the present dissertation are summarized in Table 3.

Table 3

Overview of the Goals, Research Questions, and Samples of the Three Empirical Studies of the Dissertation

<i>Study</i>	<i>Preparation & Study goals</i>	<i>Research questions</i>	<i>Sample & Statistical analyses</i>
Study 1	<p>Preparation: Development of an instrument</p> <p>Goal: Scaling and validation of the new instrument for measuring elementary school children's understanding of science</p>	<ol style="list-style-type: none"> 1. Can the understanding of the scientific inquiry cycle (SIC) be reliably and validly assessed with a new paper-and-pencil test for elementary school children? 2. What are the relations of the SIC test to cognitive abilities and epistemic beliefs in the domain of science? 	<p>$N = 878$ elementary school children ($N = 681$ from school classes, $N = 197$ from courses of the HCAP)</p> <p>IRT scaling, Model fit analyses, IFA (confirmatory item factor analyses) Correlation and multiple regression analyses</p>
Study 2	<p>Preparation: Development of an extracurricular science intervention</p> <p>Goal: Exploring the effectiveness of this extracurricular science intervention (under highly controlled conditions)</p>	<ol style="list-style-type: none"> 1. Are there effects of the intervention on the development of children's <ol style="list-style-type: none"> (a) Epistemic beliefs? (b) Epistemic curiosity? (c) Investigative interests? 	<p>$N = 65$ elementary school children (nominated for participation in the HCAP)</p> <p>$N = 3$ course instructors (scientists from the university, course developers)</p> <p>Multiple regression analyses</p>
Study 3	<p>Preparation: Writing of a course manual & development of a further training for course instructors</p> <p>Goal: Implementing the science intervention from Study 2 into practice and exploring its effectiveness (under real-world conditions)</p>	<ol style="list-style-type: none"> 1. Can the intervention be successfully implemented under real-world conditions by course instructors of the HCAP? 2. Are there intervention effects on the development of children's <ol style="list-style-type: none"> (a) Epistemic beliefs? (b) Inquiry-based methodological competencies? 3. Can the SIC-test (from study 1) be implemented successfully for the evaluation? 4. Will there be sufficient implementation fidelity? 	<p>$N = 117$ elementary school children (nominated for participation in the HCAP)</p> <p>$N = 10$ course instructors of the HCAP</p> <p>Multiple regression analyses</p>

Study 1:
Scientific Reasoning in Elementary
School Children: Assessment of the
Inquiry Cycle

Schiefer, J., Golle, J., & Oschatz, K. (2016). Scientific Reasoning in Elementary School Children: Assessment of the Inquiry Cycle. *Manuscript submitted for publication.*

This study was funded in part by the Hector Foundation II.

Abstract

Children's scientific reasoning skills are relevant for their science learning and their general understanding of the world around them. As there are hardly any paper-and-pencil tests for assessing elementary school children's scientific reasoning skills, the goal of the current study was to develop a new, reliable, and valid instrument for this age group. We focused on assessing children's understanding of the scientific inquiry cycle (SIC), which is a core element of scientific reasoning. 15 items were developed and applied in a sample of 878 third- and fourth-grade students. As confirmed by IRT modeling, the items formed a reliable scale. Furthermore, we explored the relation between children's SIC performances and their (meta)cognitive abilities. As expected, intelligence, text comprehension, experimentation strategies, and sophisticated epistemic beliefs in the domain of science were positively associated with children's SIC performance, a finding that contributes to the understanding of the construct validity of the SIC.

Keywords: scientific reasoning, inquiry cycle, development of an instrument, elementary school children, cognitive abilities, epistemic beliefs

Scientific Reasoning in Elementary School Children: Assessment of the *Inquiry Cycle*

Scientific reasoning can be broadly defined as knowledge seeking (Kuhn, 2002). Scientists and experts need scientific reasoning to be able to draw adequate conclusions in their research fields, and laymen need it to extend their knowledge of the world. Even in elementary school, children are already beginning to think scientifically (Kuhn, 2002; Zimmerman, 2007). It is assumed that scientific reasoning guides children's information seeking processes in scientific disciplines and facilitates their general understanding of the world. It supports conceptual change and science learning as well as the development of children's personal epistemology (see Kuhn, 2002; Morris, Croker, Masnick, & Zimmerman, 2012; Osborne, 2013). Due to the great importance of scientific reasoning for acquiring knowledge about the surrounding world, national and international education standards have identified scientific reasoning as a normative goal of students' science education (National Research Council, 1996; OECD, 2007).

The *scientific inquiry cycle (SIC)* is a core element of scientific reasoning (e.g., Klahr & Dunbar, 1988; Kuhn, 2002; White, Frederiksen, & Collins, 2009). In brief, the SIC includes the interrelated steps of (a) theorizing, (b) questioning and hypothesizing, (c) investigating, and (d) analyzing and synthesizing (White, Frederiksen, & Collins, 2009; Zimmerman, 2007). The understanding of these steps is essential for inquiry-based science learning approaches as well as for scientific reasoning and argumentation (Colburn, 2000; Kuhn, 2010; Kuhn & Dean, 2005).

At the intersection of cognitive development and science education, instruments for investigating scientific reasoning skills are required to describe children's competencies or to measure their progress in science learning. So far, there are hardly any paper-and-pencil tests that have been designed to measure the scientific reasoning abilities of elementary school children aged 8 to 10 years (e.g., Koerber, Mayer, Osterhaus, Schwippert, & Sodian, 2015; Mayer, Sodian, Koerber, & Schwippert, 2014).

In order to narrow this research gap, the goal of the present study was to develop a new paper-and-pencil test to assess young children's scientific reasoning skills. Thereby, we focused on measuring children's understanding of the scientific inquiry cycle. The instrument consists of 15 items (the exact structure of the SIC test will be described in the Method section) that were applied in a sample of 878 third- and fourth-grade students and scaled by IRT modeling in Mplus (Muthén & Muthén, 1998-2012).

To investigate the construct validity of the instrument, we explored the relation between children's SIC performance and their cognitive and metacognitive abilities.

Scientific Reasoning

Scientific reasoning includes “the skills involved in inquiry, experimentation, evidence evaluation, and inference that are done in the service of conceptual change or scientific understanding” (Zimmerman, 2007, p. 172). It involves a range of cognitive and metacognitive skills and is considered a cumulative and cyclical process that requires the coordination of theory and evidence (Kuhn, 2002; White, Frederiksen, & Collins, 2009). The goal of this cyclical process is the acquisition of knowledge or to produce change in already existing knowledge (see Kuhn, 2002). Scientific reasoning encompasses the ability to generate, test, and revise theories and hypotheses and to reflect on this process (Kuhn & Franklin, 2006; Zimmerman, 2007).

The SIC as a Core Element of Scientific Reasoning

The SIC can be considered the core element of scientific reasoning, and it has also served as a theoretical framework for many scientific reasoning models (e.g., scientific discovery as dual search [SDDS] model by Klahr, 2000; Kuhn, 2002; White & Frederiksen, 1998; Zimmerman, 2007). The scientific inquiry cycle includes the following steps: (a) the generation of hypotheses on the basis of a specific research question (derived from theory or a result of previous research), (b) the planning and conducting of experiments, (c) data collection, (d) analysis, (e) evaluation of evidence, and (f) the drawing of inferences. Thus, the SIC subsumes all individual components of scientific reasoning from a metaperspective and emphasizes a holistic view as the components build the basis of the cumulative and cyclical process of knowledge acquisition and change (Kuhn & Franklin, 2006; Zimmerman, 2007).

All of the steps of the SIC are arranged to represent a cycle, but as inferences from an experiment lead mostly to new research questions or hypotheses and the start of a modified inquiry process, they correspond more closely to a spiral (see Figure 1). Thereby, it should be noted that mature scientific inquiry does not necessarily proceed in the stepwise manner that is postulated (it is possible to start anywhere in the cycle) and that scientists do not necessarily proceed through these steps of inquiry in a fixed order. For instance, “analyzing data can lead to the need to do further investigation” (White,

Frederiksen, & Collins, 2009, p. 9). Nevertheless, the inquiry cycle—in which one starts with theorizing and questioning—is an effective initial model that can enable students to develop capabilities for inquiry and an understanding of its constituent processes (White & Frederiksen, 1998, 2005). Furthermore, this model represents the theory-driven deductive approach that is approved and applied by scientists in empirical investigations (see White, Frederiksen, & Collins, 2009).

So far, all steps in the SIC have usually been investigated and assessed independently of each other, although they are interdependent and interrelated (Wilhelm & Beishuizen, 2003). It can be assumed that the single components (e.g., designing experiments) can be demonstrated or trained but will not be sufficient to allow for targeted empirical research and reflection on this research (see Kuhn & Franklin, 2006). This strengthens the central role of the understanding of the SIC as a metaperspective on scientific reasoning.

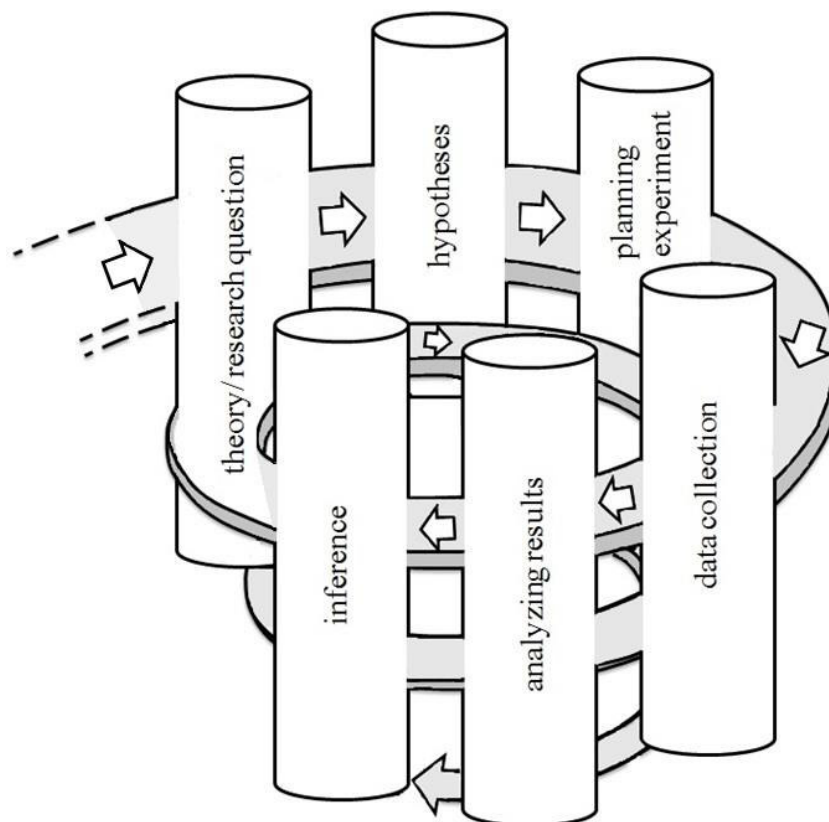


Figure 1. Steps of the scientific inquiry cycle (SIC), authors' own illustration (following Klahr & Dunbar, 1988; Kuhn, 2002; White & Frederiksen, 1998; White, Frederiksen, & Collins, 2009; Zimmerman 2007).

Scientific Reasoning in Elementary School Children

Traditionally, developmental psychologists have considered the scientific reasoning abilities of elementary school children to be deficient and have assumed that such skills emerge only during adolescence (Inhelder & Piaget, 1958). By contrast, developmental research within the last 20 years has provided evidence of children's early scientific reasoning competencies (see Bullock, Sodian, & Koerber; Zimmerman, 2007; Morris et al., 2012). Although elementary school children have trouble systematically designing controlled experiments, drawing appropriate conclusions on the basis of evidence, and interpreting evidence in general (Morris et al., 2012; Zimmerman, 2007), they do possess basic scientific reasoning skills. They are able to differentiate hypotheses from evidence, distinguish between a conclusive and an inconclusive experimental test, and do not confound the testing of hypotheses with the production of positive effects (e.g., Sodian, Zaitchik, & Carey, 1991).

Children's Assessment of Scientific Reasoning

Although a variety of task formats have been used to assess children's scientific reasoning skills, including interviews, self-directed experimentation tasks, simulations, or story problems (Bullock & Ziegler, 1999; Carey et al., 1989; Dunbar & Klahr, 1989; Kuhn et al., 1995; Schauble, 1996, for an overview, see Mayer et al., 2014), only a few paper-and-pencil tests have been developed. Questionnaires offer the simplest and most efficient way to measure abilities in group settings (e.g., school classes, science interventions), and educational research and practice has progressively focused on the development of reliable and valid questionnaires ever since national and international large-scale studies (e.g., PISA, TIMSS) have increased in importance. Developing paper-and-pencil measures to assess elementary school children's scientific reasoning skills poses a great challenge (e.g., due to children's limited reading capacities), which explains the apparent lack of instruments. Nevertheless, a recently developed instrument for fourth graders was used to assess different components of scientific reasoning (e.g., understanding theories, experimentation strategies, or data interpretation; Koerber et al., 2015; Mayer et al., 2014). Overall, the results indicated that fourth graders could be tested successfully with this instrument and that they displayed competence in different scientific reasoning components. However, the instrument was not able to assess their

understanding of the complete process of the scientific inquiry cycle (see Kuhn & Dean, 2005; White, Frederiksen, & Collins, 2009; Zimmerman, 2007).

Relations between the SIC and Other Constructs

Relations to Existing Scientific Reasoning Instruments

To determine the convergent validity of a new instrument, the use of already existing questionnaires is recommended (see Kline, 2015). As described above, there are hardly any scientific reasoning instruments for elementary school children that can be applied in group testing situations. The recently developed scientific reasoning instrument (Koerber et al., 2015; Mayer et al., 2014) had not been published when we conducted our study and was not available for the validation of the SIC test. Although no comprehensive scientific reasoning test was available for validation, commonly used scientific reasoning tasks could be applied to validate the SIC items. As research has often focused on experimentation strategies (see Chen & Klahr, 1999; Koerber et al., 2015; Zimmerman, 2007), we assessed such strategies alongside the new instrument. Because an understanding of the SIC and an understanding of experimentation strategies are associated with the scientific inquiry process, we expected a positive relation between the constructs.

Relations to Cognitive Abilities

To develop a valid measure of the understanding of the SIC, it is necessary to distinguish this competence from general cognitive abilities (e.g., intelligence or reading skills; Mayer et al., 2014), which are assumed to be necessary for the processing of (text-based) assessment tasks. There is evidence that scientific reasoning is positively related to these prerequisites (i.e., intelligence and reading comprehension) but can be measured separately from these cognitive abilities (Bullock et al., 2009; Mayer et al., 2014). To determine the discriminant validity of the understanding of the SIC, we assessed intelligence and reading comprehension as control variables (see Koerber et al., 2015; Mayer et al., 2014) and expected a positive relation between these constructs and the SIC.

Relations to Epistemic Beliefs

Beside cognitive abilities, metacognitive processes (e.g., an epistemological understanding or epistemic knowledge) are essential for scientific reasoning and inquiry

(e.g., Kuhn, 2002; Morris et al., 2012; Osborne, 2013; White, Frederiksen, & Collins, 2009) because the ability to engage in scientific reasoning requires a “meta-level knowledge of science of the epistemic features of science” (Osborne, 2013, p. 274). Epistemic beliefs are subjective beliefs about the nature of knowledge and knowing (Hofer & Pintrich, 1997). In line with Conley, Pintrich, Vekiri, and Harrisonx’s (2004) conceptualization, the four dimensions of *source*, *certainty*, *development* and *justification of knowledge* could be distinguished. According to the interdependence of scientific reasoning and epistemic beliefs, we investigated relations between the SIC and all dimensions of epistemic beliefs in the present study (see Osborne, 2013). Regarding the inquiry cycle, we expected that, in particular, sophisticated epistemic beliefs about the certainty, development, and justification of knowledge would be positively associated with the understanding of the SIC. Beliefs regarding these three dimensions refer to science as a changing and reversible discipline and might be a prerequisite for the understanding of the cyclical knowledge-seeking phases (inquiry, analysis, inference, and argument).

The Present Study

The goal of the present study was the development, scaling, and validation of a new paper-and-pencil test for assessing elementary school children’s scientific reasoning skills. Existing instruments have often focused on single aspects of scientific reasoning as experimentation strategies or on evaluating evidence (e.g., Chen & Klahr, 1999; Koerber et al., 2005; Koslowski, 1996). We focused on the assessment of the understanding of the complete scientific inquiry cycle (SIC), which is a core component of scientific reasoning that refers to a metaperspective on scientific reasoning.

To investigate the construct validity of the instrument, we investigated the relations between children’s performance on the SIC test and intelligence, reading comprehension, experimentation strategies, and epistemic beliefs. Two hypotheses guided our study. First, we hypothesized that the developed items would form a reliable scale and would measure elementary school children’s understanding of the SIC (Hypothesis 1). Second, we hypothesized that individual differences in SIC performance would be positively related to intelligence, reading skills, experimentation strategies, and sophisticated epistemic beliefs in the domain of science (Hypothesis 2).

Method

Participants and Experimental Design

The current study was based on data from 878 elementary school children in the third and fourth grades (57.4% male; 49.9% Grade 3; age: $M = 8.89$, $SD = 0.76$) in Germany. In a cross-sectional design, the SIC test and the validation instruments were assessed in 42 classes from 10 elementary schools ($n = 681$) by applying a rotational design with three versions of the questionnaires (all contained the SIC test and a combination of the following instruments: fluid and crystallized intelligence, text comprehension, epistemic beliefs, and experimentation strategies, see Table 1 for details).

Table 1

Overview of the Sample ($N = 878$) and the Multiple Booklet Design

	School classes booklet A	School classes booklet B	School classes booklet C	STEM courses booklet D
Duration of assessment	90 min	90 min	90 min	45 min
SIC test	yes	yes	yes	yes
Fluid intelligence	yes	yes	no	no
Crystallized intelligence	yes	yes	no	no
Text comprehension	yes	no	yes	no
Epistemic beliefs	no	yes	yes	no
Experimentation strategies	yes	no	yes	no
n	224	228	229	
N	681			197

Note. n = number of students who got booklets A, B, or C. N = total number of students who were tested in school classes or STEM courses. The instruments were arranged in the booklets in the presented order (from top to bottom).

Such multiple booklet designs are a common procedure in the context of large-scale assessments. This method allows representative testing of a variety of constructs to be applied but does not require any single child to answer the entire item set (see Koerber et al., 2015). Within the participating schools, the booklets were randomly assigned to classrooms. The SIC test was additionally assessed in 36 extracurricular STEM courses ($n = 197$) for third and fourth graders. Prior to testing, we obtained parents' written consent for their child's participation. Each measure was administered in a group testing situation by a trained instructor. Data collection in schools took 90 min and included the assessment of the SIC test and a choice of validation instruments. Data collection in STEM courses was limited to 45 min and included only the assessment of the SIC test. Data were collected in November and December, 2014.

Instruments

SIC test. The instrument focused on the assessment of the understanding of the complete scientific inquiry cycle. The developed tasks required (a) the active reconstruction of the sequences of all steps of the SIC and (b) an understanding of the consecutive next steps of the cycle within a given inquiry process (see White & Frederiksen, 1998; White, Frederiksen, & Collins, 2009). We believe that an understanding of all steps is required for an understanding of the complete inquiry cycle and that partial solutions do not indicate an understanding of the SIC. Furthermore, all of the steps are related to, interdependent with, and interact with each other (Klahr & Dunbar, 1988; Wilhelm & Beishuizen, 2003), relations that correspond to a holistic approach to the scientific inquiry process. Therefore, we postulated that the understanding of the SIC would be represented as a one-dimensional construct.

The SIC items were developed according to generally recommended procedures (e.g., Downing & Haladyna, 2006). Prior to application, items were repeatedly discussed with four distinguished experts in the field of scientific reasoning. The practicability and comprehensibility of the items were tested in a pilot phase with $N = 10$ third and fourth graders enrolled in an extracurricular STEM enrichment program. After revision, the items were tested again with three children using think-aloud techniques during processing (e.g., to detect possible problems in understanding the instructions or the handling of the presented materials; see Fonteyn, Kuipers, & Grobe, 1993).

The final SIC test consisted of 15 items that were dichotomously scored (0 = wrong answer, 1 = correct answer). Two different response formats were used to assess the understanding of the inquiry process: (a) sorting the steps of the SIC (three items) and (b) selecting the next step in the SIC (12 items).

(A) Three items required the sorting of the single steps of the inquiry cycle via printed labels [finding the research question (1); generating hypotheses (2); planning an experiment (3); conducting an experiment/collecting data (4); analyzing results (5); inference (6)]. Each of these three tasks required the active reconstruction of a different inquiry process. The respective research issue was introduced to the children in a short paragraph. Two out of three items were presented in a concrete everyday context (e.g., “Tom wants to find out whether his new pet has a sensitive sense of smell. How can he investigate this like a scientist?”). The third item was presented in a general context (“How do scientists proceed when they want to investigate something?”). The six single inquiry steps were read aloud to the children in a random order by the test instructors (to compensate for potential disadvantages due to some children’s poor reading abilities). Afterwards, printed labels that contained the six inquiry steps that were read to the children previously were given to them. They had to put the steps in the right order by sticking them in their questionnaire (see Figures 2 and 3 in the Appendix). The starting point—finding a research question—was given to the children. Only completely accurate solutions counted as correct because partial solutions do not indicate an understanding of the entire SIC.

(B) Twelve items were single-choice items, and the children were asked to select the respective next step in the inquiry cycle within a given inquiry process. The first six items referred to an everyday life context (e.g., “Mr. Abendstern is a famous scientist and knows exactly how a scientist has to work. He is interested in what causes tooth decay and wants to find out more about it”), the second six items referred to a general context (e.g., “Mrs. Morgenstern is a famous scientist and knows exactly how a scientist has to work”—without specifying a research topic). Children were told that they would be asked about the different working steps of the scientists. These questions and the respective working steps were not described in the correct consecutive stepwise order in which the scientist would do them. This procedure was chosen to avoid dependencies and cross-links between the children’s answers. Each of the questions referred to one of the six steps of the inquiry cycle, and the possible next steps were offered in a random order (e.g., Mrs.

Morgenstern has a hypothesis she wants to check. What is her next working step? [a] She performs an experiment, [b] She plans an experiment to verify her hypothesis, [c] She evaluates results from an experiment). The response options of these single-choice items referred either to the next step in the inquiry cycle (correct answer: [b] in the example) or to two randomly selected other steps (wrong answers: distractors [a] and [c] in the example). Every page of the questionnaire contained only one item, and the children were not allowed to move backwards through the questionnaire to correct answers they had already given. Example items for both response formats can be found in Figures 2, 3 and 4 in the Appendix.

Reading comprehension. Reading comprehension was assessed with the “text comprehension” subtest of the standardized German reading proficiency test ELFE1-6 (Lenhard & Schneider, 2006). Each of the 20 single-choice items consisted of a small section of text, a question, and four answer alternatives. Children had to choose between the right answer and three distractors. All answers were explicitly mentioned in the text. Children had 7 min to read the texts and answer the items. Sum scores were used in further analyses (Cronbach’s $\alpha = .87$).

Intelligence. Intelligence was measured with the BEFKI-short (Schipolowski, Wilhelm, & Schroeders, 2013). The first subscale fluid intelligence consisted of 16 items. Within a time limit of 15 min, children had to select two figures that completed a series of figural patterns ($\alpha = .67$). The second subscale crystallized intelligence consisted of 16 single-choice items and included questions about general knowledge. Within a time limit of 8 min, children had to choose one out of five answer alternatives ($\alpha = .64$). Sum scores were calculated separately for the two subscales and used in further analyses.

Experimentation strategies. Experimentation strategies (focusing on the control of variables strategy; Chen & Klahr, 1999) were assessed with six single-choice items with three answer alternatives (one correct, two misconceptions). Three items were taken from the unpublished scientific reasoning scale (Koerber et al., 2015; Mayer et al., 2014), and three items were newly developed in the same format. The items were presented in everyday-life contexts designed to assess domain-general experimentation skills (Mayer et al., 2014). The items were dichotomous (1 = correct answer, 0 = wrong answer). Sum scores were used in further analyses ($\alpha = .44$).

Epistemic beliefs. We assessed science-related epistemic beliefs with a 26-item instrument (Conley et al., 2004, adapted from previous work by Elder, 2002, translated

by Urhahne & Hopf, 2004). The four subscales include the dimensions: source, certainty, development, and justification of knowledge. Items are rated on a 4-point Likert scale. The scale source of knowledge (5 items, $\alpha = .59$) addresses beliefs about knowledge residing in external authorities (e.g., “Everybody has to believe what scientists say”). Certainty of knowledge (6 items, $\alpha = .58$) refers to a belief in a right answer (e.g., “All questions in science have one right answer”). Development of knowledge (6 items, $\alpha = .56$) measures beliefs about science as an evolving and changing subject (e.g., “Sometimes scientists change their minds about what is true in science”). Justification of knowledge (9 items, $\alpha = .65$) addresses the role of experiments and how individuals justify knowledge (e.g., “Good answers are based on evidence from many different experiments”). The source and certainty scales were recoded so that for each of the scales, higher scores reflected more sophisticated beliefs (a higher negation of the source and the certainty items).

Statistical Analyses

Initial item analyses were computed to explore item characteristics (means, standard deviations, item selectivity). We applied IRT modeling in Mplus to scale children’s test scores with a two-parameter logistic model for dichotomous items (2PL model; Birnbaum, 1968; Muthén & Muthén, 1998-2012). We used a maximum likelihood estimator with robust standard errors (MLR), which uses a numerical integration algorithm (Muthén & Muthén, 1998-2012). To correct for the clustering of the data (children nested in classes), we used type = complex for all analyses (Muthén & Muthén, 1998-2012). We applied confirmatory item factor analyses (IFA) using structural equation modeling (WLSMV estimator) to test the model fit and the postulated one-dimensional latent factor structure of the 2PL model (Birnbaum, 1968). We computed the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), the root mean square error of approximation (RMSEA), χ^2 , p-value, and the χ^2/df ratio (see Schermelleh-Engel, Moosbrugger, & Müller, 2003, for recommendations for model evaluation¹). To investigate the relations between children’s performance on the SIC test (EAP

¹ A nonsignificant p -value for the χ^2 statistic indicates that the tested model should be retained. A χ^2/df ratio between 0 and 2 reflects a good model fit and a ratio between 2 and 3 an acceptable fit (see Schermelleh-Engel et al., 2003, p. 52). CFI and TLI values greater than .90 and .95 (on a scale from 0 to 1) are typically taken to reflect acceptable and excellent fits to the data, respectively (Hu & Bentler, 1999). An RMSEA of less than .06 is typically taken to reflect a reasonable fit (Chen, Curran, Bollen, Kirby, & Paxton, 2008).

parameters) and intelligence, text comprehension, experimentation strategies, and epistemic beliefs (Hypothesis 2), correlation and multiple regression analyses were calculated. All variables were z-standardized prior to the analyses.

Missing data. In our study, there were hardly any missing values on the SIC items. Missing data ranged from 0% to 1.03%. Due to the multiple booklets design, the amount of planned missing data on the other scales ranged from 33.2% to 34.5%. We used the full information maximum likelihood approach implemented in Mplus to deal with missing data. To estimate the model parameters, this approach takes into account all variables from the respective models (see Schafer & Graham, 2002).

Results

Initial Item Analyses

The initial item analyses (means, standard deviations, selectivity) are presented in Table 2. As our items were scored dichotomously (1 = *correct*, 0 = *incorrect*), the means of the initial item analyses represent the solving frequencies ($.28 \leq M \leq .82$). None of the items were solved by zero or all of the children. Items were neither too difficult nor too easy. Items had sufficient variance ($.38 \leq SD \leq .50$). Initial analyses of the item selectivity (correlations between the items and the test score) revealed that three items (Items 4, 8, and 12) had values close to zero. These items were excluded from further analyses and the subsequent scaling of the items.

Scaling of the SIC Items

To address our first research question, we scaled the items with the Birnbaum (1968) measurement model as a 2PL model for dichotomous items. Model fit information criteria are summarized in Table 3. We applied confirmatory item factor analyses (IFAs) using structural equation modeling (WLSMV estimator) in Mplus to test the model fit and the postulated one-dimensional latent factor structure. This procedure tests whether observed item responses can be explained by a single continuous latent trait (see Koerber et al., 2015). The model fit of the 2PL model revealed acceptable results (RMSEA = .035;

$\chi^2/df = 2.10$, $\chi^2(54) = 113.662$, $p < .000$, CFI = 0.89, TLI = 0.86).² The SIC items showed an acceptable overall marginal EAP reliability of .64. Values above .70 can be described as good, values above .60 can be described as acceptable (comparable to Cronbach's alpha; see Field, 2009; Koerber et al., 2015). The EAP reliability is an estimate of test reliability that is obtained by dividing the variance of the individual expected a posteriori ability estimates (EAPs) by the estimated total variance of the latent ability (Kim, 2012). The item properties (difficulty, discrimination) for the 2PL model are summarized in Table 4.³

Table 2

Results of the Initial Item Analyses (Sorted from Most Difficult to Least Difficult)

Item	<i>N</i>	<i>M</i>	<i>SD</i>	<i>r.cor</i>
8	875	.28	.45	.04
12	869	.28	.45	.06
14	874	.32	.47	.17
2	878	.33	.47	.26
4	872	.35	.48	.10
3	878	.35	.48	.37
11	874	.49	.50	.40
6	873	.57	.50	.21
13	870	.61	.49	.41
1	878	.62	.49	.27
7	875	.66	.48	.45
5	870	.67	.47	.45
10	874	.71	.45	.33
15	872	.79	.41	.38
9	869	.82	.38	.45

Note. *r.cor* = correlation between item and test score.

² A closer look at the residuals of the latent response variables revealed that some residuals (Item 13 with Item 14; Item 2 with Item 3) were correlated and might therefore have contributed to the selective misfit of the model. Analyzing the model fit with modifications (correlations of the residuals) revealed a very good fit of the 2PL model: RMSEA = .023; $\chi^2/df = 1.45$, $\chi^2(52) = 75.205$, $p = .019$, CFI = .96, TLI = .94.

³ Results of the differential item analyses (see Holland & Wainer, 2012) revealed no differential item functioning (DIF; item difficulty, item discrimination) for gender (boys vs. girls), intelligence, grade level (Grade 3 vs. 4), or age.

Table 3

Model Fit Information for the 2PL Model

Number of free parameters	24
Log-likelihood (LL)	-6262.92
Scaling correction factor for MRL	1.6314
AIC	12573.84
BIC	12688.50
Sample-size adjusted BIC	12612.28
Deviance (-2*LL) (df)	12525.84 (4044)

Note: AIC = Akaike's Information Criterion. BIC = Bayes Information Criterion.

Table 4

Item Properties for the 2PL Model (Items Ordered from Most Difficult to Least Difficult)

Item	Item format	Percent correct (%)	Difficulty	Discrimination
14	SC	31.7	1.724	0.468
2	ST	33.3	1.073	0.723
3	ST	34.6	0.748	1.033
11	SC	48.7	0.060	0.962
6	SC	56.6	-0.566	0.491
13	SC	60.6	-0.520	0.980
1	ST	61.6	-0.864	0.592
7	SC	65.6	-0.758	1.040
5	SC	66.8	-0.722	1.260
10	SC	71.3	-1.242	0.837
15	SC	79.0	-1.539	1.035
9	SC	82.0	-1.488	1.040

Note. SC = single-choice item; ST = sorting task; $N = 878$; EAP reliability = .64

***Relations between SIC Performance, Cognitive Abilities, Experimentation Strategies,
and Epistemic Beliefs***

To address the second research question, correlations were calculated between SIC performance (latent EAP estimates), cognitive abilities, experimentation strategies, and epistemic beliefs (see Table 5). Apart from source of knowledge, all variables were positively correlated with SIC performance. Correlation coefficients ranged from .17 for development of knowledge to .49 for text comprehension (all $ps < .01$). Text comprehension, experimentation strategies, and fluid and crystallized intelligence had the highest positive relations with the SIC test. This means that children with a higher level of text comprehension, experimentation strategies, and intelligence scored higher on the SIC test than children with lower scores on those scales. Correlations between the SIC test and epistemic beliefs were low to moderate. Children with more sophisticated beliefs about certainty, development, and justification of knowledge scored higher on the SIC test than children with less sophisticated beliefs in these dimensions.

In a second step, we computed hierarchical multiple regression models to predict SIC performance (see Table 6). The predictors were the z-standardized measures of text comprehension, crystallized intelligence, and fluid intelligence. The dependent variable was the z-standardized SIC EAP score in all models. Results for the first model (including cognitive abilities) revealed that text comprehension ($\beta = .28, p < .001$), crystallized intelligence ($\beta = .18, p = .019$), and fluid intelligence ($\beta = .23, p < .001$) were positively associated with SIC performance and explained 30% of its variance. When holding all other variables constant, children with a higher level of text comprehension, fluid intelligence, and crystallized intelligence performed better on the SIC test than children with a lower level of text comprehension or fluid intelligence. The results for Model 2 (adding the z-standardized measures of experimentation strategies to the first model) indicated that experimentation strategies ($\beta = .28, p < .001$) were also positively associated with SIC performance. When holding all other measures of cognitive abilities constant, children with a good understanding of experimentation strategies performed better on the SIC test than children with a lower understanding of experimentation strategies.

The third model predicted SIC performance from epistemic beliefs. Predictors were the z-standardized measures of the four dimensions: source, certainty, development, and justification of knowledge. Results showed that sophisticated beliefs about certainty

of knowledge ($\beta = .25, p = .001$) and justification of knowledge ($\beta = .21, p < .001$) were positively associated with SIC performance. When holding all dimensions of epistemic beliefs constant, children with more sophisticated stances on certainty and justification of knowledge performed better on the SIC test than children with less sophisticated beliefs in those scales.

Considering the whole model (including all z-standardized variables from Models 1, 2 and 3), text comprehension ($\beta = .20, p < .001$), fluid intelligence ($\beta = .18, p = .002$), experimentation strategies ($\beta = .28, p < .001$), and sophisticated beliefs about certainty of knowledge ($\beta = .18, p = .001$; beliefs about *less* certainty, see reversal of items) were positively associated with SIC performance and explained 38% of the variance in SIC performance. When controlling for all variables, the effects of crystallized intelligence as well as justification of knowledge decreased to nonsignificant values.

Table 5
Correlation Coefficients, Means (M), Standard Deviations (SD), and Intra-class Correlations (ICCs) between All Measures

Construct	M	SD	ICC	N	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Scientific inquiry (SIC) ^a	0.00	.80	.16	878	1								
(2) Intelligence - fluid	8.00	2.91	.03	446	.40***	1							
(3) Intelligence - crystallized	10.32	2.86	.14	451	.43***	.39***	1						
(4) Text comprehension	13.26	4.03	.20	450	.49***	.38***	.56***	1					
(5) Experimentation strategies	3.07	1.49	.16	452	.46***	.32***	.45***	.46***	1				
(6) Source of knowledge ^b	2.43	.60	.11	456	.06	.01	.04	.05	.06	1			
(7) Certainty of knowledge ^b	2.46	.57	.09	455	.20**	.08	.10*	.09	.13*	.52***	1		
(8) Development of knowledge	3.12	.53	.02	455	.17**	.24***	.23***	.18**	.22***	-.06	-.02	1	
(9) Justification of knowledge	3.40	.41	.09	456	.22***	.22***	.32***	.31***	.16**	-.16*	-.12*	.40***	1

^aItems were scaled by a 2PL model for dichotomous items. The mean of the latent factor was fixed to 0.

^bItems were reversed so that for these scales, higher scores reflected more sophisticated beliefs (a strong negation of the source and certainty items).
^{*} $p < .05$. ^{**} $p < .01$. ^{***} $p < .001$. (two-tailed)

Table 6

Regression Models for Predicting the SIC Performances (EAP scores)

Model	M1		M2		M3		M4	
	β	SE	β	SE	β	SE	β	SE
Text comprehension ^a	.28***	(.05)	.20***	(.05)			.20***	(.05)
Intelligence - crystallized ^a	.18*	(.08)	.11	(.09)			.09	(.08)
Intelligence - fluid ^a	.23***	(.05)	.19**	(.06)			.18**	(.06)
Experimentation strategies ^a			.28***	(.07)			.28***	(.07)
Source of knowledge ^a					-.03	(.06)	-.06	(.06)
Certainty of knowledge ^a					.25**	(.07)	.18**	(.05)
Development of knowledge ^a					.10	(.06)	-.03	(.04)
Justification of knowledge ^a					.21***	(.05)	.06	(.05)
R ²	.30		.36		.12		.38	

Note. $N = 878$. Standardized coefficients are reported.

^aVariables were z-standardized prior to analyses.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This study focused on the development, scaling, and validation of a new instrument for assessing the understanding of the scientific inquiry cycle for elementary school children. The SIC can be considered a core element of scientific reasoning that had not previously been assessed in this age group. The items required students to reconstruct the associated sequences and to answer single-choice questions that referred to consecutive steps of the SIC.

Scaling of the SIC Items

The results of the 2PL scaling indicated that our instrument could reliably measure the understanding of the inquiry cycle in elementary school age children (Hypothesis 1). The results of the item response modeling indicated that the items formed a reliable and feasible scale and fulfilled—despite the selective misfit of the model—the overall psychometric affordances of a dichotomous 2PL model (Birnbaum, 1968). The SIC test had an EAP reliability of .64, which is comparable to the reliability of existing scientific reasoning tests at an elementary school level (Koerber et al., 2015; Mayer et al., 2014).

Relations between SIC Performance, Cognitive Abilities, Experimentation Strategies, and Epistemic Beliefs

To investigate the construct validity of the SIC test, we analyzed relations between the children's SIC performance and their cognitive and metacognitive abilities (Hypotheses 2). The results of the correlation analyses suggested that besides the components of verbal and nonverbal intelligence, experimentation strategies and sophisticated epistemic beliefs were associated with SIC performance. The results were consistent with the expectation that cognitive ability constructs would be related to children's SIC performance. Due to the moderate relation, it can be concluded that SIC performance can be differentiated from cognitive ability variables as a separate construct, thus pointing to discriminant validity (Kline, 2015). This differentiation is important as reading comprehension and intelligence generally play an important role in the processing of written test instruments and influence test performance, especially at the elementary school level (Koeppen et al., 2008).

The correlations of the SIC test with epistemic beliefs corresponded with our prediction that sophisticated beliefs about the nature of knowledge and knowing (see

Hofer & Pintrich, 1997) would be positively associated with students' understanding of the SIC. The results provide empirical evidence for the importance of metacognitive processes and metacognitive knowledge for the completion of scientific reasoning tasks (e.g., Kuhn, 2002; Morris et al., 2012). In line with our expectations, sophisticated epistemic beliefs about the certainty, development, and justification of knowledge were positively correlated with children's SIC performance. These epistemic beliefs refer to science as a changing and reversible discipline (Conley et al., 2004) and might be prerequisites for the understanding of the cyclical and cumulative knowledge-seeking phases (Koslowski, 1996; Kuhn, 2002; Kuhn & Franklin, 2006). The goal of these inquiry phases is not the separate collection of knowledge but the ongoing and continuing generation, testing, and revision of theories and hypotheses. An understanding of this process of knowledge acquisition and change requires an understanding of the need for change and the constant further development of scientific knowledge.

The results of the multiple regression analyses revealed that text comprehension, fluid intelligence, experimentation strategies, and sophisticated epistemic beliefs about the certainty of knowledge explained most of the variance in children's performance on the SIC test. Besides cognitive abilities and experimentation strategies, further developed beliefs about the certainty of knowledge were positively associated with SIC performance. Due to the recoding of the items (see Conley et al., 2004), this indicates that beliefs in a high level of uncertainty in scientific knowledge are positively associated with SIC performance.

Implications

Our results have important implications for educational research and practice. First, the present study showed that the items on the new paper-and-pencil instrument form a reliable scale and can be used to assess third- and fourth-grade students' understanding of the inquiry cycle. The SIC test complements and broadens existing tasks and tests for elementary school children (e.g., Bullock & Ziegler, 1999; Koerber et al., 2015; Mayer et al., 2014). It assesses a metaperspective on scientific reasoning, which had yet to be investigated.

Second, our study demonstrated that 8-to-10-year-old children were competent in solving the SIC tasks and had an understanding of the process of scientific inquiry. Thus, they might possess an early understanding of the deductive hypothesis-driven process of

knowledge seeking and change (Kuhn & Franklin, 2006; Zimmerman, 2007). Our findings strengthen the recommendation that educators should incorporate scientific inquiry methods into science curricula (see education plans; National Research Council, 1996). As there is evidence that already elementary school children can understand the processes and goals of inquiry, they might under the guidance of a teacher be able to plan, conduct, and interpret experiments independently (see Colburn, 2000).

Limitations and Future Research

Although our study demonstrated that the understanding of the SIC can be reliably and validly measured in elementary school children, some limitations should be considered when interpreting the results. First, our study was narrowly focused on a specific sample of third and fourth graders and investigated their competences on the SIC test with a cross-sectional design. It might be promising to administer our instrument to younger or older age groups or to investigate students' achievement on the SIC test longitudinally. According to Mayer et al. (2014), this might provide further insight into "age differential developmental changes [...] and the early impact of prerequisites contributing to development" (p. 50). Given that so far there are also no instruments for adults, it might be promising to assess the SIC competences of university students or adults. This might provides insight into whether, for example, trainee teachers or laypeople have an understanding of the inquiry process.

Second, the limited reliability and the selective misfit of the SIC items should be taken into account. It might be promising to develop more items for each of the steps of the inquiry cycle to increase the reliability of the instrument (Kline, 2009). However, it has to be noted that the reliability of the SIC test is in line with the reliability of other scientific reasoning tests for 8-to-10-year-old children (Koerber et al., 2015; Mayer et al., 2014). Furthermore, the existence of more items might enable a detailed analysis of intermediate steps in the inquiry cycle. Although we considered the understanding of the inquiry cycle from a holistic point of view, it might be promising to gain insight into children's difficulties with specific steps of the inquiry cycle or their solution strategies. It can be assumed that certain sequences (e.g., analyzing data after data collection) might be easier for some children than others (e.g., closing the inquiry cycle by starting a new phase of inquiry after drawing inferences). On the basis of children's response patterns, it might be possible to reconstruct how the understanding of the SIC develops.

Third, further validation of the SIC items with other instruments or tasks is needed. It might be interesting to investigate the relation between the SIC test and existing scientific reasoning tasks that were not available when we conducted our study (e.g., Koerber et al., 2015; Mayer et al., 2014). Validation with further cognitive variables that have been found to be related to scientific reasoning (e.g., problem-solving, inhibition, spatial abilities; see Mayer et al., 2014), metacognitive processes (e.g., planning skills, strategy use), motivational factors, or sociocultural background might be promising for explaining individual differences. Furthermore, validation with science grades as well as hands-on scientific practices might be important for investigating the criterion validity of the SIC test.

Conclusion

Taken together, the present study showed that it is possible to assess elementary school children's understanding of the scientific inquiry cycle with a paper-and-pencil test. This broadens the pool of existing instruments for measuring scientific reasoning in this age group. In addition to recent studies (e.g., Koerber et al., 2015; Mayer et al., 2014), the present investigation provides further empirical evidence that elementary school children are competent in scientific reasoning. The presented SIC test can assess the core of scientific reasoning and might be implemented in future research for assessing elementary school children's scientific reasoning competencies or the measurement of educational progress in the area of science learning (e.g., within the evaluation of science interventions).

References

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley.
- Bullock, M., Sodian, B., & Koerber, S. (2009). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. *Human development from early childhood to early adulthood: Findings from the Munich longitudinal study*, 20, 173-198.
- Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich longitudinal study* (pp. 38-54). New York, NY: Cambridge University Press.
- Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). An experiment is when you try it and see if it works. A study of junior high school students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11, 514-529.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods and Research*, 36(4), 462-494.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 1098-1120.
- Colburn, A. (2000). An inquiry primer. *Science Scope*, 23(6), 42-44.
- Conley, A. M., Pintrich, P. R., Vekiri, I., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology*, 29, 186-204.
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. *Complex information processing: The impact of Herbert A. Simon*, 109-143.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage publications.
- Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research*, 3(4), 430-441.


- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88-140.
- Holland, P. W., & Wainer, H. (Eds.). (2012). *Differential item functioning*. New York: Routledge.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Kim, S. (2012). A note on the reliability coefficients for item response model-based ability estimates. *Psychometrika*, 77(1), 153-162.
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1-48.
- Kline, P. (2015). *A Handbook of Test Construction (Psychology Revivals): Introduction to Psychometric Design*. London: Routledge.
- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie/Journal of Psychology*, 216(2), 61-73.
- Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development*, 86(1), 327-336.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16(11), 866-870.
- Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how). In W. Damon & R. M Lerner (Eds.), *Handbook of Child Psychology* (pp. 517-550), New York: Wiley.

- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 371-393). Oxford: Blackwell.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, 94(5), 810-824.
- Kuhn, D., Garcia-Mila, M., Zohar, A., Andersen, C., White, S. H., Klahr, D., & Carver, S. M. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, 60(4), 1-128.
- Lazonder, A. W., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction*, 22(6), 458-464.
- Lenhard, W., & Schneider, W. (2006). *Ein Leseverständnistest für Erst- bis Sechstklässler: ELFE 1-6 [A reading comprehension test for first to sixth graders]*. Göttingen: Hogrefe.
- Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, 29, 43-55.
- Morris, B. J., Croker, S., Masnick, A. M., & Zimmerman, C. (2012). The emergence of scientific reasoning. In Kloos, H., Morris, B. J., & Amaral, J. L. (Eds), *Trends in Cognitive Development*, (pp. 61-82). Rijeka, Croatia: InTech.Croatia, InTech Publishing.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.
- National Research Council (Ed.). (1996). *National science education standards*. Washington, D.C.: National Academy Press.
- OECD – Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (2007). *PISA 2006: Naturwissenschaftliche Kompetenzen für die Welt von Morgen*. URL: <http://www.oecd.org/dataoecd/59/10/39731064.pdf> (05.08.2009).
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265-279.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32(1), 102-119.

- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.
- Schipolowski, S., Wilhelm, O., & Schroeders, U. (2013). BEFKI GC-K. Berliner Test zur Erfassung fluider und kristalliner Intelligenz – GC-Kurzskala [BEFKI GC-K. Berlin test for assessing fluid and crystallized intelligence – short version]. In C. J. Kemper, E. Brähler & M. Zenger (Hrsg.), *Psychologische und sozialwissenschaftliche Kurzskalen – Standardisierte Erhebungsinstrumente für Wissenschaft und Praxis* (S. 30-34) [*Short scales in psychology and the social sciences – Standardized assessment instruments for research and practice*]. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Urhahne, D., & Hopf, M. (2004). Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien [Epistemological beliefs in science and their relations to motivation, self-concept, and learning strategies]. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 71-87.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3-118.
- White, B., & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, 40(4), 211-223.
- White, B., Frederiksen, J., & Collins, A. (2009). The interplay of scientific inquiry and metacognition: More than a marriage of convenience. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 175–205). New York, NY: Routledge.
- Wilhelm, P., & Beishuizen, J. J. (2003). Content effects in self-directed inductive learning. *Learning and Instruction*, 13(4), 381-402.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172-223.

Appendix

Tom has a new pet, an aardvark. He doesn't know much about aardvarks and wants to find out more about them. How should Tom proceed best if he wants to work like a scientist?



These are his working steps. They are not in the correct order.

He is considering an attempt to find out if the aardvark can detect sweet smells better than his dog.	He gathers information and reads in an animal encyclopedia that aardvarks have a good nose for sweet smells. He wants to check this.	He hides a small piece of cake in the garden and lets the dog and the aardvark look for the cake 10 times alternately.	He has evidence that the aardvark has a keener sense of smell for sweet smells than his dog.
He calculates how long each animal needed to find the cake. The aardvark was faster than the dog.	He wants to find out more about aardvarks' good noses and gathers information about whether aardvarks have a good nose for musty smells, too.	He is interested in aardvarks and wants to find out more about them.	

→ Put the working steps in the right order. The first step is already predetermined.

→ Glue the cards only if you are sure about the order you want to put them in.

Figure 2. Item example (sorting task part 1).

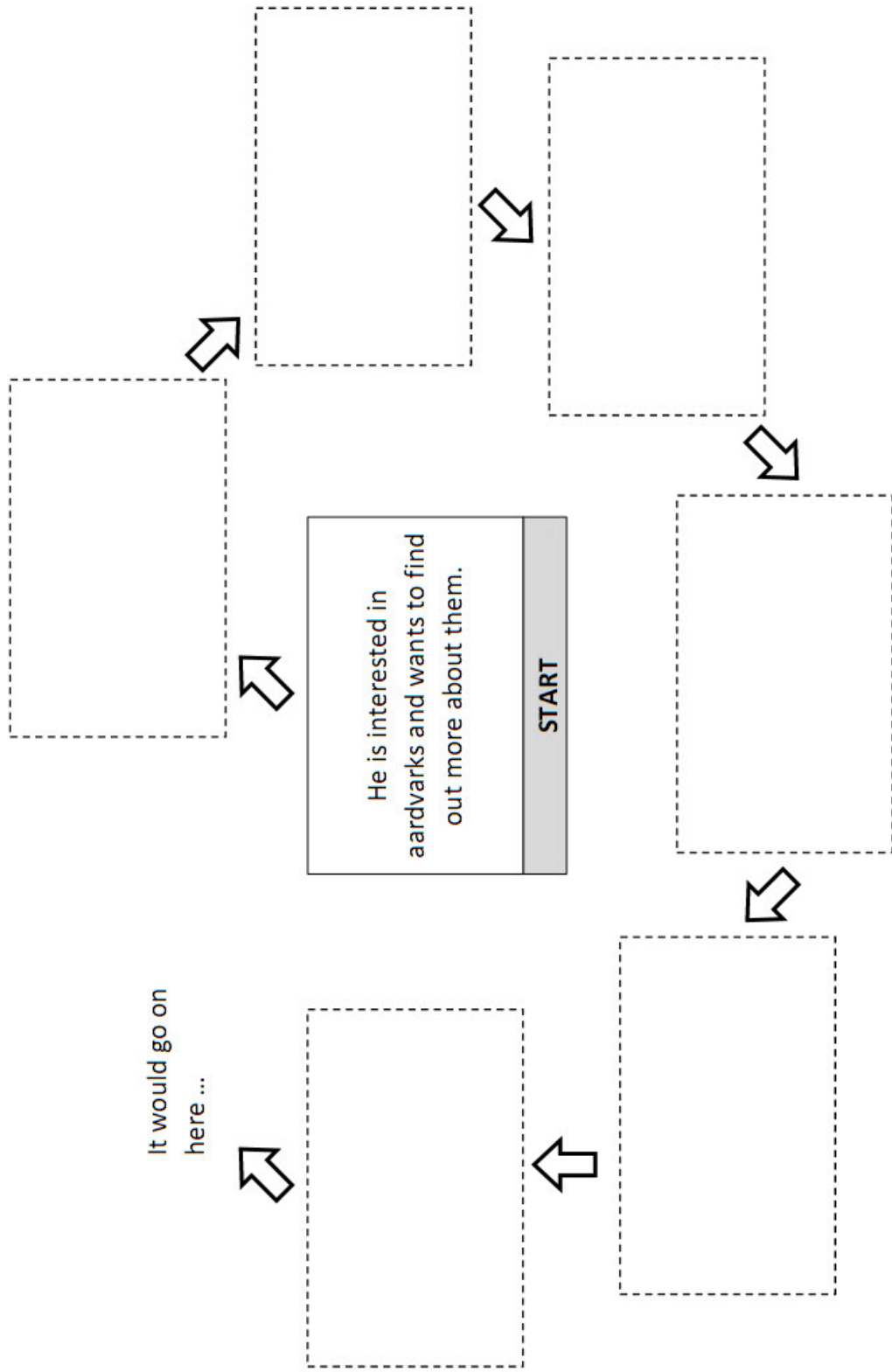


Figure 3. Item example (sorting task part 2).


<p>Mr. Abendstern is a famous scientist and he knows exactly how a scientist has to work.</p>	
<p>Now it's up to you. Please answer the following questions about Mr. Abendstern's working steps.</p> <p>The working steps for Mr. Abendstern are not in the correct order.</p> <p>Mark the <u>best</u> answer with a cross.</p>	
<p>Mr. Abendstern is interested in what causes tooth decay and wants to find out more about it. What is his next step?</p>	
<p><input type="checkbox"/> He plans an experiment to find out if chocolate causes tooth decay.</p>	
<p><input type="checkbox"/> He evaluates results. He observes whether people who only ate only apples for one week had worse teeth than those who ate only chocolate.</p>	
<p><input type="checkbox"/> He establishes a hypothesis: he believes that chocolate causes tooth decay.</p>	

Figure 4. Item example (single-choice item).

Study 2:
Fostering Epistemic Beliefs, Epistemic
Curiosity, and Investigative Interests in
Elementary School Children: A
Randomized STEM Intervention Study

Schiefer, J., Golle, J., Tibus, M., Herbein, E., Hoehne, V., Trautwein, U., & Oschatz, K. (2016). Fostering Epistemic Beliefs, Epistemic Curiosity, and Investigative Interests in Elementary School Children: A Randomized STEM Intervention Study. *Manuscript submitted for publication.*

This study was funded in part by the Hector Foundation II.

Abstract

The fostering of young students' competencies in the so-called STEM subjects is of increasing interest in many countries around the world. The present study investigated the effectiveness of a newly developed science-focused STEM intervention for elementary school children. The intervention included an inquiry-based approach as well as reflections on epistemological issues and focused on the fostering of children's epistemic beliefs, epistemic curiosity, and investigative interests. The effectiveness of the program was investigated by applying a randomized control group design with repeated measures. Sixty-five third- and fourth-grade students participated in this study, which was conducted as part of an extracurricular enrichment program. The results revealed that the children assigned to the intervention developed more sophisticated epistemic beliefs and a higher level of epistemic curiosity than the children assigned to the control condition. This finding indicates that the fostering of epistemic beliefs should be considered already in young children's comprehensive science learning.

Keywords: Science-focused STEM intervention, science epistemology, epistemic beliefs, epistemic curiosity, elementary school children, randomized controlled study

Fostering Epistemic Beliefs, Epistemic Curiosity, and Investigative Interests in Elementary School Children: A Randomized STEM Intervention Study

The promotion of students' achievements and competencies in the so-called STEM disciplines (science, technology, engineering, mathematics) is one cornerstone of current educational research and practice (European Commission, 2007; National Research Council, 2011; NSB, 2010; OECD, 2011). It has even been argued that a country's long-term economic growth depends on its success in fostering students' achievements in the STEM subjects (Sawyer, 2008). Not only because of the social debate about the imminent shortage of engineers and natural scientists, but also due to declines in students' interest in science subjects over the course of their school careers (Krapp, 1998; Pratt, 2007), questions have been raised about how to successfully promote students' interests and competencies in STEM disciplines across all grades (BMBF, 2013; Carnevale, Smith, & Melton, 2015; European Commission, 2007; Sawyer, 2008).

A variety of intervention programs have been developed to foster students' STEM-related interests, achievements, and later career decisions (for recent reviews, see Tsui, 2007; Valla & Williams, 2012; Veenstra, Padró, & Furst-Bowe, 2012). In the field of science learning, inquiry-based approaches in particular are recommended for the development of a profound understanding of science (see Elder, 2002; European Commission, 2007; Lederman, 1992, 2007; National Research Council, 2011). Although it is essential to evaluate the effectiveness of such programs, Valla and Williams (2012) detected a lack of useful, empirically valid evaluation studies in their review of existing research. They reported that existing investigations of STEM interventions often show methodological shortcomings such as a lack of proper control groups, a lack of randomization, or the failure to administer a baseline (Brody, 2006; Valla & Williams, 2012). Especially at the elementary school level, only a few empirically supported interventions have focused on the promotion of fundamental aspects of science (European Commission, 2007; Valla & Williams, 2012).

The goal of the current study was to investigate the effectiveness of a new extracurricular science-focused STEM intervention for elementary school children in Grades 3 and 4. We focused on the targeted fostering of children's epistemic beliefs (conceptions about scientific knowledge and its development), epistemic curiosity, and investigative interests. The intervention concept included an inquiry-based approach to the natural sciences as an enterprise as well as reflections on epistemological issues. We

examined the effectiveness of the intervention by employing a randomized control group design with repeated measures (Torgerson & Torgerson, 2001, 2008; Valla & Williams, 2012). We used a sample of 65 third- and fourth-grade students who were nominated to attend an extracurricular enrichment program.

STEM Interventions and their Effectiveness

Following the call for an effective promotion of students' competencies in the STEM disciplines (European Commission, 2007; National Research Council, 2011; NSB, 2010; OECD, 2011), a variety of approaches have been explored (Tsui, 2007; Valla & Williams, 2012). STEM competencies are defined as "the set of core cognitive knowledge, skills, and abilities that are associated with STEM occupations and the non-cognitive work interests and work values associated with STEM occupations" (Carnevale et al., 2011, p. 97). Because students' interest in and enthusiasm for science have been shown to continuously decrease during their school years (Krapp, 1998; Pratt, 2007), it may be important to promote children's STEM competencies as early as elementary school, when children are at an age that has been described as the "curiosity golden age" (European Commission, 2007, p. 12). Interventions are believed to help counteract this decrease by maximizing the cumulative learning process that is critical for talent development (Keeley, 2009). The intention is to encourage a high level of engagement in science learning as early as possible to increase the chances that long-lasting intervention effects will affect children across all grades (e.g., Brandwein, 1995; Maltese & Tai, 2010; Metz, 2008). Intervention programs have been implemented in many different formats such as after school courses, summer camps, and residential plans (for an overview, see Valla & Williams, 2012). They focus on students' interests, abilities, skills, knowledge, or on values in the STEM disciplines (Carnevale et al., 2011). Their educational goals vary from providing positive experiences in science and math, to exposing students to STEM role models, to instilling and maintaining interest or self-confidence, up to preparing students for professional STEM careers.

Science-focused STEM interventions and promotion programs across different age groups have been shown to affect students' achievement, later occupational attainment, and interest (Benbow, Lubinski, & Sanjani, 1995; Dorsen, Carlson, & Goodyear, 2006; Valla & Williams, 2012; Veenstra, Padró, & Furst-Bowe, 2012; Wai, Lubinski, Benbow, & Steiger, 2010). Positive effects have been reported on elementary-

school-age children's science process skills, science concepts, and content knowledge (Cotabish, Dailey, Robinson, & Hughes, 2013; Valla & Williams, 2012). However, there is a lack of intervention studies that have focused on the promotion of more fundamental aspects of science such as the understanding of the epistemology of science (Elder, 2002; European Commission, 2007; Lederman, 1992, 2007). Students' epistemic beliefs are fundamental for science learning and for a deep understanding of the genesis, constitution, and change in scientific knowledge (Elder, 2002; Jehng, Johnson, & Anderson, 1993; Lederman, 2007). For this reason, in the present study, we focused on promoting elementary school children's epistemic beliefs.

Theoretical Conceptualization of Epistemic Beliefs

Epistemic beliefs are defined as subjective beliefs about the nature of knowledge and knowing (Hofer & Pintrich, 1997). They refer to the epistemology of science or epistemic beliefs inherent in scientific knowledge and its development (Lederman, 1992, 2007). Epistemic beliefs have been described domain-specific, domain-general, or both (Hammer & Elby, 2002; Muis, Bendixen, & Haerle, 2006; Pintrich, 2002). We follow the domain-specific perspective and refer (unless otherwise stated) to investigations of epistemic beliefs in the domain of science.

To date, there is no standard conceptualization of epistemic beliefs. In this study, we followed Conley et al.'s (2004) conceptualization, which is based on previous work by Elder (2002) and Hofer and Pintrich (1997). The corresponding dimensions refer to the nature of knowledge and the nature of knowing: (a) source, (b) certainty, (c) development, and (c) justification of scientific knowledge. The source and justification dimensions reflect beliefs about the nature of knowing, whereas the certainty and development dimensions reflect beliefs about the nature of knowledge.

The source dimension addresses beliefs about knowledge that resides in external authorities. In less sophisticated stances, knowledge is conceptualized as "external to the self, originating and residing in outside authorities" (Conley et al., 2004, p. 190). More sophisticated stances view knowledge as a product of experimental evidence, thinking, or interacting with others. The certainty dimension reflects a less sophisticated stance that involves "the belief in a right answer" (Conley et al., 2004, p. 194). By contrast, more sophisticated views can be identified by statements such as "there may be more than one answer to complex problems" (Conley et al., 2004, p. 190). The development dimension

is associated with beliefs that recognize science as an evolving subject. Less sophisticated stances regard scientific ideas and theories as unchangeable. Stances that are more sophisticated include statements about how scientific ideas are continuously changing (e.g., due to new discoveries or data). Finally, the justification dimension refers to the role of experiments and how students evaluate claims. Less sophisticated stances include assumptions about absolute or nonreflected judgments. Stances that are more sophisticated include justified judgments and the acceptance of a variety of explanations for scientific phenomena (Conley et al., 2004).

For many years, elementary school children's epistemic beliefs were not in the focus of educational research (Elder, 2002; Smith et al., 2000). Because evidence from the cognitive development literature gave rise to the idea that young children have already developed an understanding of the epistemology of science (e.g., Montgomery, 1992; Wellman, 1990), some research has focused on investigating the epistemic beliefs of elementary school children. Using a cross-sectional design, Elder (2002) analyzed the epistemic beliefs of fifth graders who studied hands-on, inquiry-based science. Their epistemic beliefs reflected a mixture of naive and sophisticated understandings. On the one hand, the children tended to regard scientific knowledge as a developing, changing construct that is created by reasoning and testing. On the other hand, they displayed naive notions of science as a mere activity rather than as directed by aims to explain phenomena in the world.

The Promotion of Epistemic Beliefs

Epistemic beliefs have a crucial influence on science learning (e.g., Nussbaum, Sinatra, & Poliquin, 2008; Qian & Alvermann, 1995). There is evidence that it is possible to change epistemic beliefs through new impressions and learning experiences (Perry, 1970). More specifically, research has detected effective teaching strategies and teaching concepts that are essential for enhancing the understanding of science as well as epistemic beliefs. Positive effects have been reported through Inquiry-Based Science Education (e.g., Blanchard, Southerland, Osborne, Sampson, Annetta, & Granger, 2010; Elder, 2002; Minner, Levy, & Century, 2010). In addition, the use of material involving conflicting information (e.g., conflicting evidence or refutations of epistemological instruction) and learning materials that were based on an explicit reflexive approach—where students' attention was actively directed toward relevant aspects of the

epistemology of science via discussions, instruction, or critical scrutiny—has revealed positive effects on the enhancement of epistemic beliefs (Akerson & Hanuscin, 2007; Kienhues, Bromme, & Stahl, 2008).

To date, very few studies have investigated the systematic promotion of epistemic beliefs in the domain of science at the elementary school level. We were able to locate two studies, one focusing on sixth graders and one on fifth graders. Smith et al. (2000) tested whether sixth graders could develop more sophisticated epistemic beliefs in a constructivist compared with a traditional science classroom. In the constructivist classroom, students were allowed to direct their own inquiries, were promoted in their metacognitive understandings, and were engaged in deep domain-specific issues in science. The traditional science classroom presented students with factual learning or concrete problem solving. Only students in the constructivist classroom developed an appropriate epistemological stance toward science that focused on the central role of ideas in the knowledge acquisition process and the mental, social, and experimental work that is involved in this process. Due to the lack of randomization, additional factors besides the curricula may have contributed to the group differences.

Conley et al. (2004) investigated whether the epistemic beliefs of fifth-grade students could be enhanced during a 9-week science course. Results showed that students became more sophisticated in their beliefs about the source and certainty of knowledge over time according to a comparison between posttest and pretest measurements. However, no reliable changes were found in the development and justification dimensions. As no control group was used in the study, alternative explanations for the effects (e.g., maturation) could not be excluded. It has to be pointed out—in accordance with Valla and Williams (2012) or Brody (2006)—that existing intervention studies on the promotion of elementary school children’s epistemic beliefs have suffered from methodological limitations such as the absence of control groups or a lack of randomized group allocation (see Conley et al., 2004; Smith et al., 2000).

Epistemic Curiosity and Investigative Interests

Beyond epistemic beliefs, the promotion of epistemic curiosity and investigative interests is an important aim of science-focused STEM interventions, especially at the elementary-school level (Carnevale et al., 2011; Pratt, 2007). Epistemic curiosity is defined as the desire for knowledge that motivates individuals to learn new ideas, to

eliminate information gaps, and to solve intellectual problems (Litman & Spielberger, 2003; Litman, 2008). It has been positively related to epistemic beliefs (Richter & Schmid, 2010), exploratory behavior, and the closure of gaps in knowledge (Litman, Hutchins, & Russon, 2005). The new science-related STEM intervention that we are introducing in this study includes intellectually challenging elements that are intended to inspire children to enjoy thinking and to reflect on scientific problems and should help eliminate information gaps. Due to these elements, we expected that in addition to a positive effect on epistemic beliefs, children's epistemic curiosity would be positively affected by attending the intervention.

As our intervention also included several aspects of science learning as active investigations or inquiry-based approaches, we expected that children's investigative interests would also be fostered through their participation. The promotion of investigative interests has been a common aim of science-focused STEM interventions (Carnevale et al., 2011; European Commission, 2007; National Research Council, 2011). Students with a high level of investigative interests prefer activities that involve thought, observation, investigation, exploration, and discovery. They like to solve problems, perform experiments, and conduct research (Holland, 1997). A high level of investigative interest has been found to be positively related to the choice to pursue a STEM occupation. It can predict transitions from school to studying a STEM subject at university, career decisions (e.g., working in a STEM occupation for many years), as well as a sustained commitment to scientific pursuits (Carnevale et al., 2011; Lubinski & Benbow, 2006; Rounds & Su, 2014). An early promotion of investigative interests has high relevance and might support students' later tenure in a STEM occupation (Nye, Su, Rounds, & Drasgow, 2012).

The Present Study

The goal of the present study was to investigate the effectiveness of a newly developed science-focused STEM intervention for elementary school children with regard to promoting children's epistemic beliefs, epistemic curiosity, and investigative interests. We used a randomized control group design with repeated measures to estimate the average causal effect of our program (Torgerson & Torgerson, 2001, 2008).

The European Commission diagnosed a lack of intervention studies that go beyond the teaching of science-content knowledge (Andrés et al., 2010; European

Commission, 2007). Therefore, we focused on promoting epistemic beliefs, which are essential for the development of a basic understanding of science (Elder, 2002; Lederman, 1992, 2007). Our intervention included learning settings that allowed students to participate actively (e.g., inquiry-based science education) as well as to reflect on epistemological issues. Thus, we expected to find positive effects of the intervention on epistemic beliefs. In addition, the intervention included several aspects of science learning and experimentation and required the elimination of information gaps. These aspects are associated with investigative interests and epistemic curiosity. Therefore, we expected to find positive intervention effects on those constructs as well.

Moreover, we investigated whether epistemic beliefs could already be fostered in elementary school children in Grades 3 and 4. Previous research has mostly focused on elementary school children from the fifth grade onwards or secondary school students (e.g., Conley et al., 2004; Smith et al., 2000). Children are commonly exposed to formal instruction in science during elementary school for the first time and acquire an understanding of the world around them (Bruer, 1993). Thus, they might develop some “comprehension of the nature of scientific knowledge” (Elder, 2002, p. 347). Therefore, we surmised that third and fourth graders would be ready to benefit from an intervention focusing on the promotion of conceptions about the nature of knowledge and knowing (see Hofer & Pintrich, 1997).

Method

Sample

Data were collected from 65 elementary school children (58.46% male, age: $M = 8.73$, $SD = 0.60$, Grade 3: $N = 33$, Grade 4: $N = 32$). All of them participated in a voluntary extracurricular enrichment program (Hector Children’s Academy Program, HCAP) in the German state of Baden-Württemberg. To take part in the program, the children had to be nominated by their class teacher. The intention is that the 10% most talented or best-performing children in an age cohort will be given the opportunity to participate in this statewide program. Besides school performances, a high level of motivation and interest can be taken into account for the nominations. After admission, children can choose from a variety of courses.

Table 1

Descriptive Statistics for the Scales: Means, Standard Deviations, Internal Consistencies, Numbers of Items, and Examples

Construct	T1			T2			Number of items and example				
	N	M	SD	α	N	M		SD	α		
Source of knowledge ^a	IG	32	2.19	0.53	.73	30	2.62	0.58	.74	5	Everybody has to believe what scientists say.
	CG	31	2.20	0.62		26	2.50	0.67			
Certainty of knowledge ^a	IG	32	2.60*	0.66	.73	30	2.92	0.67	.75	6	All questions in science have one right answer.
	CG	31	2.23*	0.57		26	2.44	0.53			
Development of knowledge	IG	32	3.27*	0.40	.58	30	3.50	0.40	.73	6	Ideas in science sometimes change.
	CG	31	3.49*	0.38		26	3.31	0.55			
Justification of knowledge (Conley et al., 2004)	IG	32	3.51	0.33	.58	30	3.55	0.32	.64	9	It is good to have an idea before you start an experiment.
	CG	31	3.57	0.36		26	3.41	0.39			
Epistemic curiosity (Litman, 2003)	IG	32	3.10	0.48	.86	31	3.17	0.46	.86	14	I always like to learn something new.
	CG	32	3.19	0.42		25	3.07	0.38			
Investigative interests (Holland, 1997)	IG	31	4.00	0.81	.78	28	4.15	0.48	.57	7	How much are you interested in mixing different liquids?
	CG	33	4.15	0.63		22	4.14	0.46			
Fluid intelligence ^b (Weiß, 2006)	IG	30	119.07	15.06	.77	-	-	-	-	56	
	CG	28	118.32	15.40		-	-	-	-		

Note. N = Number of participating children, M = mean, SD = standard deviation, α = Cronbach's alpha. Measurement time points: T1 = November 2013, T2 = February 2014. IG = Intervention group, CG = control group. *t* tests for independent samples (IBM SPSS, version 22) were computed to test for significant differences between the IG and the CG at T1.

^aItems were recoded so that higher scores reflected more sophisticated beliefs. ^bFluid intelligence was measured once.

* $p < .05$.

Four academies in different regions participated in the study. Overall, the participating children were from 24 different schools and 26 classes. They had a mean IQ of 119 ($SD = 15.10$). The intervention group consisted of 32 children (62.5% male, 56.25% Grade 3, age: $M = 8.74$, $SD = 0.58$), the control group consisted of 33 children (54.54% male, 45.45% Grade 3, age: $M = 8.75$, $SD = 0.58$). The number of applications for the course determined the sample size. The baseline demographics for each group can be found in Table 1.

Experimental Design

We investigated the effectiveness of the intervention by using a randomized control group design with repeated measures (pretest [T1], posttest [T2]; see Figure 1).

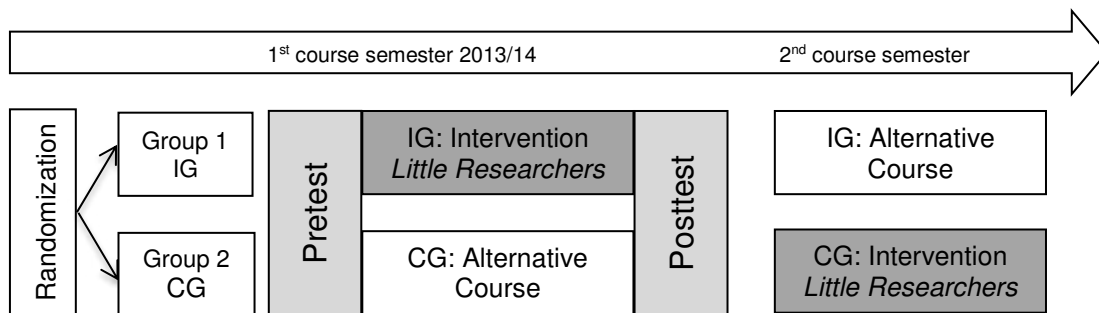


Figure 1. Design of the intervention study. IG = Intervention group, CG = control group.

The intervention consisted of the course titled *Little Researchers—We Work Like Scientists* (see below for the description of the intervention). The children met for 90 min once a week for 10 weeks. The control condition was an alternative course—a public speaking training covering speech anxiety, nonverbal communication, and comprehensibility—where children were trained to give a competent presentation about a scientific topic of their choice. This course always took place at the same time as the intervention, and like the intervention, it was developed and held by a researcher from the university. In order to enable randomization, the intervention and the control course were offered as two parts of the same course titled Talking about science—With others and to others. Furthermore, by combining the courses, we could assume that the intervention group did not differ from the control group in their interest in science and scientific topics. This study design offers the advantage that all participants received the treatment, a practice that corresponds to the ethical principles of intervention studies (Emanuel,

Wendler, & Grady, 2000). It also minimized the risk of attrition between the pretest and posttest and enabled us to control the activities of the children in the control group. Furthermore, the intervention effects could be reduced to the specific methods of the science intervention (e.g., inquiry-based learning and reflecting on epistemological issues) because the control group also worked with scientific topics while preparing their presentations.

The courses took place over two semesters (first semester from November, 2013, to February, 2014; second semester from March to July, 2014). After registering for the course, and within each of the four participating academies, the children were randomly assigned to either the intervention or control group. By using a random number generator in Excel, the randomization was conducted by a neutral person before the intervention began and did not involve any restrictions (e.g., grade level or gender). The children who participated in the intervention course during the first semester participated in the control course in the second semester and vice versa. Data collection was integrated into the first two and the last two parallel course sessions of the first semester. The courses consisted of five to 10 children. Trained research assistants and scientists belonging to the university administered questionnaires. Prior to testing, we obtained parents' written consent for their child's participation.

Description of the Intervention

The intervention Little Researchers—We Work Like Scientists was developed and implemented by scientists in the field of education sciences, psychology, and science education. The course instructors were scientists from the university who worked with a detailed manual and time schedule for the single course units to ensure the fidelity of the implementation (O'Donnell, 2008). Overall, there were only slight differences in the implementations of the program (e.g., deviations from the timetables between 5 to 10 min). All deviations were documented and discussed in regular meetings.

Single course units (see Figure 2) were arranged in such a way that children experienced and applied the cumulative and cyclical process of scientific research in each course unit (Conley et al., 2004; Kuhn, 2002; White, Frederiksen, & Collins, 2009). On the basis of theories and derived hypotheses about the phenomena of interest, the students conducted experiments, analyzed data, evaluated evidence, presented results, and drew

inferences with the goal of generating or revising the theories. This process was intended to give them insight into scientific working methods. Thereby, they experienced science as a hypothesis-driven and changing subject, which is relevant for the development of sophisticated epistemic beliefs. This was strengthened through the explicit integration of conflicting information and results (i.e., due to the use of different materials or different methods of investigation). This was intended to support children's critical thinking about the subjectivity as well as the boundaries of empirical research (see Kienhues, Bromme, & Stahl, 2008). Further empirically supported elements for the enhancement of epistemic beliefs from were adapted for third and fourth graders (i.e., Akerson & Hanuscin, 2007; Blanchard et al., 2010; Conley et al., 2004). Those were reflections of relevant aspects of the epistemology of science by means of discussions, science communication, teaching, and critical scrutiny.

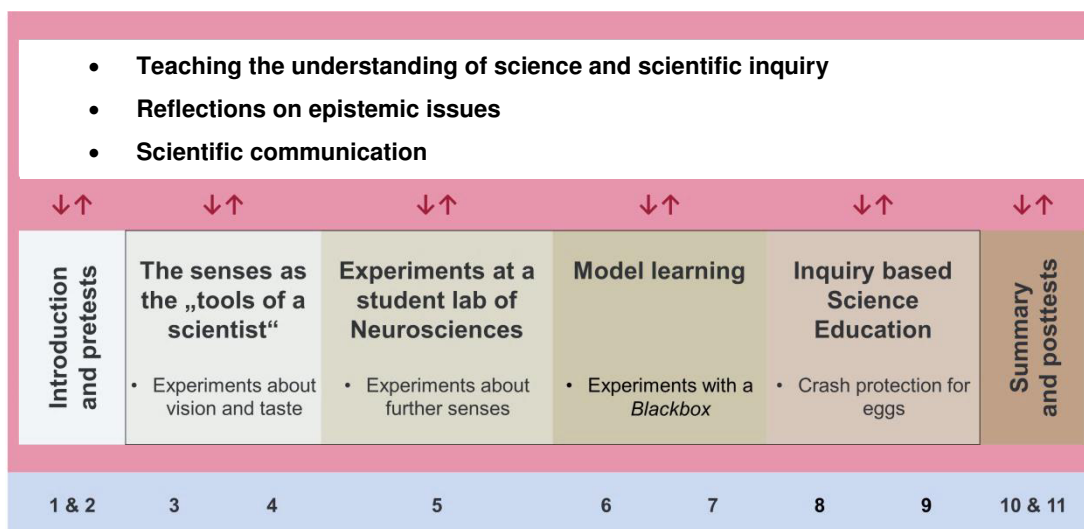


Figure 2. Course concept of the intervention *Little Researchers—We Work Like Scientists*.

As hands-on activities as well as inquiry-based learning (Blanchard et al., 2010; Elder, 2002) revealed positive effects on the enhancement of epistemic beliefs, these elements were integrated into different course units (e.g., experiments on the human senses, experiments in a student lab for neuroscience, examination of an unknown object—a so-called “black box” [Frank, 2005], and simple physical experiments). Children were also encouraged to formulate, present, and discuss their hypotheses and the results of their “research studies” in small groups and simulated “research congresses.” This was supposed to foster critical and reflexive thinking as well as

communication about science. In all course units, the children experienced the principles as well as the boundaries of scientific inquiry. To meet this target, the children were given the opportunity to carry out practical experiments and reflect on their ideas and observations. The publication of the course concept—including all materials—is in planning to provide the concept for teachers or course instructors of the HCAP.

Measures

All administered scales and the corresponding descriptive statistics, Cronbach's alphas, numbers of items, and examples can be found in Table 1. Epistemic beliefs were assessed with a 26-item instrument (Conley et al., 2004, adapted from previous work by Elder, 2002, translated by Urhahne & Hopf, 2004). Four subscales reflect the dimensions source, certainty, development, and justification of knowledge. Items were rated on a 4-point Likert scale ranging from 1 (*I completely disagree*) to 4 (*I completely agree*), presented via stars of increasing size. The scale source of knowledge addresses beliefs about knowledge residing in external authorities. Certainty of knowledge refers to the tendency to believe in a right answer. Development of knowledge measures beliefs about science as an evolving and changing subject. Justification of knowledge addresses the role of experiments and how individuals justify knowledge. The source and certainty scales were recoded for the analyses. Thus, for each scale, higher scores reflected more sophisticated beliefs.

Epistemic curiosity was assessed with an instrument developed by Litman (2003). The items address the desire for knowledge that motivates individuals to learn something new and to solve intellectual problems. Items were rated on a 4-point Likert scale ranging from 1 (never) to 4 (always). Seven items from the RIASEC questionnaire (Holland, 1997) were administered to assess the investigative interests of the participants. The items were rated on a 5-point Likert scale ranging from 1 (*not at all*) to 5 (*very much*). Children filled out the complete RIASEC at home after the first course session because there was not enough time to administer this questionnaire during the first two course sessions. Children's fluid intelligence was measured with a nonverbal intelligence test (Culture Fair Test - CFT 20-R; Weiß, 2006), including the four subscales continuing series, classifications, matrices, and topological conclusions.

Statistical Analyses

Due to the small sample size, possible group differences at T1 were analyzed with t tests for independent samples in IBM SPSS (version 22). The effectiveness of the intervention was analyzed with multiple regression analyses in Mplus (Muthén & Muthén, 1998-2012). All analyses used the robust maximum likelihood estimator, which corrects standard errors for the non-normality of the variables (Muthén & Muthén, 1998-2012). The dependent variables were the z-standardized posttest measures from the previously described scales. The predictors in our regression models were group assignment (0 = control, 1 = intervention), and for each dependent variable, the corresponding z-standardized score on the pretest (see Enders & Tofighi, 2007). Due to the standardization of the dependent variables, the multiple regression coefficient of the group variable indicated the standardized intervention effect (effect size) controlled for the score on the corresponding pretest.

Missing data. No children from the intervention group or from the control group discontinued their participation in the study during the first semester (no dropout). However, due to illness or other reasons, some children missed single course sessions and were not able to participate in all of the surveys. Therefore, missing values occurred across all variables at rates of between 6.25% and 21.88% (see Table 1 for the exact number of participants for each measure). Because one questionnaire was filled out at home, the RIASEC scales had a rate of missing values of 33%. To analyze the intervention effects, we used the full information maximum likelihood approach implemented in Mplus to deal with the missing values (Muthén & Muthén, 1998-2012). All measured variables were taken into account to estimate the model parameters (see Schafer & Graham, 2002).

Results

Descriptive Statistics and Bivariate Correlations

In a first step, we analyzed the characteristics and differences between the intervention and control groups at T1 (see Table 1). Participants had a mean IQ of 119 ($SD = 15.10$), which is above average because the sample consisted of children who were nominated for the HCAP. There were no IQ or gender differences between the intervention and control groups. There were also no differences between the groups in

source of knowledge, justification of knowledge, epistemic curiosity, and investigative interests. However, we found differences between the two groups at T1 on certainty of knowledge, $t(61) = 2.45, p = .017$, in favor of the intervention group, and development of knowledge, $t(61) = -2.28, p = .026$, in favor of the control group. In all regression analyses, we controlled for the pretest score on each scale.

Before the intervention started, children showed a high level of investigative interests ($M = 4.08, SD = 0.72$, scale ranged from 1 to 5). Their epistemic curiosity ($M = 3.15, SD = 0.45$, Range = 1 to 4) was also above the middle of the scale. They also had quite sophisticated understandings of development ($M = 3.38, SD = 0.40$) and justification of knowledge ($M = 3.54, SD = 0.34$). In comparison with those dimensions, they scored lower on source ($M = 2.19, SD = 0.57$) and certainty of knowledge ($M = 2.42, SD = 0.63$, each scale ranged from 1 to 4).

Intercorrelations of all outcome variables are shown in Table 2. At T1 and T2, source of knowledge was positively correlated with certainty of knowledge, and development of knowledge was positively correlated with justification of knowledge (see the coding of the items). At T1, epistemic curiosity was positively correlated with beliefs about development and justification of knowledge, and at T2, with justification of knowledge and investigative interests. Investigative interests were (with the exception of certainty) positively correlated with epistemic beliefs and curiosity at T1. All measures of epistemic beliefs, interests, and curiosity were independent of children's intellectual abilities (except for certainty of knowledge) at T2.

Effects of the Intervention on the Development of Epistemic Beliefs

The first research question concerned the intervention's promotion of epistemic beliefs. Regression models for each epistemic belief subscale were used to assess the general effectiveness of the program. The results of the multiple regression analyses are presented in Table 3. The findings showed that the children assigned to the intervention exhibited more sophisticated epistemic beliefs than the children assigned to the control condition at the end of the first semester. Overall, three out of four scales were positively affected by the intervention. Controlling for the initial level of the respective outcome, the results revealed that children in the intervention compared with the control group scored significantly higher on the posttest measures of certainty of knowledge ($B = .45, p = .025$), development of knowledge ($B = .61, p = .010$), and justification of knowledge

Table 2

Intercorrelations between the Scales

Construct	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Source of knowledge T1	1												
(2) Certainty of knowledge T1	.63**	1											
(3) Development of knowledge T1	-.14	-.06	1										
(4) Justification of knowledge T1	-.22	-.16	.52**	1									
(5) Epistemic curiosity T1	-.05	-.14	.45**	.47**	1								
(6) Investigative interests T1	.30*	.20	.38**	.35**	.37**	1							
(7) Source of knowledge T2	.47**	.36**	-.11	-.14	-.11	.05	1						
(8) Certainty of knowledge T2	.29*	.59**	-.19	-.17	-.14	-.12	.61**	1					
(9) Development of knowledge T2	.25	.15	.25	.38**	.43**	.25	.06	.12	1				
(10) Justification of knowledge T2	-.04	.12	.22	.45**	.24	.19	-.10	.08	.55**	1			
(11) Epistemic curiosity T2	.08	.05	.32*	.29*	.69**	.35**	-.10	-.12	.19	.46**	1		
(12) Investigative interests T2	.14	.02	.36*	.21	.38**	.42**	.11	.08	.23	.05	.33*	1	
(13) Fluid intelligence	.06	.15	-.16	.07	-.08	.12	.18	.34*	.25	.26	-.01	-.06	1

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3
Course Effects on Epistemic Beliefs: Predicting Children's Posttest Measures

Variables	Source of Knowledge (T2) ^a		Certainty of Knowledge (T2) ^a		Development of Knowledge (T2) ^a		Justification of Knowledge (T2) ^a	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Model 1								
Treatment ^b	.18	(.23)	.45*	(.23)	.61*	(.26)	.43*	(.24)
Pretest score ^a	.49***	(.11)	.53***	(.11)	.38***	(.10)	.44***	(.13)
Explained variance (R ²)	.25		.40		.17		.23	
Model 2								
Treatment ^b	.17	(.23)	.36*	(.22)	.64*	(.27)	.43*	(.24)
Pretest score ^a	.65***	(.23)	.57	(.10)	.51	(.19)	.38*	(.19)
Treatment ^b x pretest score ^a	-.34	(.17)	-.25*	(.10)	-.20	(.23)	.12	(.24)
Explained variance (R ²)	.30		.43		.20		.23	

Note. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention.

* $p < .05$. ** $p < .01$. *** $p < .001$.

($B = .43, p = .038$). Thus, participants exhibited more sophisticated views regarding the certainty of knowledge (e.g., an understanding that there might be more than one answer to complex problems) and beliefs that recognize science as an evolving and changing subject. They also reported more sophisticated stances regarding the role of experiments and how individuals justify knowledge (acceptance of a variety of explanations for scientific phenomena). The regression coefficients can be interpreted as effect sizes and are summarized in Figure 3. They represent the z-standardized differences between the intervention and the control group controlled for the initial level of each corresponding scale. No intervention effect was found for source of knowledge ($B = .18, p = .223$). After the intervention, there was no difference between the two groups on the development of children's beliefs about knowledge residing in external authorities. For all variables, the pretest values had significant positive effects on the posttest values.

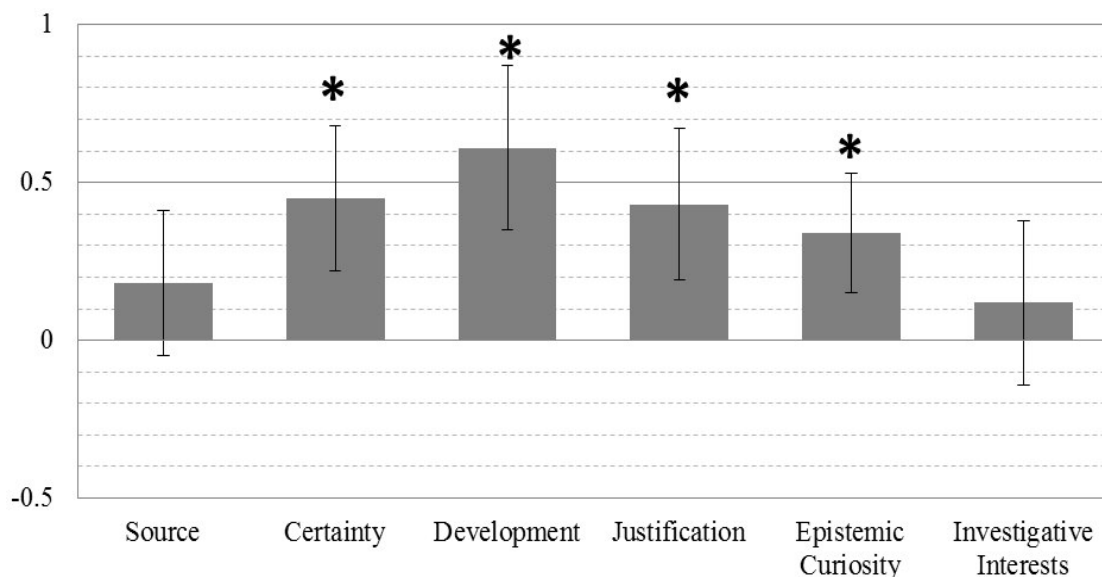


Figure 3. Effect sizes: Bars represent the z-standardized differences between the development of the intervention and the control group (posttest differences controlled for the pretest values). Error bars indicate standard errors. * $p \leq .05$.

To investigate whether there were differential intervention effects that depended on children's initial epistemic belief scores, interactions between the group variable and each pretest score were included in the regression analyses (see Table 3, Model 2). The analyses revealed only one significant interaction between course participation and the certainty of knowledge pretest scores ($B = -.25, p = .017$), that is, the lower the pretest scores of the children in the intervention group, the higher the benefit for the children in

the intervention group compared with the children in the control group. The results remained stable when we additionally controlled for participants' intelligence and gender (see Table 4). Overall, these findings provide evidence that it is possible to promote epistemic beliefs in elementary school children in Grades 3 and 4.

Effects of the Intervention on Epistemic Curiosity and Investigative Interests

Our second research question concerned whether children's epistemic curiosity and investigative interests could be fostered through their participation in the intervention. The results of the multiple regression analyses are presented in Table 5. The results revealed that children assigned to the intervention compared with the control condition scored significantly higher on epistemic curiosity ($B = .34, p = .041$). They reported a greater desire for knowledge and more motivation to learn something new and to solve intellectual problems.

No intervention effect was found for investigative interests ($B = .12, p = .318$). After the intervention, there was no difference between the two groups on the development of children's interests in solving problems, performing experiments, conducting research, or activities involving thought, observation, investigation, exploration, or discovery. The analyses also did not reveal any significant interactions between course participation and the epistemic curiosity and investigative interest pretest scores (see Table 5, Model 2). The results remained stable when we additionally controlled for participants' intelligence and gender (see Table 6).

Table 4

Course Effects on Epistemic Beliefs: Predicting Children's Posttest Measures (Controlling for the Pretest Measures, Gender, and IQ)

Variables	Source of Knowledge (T2) ^a		Certainty of Knowledge (T2) ^a		Development of Knowledge (T2) ^a		Justification of Knowledge (T2) ^a	
	B	SE	B	SE	B	SE	B	SE
Treatment ^b	.11	(.22)	.35 [†]	(.22)	.63 [*]	(.28)	.43 [*]	(.23)
Pretest score ^a	.71 ^{***}	(.14)	.58 ^{***}	(.10)	.49 [*]	(.20)	.37 [*]	(.18)
Treatment ^b x pretest score ^a	-.38 [*]	(.19)	-.19	(.10)	-.19	(.23)	.13	(.23)
Gender ^c	.14	(.22)	-.33	(.20)	.09	(.25)	-.07	(.23)
IQ ^a	.13	(.08)	.22	(.11)	-.04	(.11)	.03	(.20)
Explained variance (R ²)	.33		.51		.19		.23	

Note. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention. ^cGender was dummy-coded 0 = girls, 1 = boys.

[†] $p < .10$. ^{*} $p < .05$. ^{**} $p < .01$. ^{***} $p < .001$.

Table 5

Course Effects on Epistemic Curiosity and Investigative Interests: Predicting Children's Posttest Measures

Variables	Epistemic Curiosity (T2) ^a		Investigative Interests (T2) ^a	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Model 1				
Treatment ^b	.34*	(.19)	.12	(.26)
Pretest score ^a	.69***	(.09)	.42***	(.16)
Explained variance (R ²)	.49		.18	
Model 2				
Treatment ^b	.35*	(.19)	.15	(.27)
Pretest score ^a	.56***	(.14)	.57	(.31)
Treatment ^b x pretest score ^a	.20	(.18)	-.21	(.36)
Explained variance (R ²)	.49		.20	

Note. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6

Course Effects on Epistemic Curiosity and Investigative Interests: Predicting Children's Posttest Measures (Controlling for the Pretest measures, Gender, and IQ)

Variables	Epistemic Curiosity (T2) ^a		Investigative Interests (T2) ^a	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
Treatment ^b	.42*	(.18)	.20	(.28)
Pretest score ^a	.59***	(.14)	.61*	(.32)
Treatment ^b x pretest score ^a	.15	(.19)	-.26	(.38)
Gender ^c	-.19	(.19)	-.30	(.28)
IQ ^a	-.16	(.11)	-.08	(.13)
Explained variance (R ²)	.52		.23	

Note. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention. ^cGender was dummy-coded 0 = girls, 1 = boys.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This study tested whether elementary school children's epistemic beliefs, epistemic curiosity, and investigative interests could be promoted by a new science-focused STEM intervention. The intervention was developed specifically for third- and fourth-grade students. With inquiry-based learning elements and reflections on epistemological issues, it included some of the most promising and empirically supported elements for the enhancement of epistemic beliefs (Akerson & Hanuscin, 2007; Blanchard et al., 2010; Conley et al., 2004; Elder, 2002; Kienhues et al., 2008). The analyses revealed positive effects of the course on the enhancement of certainty of knowledge, development of knowledge, justification of knowledge, and epistemic curiosity. The results supported the effectiveness of the intervention.

Effects of the Intervention on Epistemic Beliefs

The current study revealed that our STEM intervention positively affected elementary school children's epistemic beliefs. Overall, epistemic beliefs were enhanced in the dimensions certainty of knowledge, development of knowledge, and justification of knowledge for children who participated in the intervention compared with children in the control condition (a course on speech training held at the same time). They became more sophisticated in their stance on certainty, which is the assumption that there may be more than one possible response to complex problems (see Conley et al., 2004). Referring to the development dimension, their stances moved toward recognizing science as an evolving subject, meaning that scientific ideas and theories can change on the basis of new data and evidence. Referring to the justification dimension, participants' stances moved toward using a variety of options and justifications for their judgments when using evidence or evaluating claims. In previous intervention studies (e.g., Conley et al., 2004), no changes had been found in the development or justification of knowledge dimensions.

Our study contributes to answering the question of whether epistemic beliefs can be successfully promoted in children as young as 8 to 10 years old. Researchers espousing the Piagetian hypothesis (Inhelder & Piaget, 1958) that elementary school children are "concrete thinkers" (Smith et al., 2000, p. 400) would not expect intervention effects on complex dimensions such as development or justification of knowledge. They would argue that the stimulation of these demanding dimensions of epistemic beliefs requires abstract thinking abilities and metacognitive activation that are not yet present before

secondary school. Against this background, rather small or no promotion effects would have to be expected. However, our results are in line with the propositions of researchers who believe that developmental leaps can be induced in specific domains by promotion, environmental circumstances, or supportive scaffolding (see Berk & Winsler, 1995; Fischer, 1980; Vygotsky, 1980). Our intervention can be considered a highly supportive learning context (see skill theory by Fischer, 1980) that enabled children to reach an optimal level of abstraction and reflection. In the course, the children were encouraged to argue about ideas and hypotheses using empirical evidence or to critically reflect on the results of their own investigations. It can be concluded that such elements successfully emphasized argumentation, reflection, and the development of a profound understanding of science (see Elder, 2002; Lederman, 1992, 2007). The items from the epistemic beliefs questionnaire (by Conley et al., 2004) were not explicitly addressed or discussed in the course. Thus, the children were not simply taught to do well on the questionnaire (“teaching to the test”; see Longo, 2010).

Effects of the Intervention on Epistemic Curiosity and Investigative Interests

As expected, we found positive effects on the development of children’s epistemic curiosity. According to this finding, it can be concluded that participating in the intervention caused a greater desire to obtain new knowledge or to solve intellectual problems (see the definition of epistemic curiosity by Litman & Spielberger, 2003). In the course, the children were given many opportunities to generate hypotheses, perform experiments, analyze results, and draw conclusions. Such elements might have activated their enjoyment of thinking and might have given them insights into new issues. The positive effect on epistemic curiosity is all the more impressive because epistemic curiosity has sometimes been conceptualized as a personality trait (Litman, 2008) and is therefore believed to be a rather stable person characteristic. However, there is increasing evidence that personality traits develop in response to environmental factors and intervention studies (e.g., Caspi & Roberts, 2001; Magidson, Roberts, Collado-Rodriguez, & Lejuez, 2014). Our results provide some further evidence for intervention effects on presumably stable person characteristics. However, it is important to note that, given the pre-posttest design of our study, we do not know how long-lasting the effects are.

No intervention effect was found on children's investigative interests (RIASEC; Holland, 1997). Children who attended the intervention course compared with children who attended the control course showed no statistically significant difference in their vocational interests in solving problems, performing experiments, or conducting research. The items on this questionnaire may have been too abstract for elementary school children and may have included topics (e.g., interest in reading the newspaper or mixing liquids) that the children could not associate with the contents of the STEM intervention.

Implications

Our results have important implications for educational research and practice. First, our STEM intervention program was able to positively affect elementary school children's epistemic beliefs and epistemic curiosity. It can be concluded that the conception of the intervention and the combination of the single course elements were successful and positively affected children's thinking about epistemological issues and their thirst for knowledge (Elder, 2002; Lederman, 1992, 2007).

Second, to the best of our knowledge, this is the first study to show that it is possible to successfully foster epistemic beliefs in the domain of science in children below Grade 5. Our finding corresponds with suggestions by Smith et al. (2000) who concluded, "elementary school children are more ready to formulate sophisticated epistemological views than many have thought" (p. 350). It has educational implications for science learning. Our study revealed that conceptions about scientific knowledge and its development should be taken into account in research on comprehensive science learning in elementary-school-age children.

Limitations and Future Research

Although our study demonstrated beneficial effects of a new extracurricular science-focused STEM intervention for elementary school children, some limitations should be considered when interpreting its results. First, the generalizability of our study is limited. We had a sample of 65 third and fourth graders who were nominated by their class teacher to participate in an enrichment program for high-ability learners. They had an average IQ of 119 ($SD = 15.10$). Besides their cognitive abilities, the children's interest and motivation were considered in these nominations. It can therefore be assumed that the children had a high level of interest in science as well as a high level of motivation,

especially given that they voluntarily participated in the program after school. Therefore, it cannot be concluded that the intervention would have similar positive effects in other learning groups (e.g., school classes). Future research should focus on replicating this study with different samples or on implementing the course elements into the school context. However, implementing this STEM intervention in other samples or school classes might require some changes in terms of the specific course contents.

With respect to the generalizability of the program, the fact that the course was taught by researchers from the university should be considered. This represents a limitation in terms of transferring the results of the intervention to other course instructors. In return, the choice of instructors ensured a high level of fidelity in the implementation that was particularly important for a first evaluation of the program (see Carroll et al., 2007). Scale-up studies should investigate whether the intervention when implemented by teachers or external course instructors (with a background in natural sciences) will have the same effect as the intervention implemented by researchers.

Second, no conclusions can be drawn about the effectiveness of the single course elements as the data were collected only at the beginning and the end of the intervention. It might be promising to include intermediate surveys in future research to identify course elements that work better than others as well as to determine the minimum number of course units required to enhance children's epistemic beliefs. As our intervention was intended to combine several promising elements for promoting epistemic beliefs, it was initially most important for the overall course program to reveal positive effects. Further research can examine the processes through which the intervention works. Qualitative analyses (e.g., think-aloud methods during specific course elements) might be useful for clarifying how children can be encouraged to think about epistemological issues.

Third, it was difficult to find appropriate instruments for this age group. Few questionnaires were available for elementary school children, and there has been little research on the characteristics and measurement of the epistemic beliefs of children below Grade 5. We used an instrument that was developed for older age groups (the questionnaire on epistemic beliefs by Conley et al., 2004). Some subscales had moderate reliabilities (especially the pretest measures of development and justification of knowledge), which may have decreased the potential of the study to find substantial intervention effects with these scales. Nevertheless, we found positive course effects and were able to successfully implement the instrument in elementary school children. Future

research could focus on adapting the existing instruments for elementary school children or on developing comprehensive new instruments that measure young children's epistemic beliefs.

Finally, no conclusions can be drawn about the long-term effects of our intervention. Therefore, future studies should follow students' development for a longer period of time and should also take other outcomes into account. Longitudinal studies could investigate whether secondary school children show benefits from participating in a science-focused STEM intervention in elementary school and how their epistemic beliefs affect their later science achievements or vocational choices.

Conclusion

The present study investigated the effectiveness of a newly developed science-focused STEM intervention for elementary school children in Grades 3 and 4. It included an inquiry-based approach as well as reflections on epistemological issues. We used a randomized controlled trial, thus fulfilling the gold standard for investigating the effectiveness of intervention programs. The results of this high-quality intervention study point to the effectiveness of the intervention and suggest that it was possible to promote young children's epistemic beliefs and epistemic curiosity. This indicates that the fostering of epistemic beliefs can be taken into account in research on comprehensive science learning at an early age. Future research should focus on large-scale implementations of the intervention as well as the investigation of the long-term effects of promoting young children's epistemic beliefs.

References

- Akerson, V. L., & Hanuscin, D. L. (2007). Teaching nature of science through inquiry: Results of a 3 year professional development program. *Journal of Research in Science Teaching*, 44(5), 653-680.
- Andrés, S. H., Steffen, K., & Ben, J. (2010). *TALIS 2008 technical report*. Paris: OECD, Teaching and Learning International Survey.
- Benbow, C. P., Lubinski, D., & Sanjani, H. E. (1995). Our future leaders in science: Who are they? Can we identify them early? *Talent Development III*, 59-70. Scottsdale, AZ: Great Potential Press.
- Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education. NAEYC Research into Practice Series. Volume 7*. National Association for the Education of Young Children, 1509 16th Street, NW, Washington, DC 20036-1426 (NAEYC catalog# 146).
- Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, 94(4), 577-616.
- Brandwein, P. F. (1995). *Science talent in the young expressed within ecologies of achievement* (Report No. EC 305208). Storrs, CT: National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED402700).
- Brody, L. (2006, September 26). Measuring the effectiveness of STEM talent initiatives for middle and high school students. Paper prepared for *Identifying and Developing STEM Talent: A Planning Meeting*. Washington, DC: National Academy of Sciences.
- Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. Cambridge: MIT press.
- Bundesministerium für Bildung und Forschung (Hrsg.) (2013). *Perspektive MINT. Wegweiser für MINT-Förderung und Karrieren in Mathematik, Informatik, Naturwissenschaften und Technik [Perspective MINT. Guide for promotion and careers in mathematics, informatics, natural sciences, and technology]*. Berlin: BMBF Bildung.

- Carnevale, A. P., Smith, N., & Melton, M. (2011). *STEM: Science Technology Engineering Mathematics*. Washington DC: Georgetown University Center on Education and the Workforce.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implement Science*, 2(1), 40-49.
- Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, 12(2), 49-66.
- Conley, A. M., Pintrich, P. R., Vekiri, I. & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology*, 29, 186-204.
- Cotabish, A., Dailey, D., Robinson, A., & Hughes, G. (2013). The effects of a STEM intervention on elementary students' science knowledge and skills. *School Science and Mathematics*, 113(5), 215-226.
- Dorsen, J., Carlson, B., & Goodyear, L. (2006). *Connecting informal STEM experiences to career choices: Identifying the pathway*. Waltham: ITEST Learning Resource Center.
- Elder, A. (2002). Characterizing fifth grade students' epistemological beliefs in science. In B. K. Hofer & P. R. Pintrich (Eds.), *The psychology of beliefs about knowledge and knowing* (pp. 347-363). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Emanuel, E. J., Wendler, D., & Grady, C. (2000). What makes clinical research ethical? *Jama*, 283(20), 2701-2711.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138.
- European Commission. (2007). *Science education now: A renewed pedagogy for the future of Europe*. Community Research Expert Group: Vol. 22845. Luxembourg: EUR-OP.
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87(6), 477-531.
- Frank, A. (2005). Naturwissenschaftliches Arbeiten mit der Blackbox [Scientific work with a blackbox]. *Unterricht Biologie 307/308 Beilage*.

- Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*, (pp. 169-190). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88-140.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). Psychology Press.
- Jehng, J.-C., Johnson, S. D., & Anderson, R. C. (1993). Schooling and students' epistemological beliefs about learning. *Contemporary Educational Psychology*, 18(1), 23-35.
- Keeley, P. (2009). Elementary science education in the K-12 system. *Science and Children*, 46(9), 8-9.
- Kienhues, D., Bromme, R., & Stahl, E. (2008). Changing epistemological beliefs: The unexpected impact of a short-term intervention. *British Journal of Educational Psychology*, 78(4), 545-565.
- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht [Development and promotion of interests in school lessons]. *Psychologie in Erziehung und Unterricht*, 44(3), 185-201.
- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Eds.), *Handbook of childhood cognitive development* (pp. 371-393). Oxford: Blackwell.
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29(4), 331-359.
- Lederman, N. G. (2007). Nature of Science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 831-879). New York: Routledge.

- Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences, 44*(7), 1585-1595.
- Litman, J., Hutchins, T., & Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition and Emotion, 19*(4), 559-582.
- Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment, 80*, 75–86.
- Longo, C. (2010). Fostering creativity or teaching to the test? Implications of state testing on the delivery of science instruction. *The Clearing House, 83*(2), 54-57.
- Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science, 1*(4), 316-345.
- Magidson, J. F., Roberts, B. W., Collado-Rodriguez, A., & Lejuez, C. W. (2014). Theory-driven intervention for changing personality: Expectancy value theory, behavioral activation, and conscientiousness. *Developmental Psychology, 50*(5), 1442-1450.
- Maltese, A. V., & Tai, R. H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education, 32*(5), 669-685.
- Metz, K. E. (2008). Narrowing the gulf between the practices of science and the elementary school science classroom. *The Elementary School Journal, 109*(2), 138-161.
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474-496.
- Montgomery, D. E. (1992). Young children's theory of knowing: The development of a folk epistemology. *Developmental Review, 12*(4), 410-430.
- Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review, 18*(1), 3-54.
- Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.

- National Research Council (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. Washington, DC: The National Academies Press.
- National Science Board (2010). *Preparing the next generation of STEM innovators: Identifying and developing our nation's human capital* (NSB-10-33). Retrieved from <http://www.nsf.gov/nsb/publications/2010/nsb1033>.
- Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education, 30*(15), 1977-1999.
- Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance a quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7*(4), 384-403.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84.
- OECD. (2011). *Education at a glance, 2011: OECD indicators*. Paris: OECD.
- Perry, W. G. JR. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart and Winston. Cerca con Google.
- Pintrich, P. R. (2002). Future challenges and directions for theory and research on personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *The psychology of beliefs about knowledge and knowing* (pp. 389-414). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Pratt, H. (2007). Science education's 'overlooked ingredient': Why the path to global competitiveness begins in elementary school. *Education Week, 27*(7), 26-32.
- Qian, G., & Alvermann, D. (1995). Role of epistemological beliefs and learned helplessness in secondary school students' learning science concepts from text. *Journal of Educational Psychology, 87*(2), 282-292.
- Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning, 5*(1), 47-65.
- Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science, 23*(2), 98-103.
- Sawyer, R. K. (2008). Optimising learning implications of learning sciences research. *Innovating to learn, learning to innovate, 45-65*.

- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Smith, C., Maclin, D., Houghton, C. & Hennessey, M. G. (2000). Sixth grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction* 18(3), 349-422.
- Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences: An introduction*. Houndmills: Palgrave Macmillan.
- Torgerson, C. J., & Torgerson, D. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, 49(3), 316-328.
- Tsui, L. (2007). Effective strategies to increase diversity in STEM fields: A review of the research literature. *The Journal of Negro Education*, 555-581.
- Urhahne, D. & Hopf, M. (2004). Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien [Epistemological beliefs in science and their relations to motivation, self-concept, and learning strategies]. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 71-87.
- Valla, J. M., & Williams, W. M. (2012). Increasing achievement and higher-education representation of under-represented groups in science, technology, engineering, and mathematics fields: A review of current K-12 intervention programs. *Journal of Women and Minorities in Science and Engineering*, 18(1), 21-53.
- Veenstra, C. P., Padró, F. F., & Furst-Bowe, J. A. (Eds.). (2012). *Advancing the STEM Agenda: Quality improvement supports STEM: Selected papers from the 2011 Advancing the STEM Agenda in Education, the Workplace and Society Conference at the University of Wisconsin-Stout, July, 2011*. Milwaukee: ASQ Quality Press.
- Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard university press.
- Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology*, 102(4), 860-871.

- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 – Revision (CFT 20-R)* [Basic intelligence test scale 2 - revision]. Göttingen: Hogrefe.
- Wellman, H. M. (1990). *The child's theory of mind* (Vol. 37). Cambridge, MA: MIT press.
- White, B., Frederiksen, J., & Collins, A. (2009). The interplay of scientific inquiry and metacognition: More than a marriage of convenience. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 175-205). New York: Routledge.

Study 3:
Elementary School Children's
Understanding of Science:
The Implementation of an
Extracurricular Science Intervention

Schiefer, J., Golle, J., Tibus, M., Trautwein, U., & Oschatz, K. (2016). Elementary School Children's Understanding of Science: The Implementation of an Extracurricular Science Intervention. *Manuscript submitted for publication.*

This study was funded in part by the Hector Foundation II.

Abstract

The promotion of students' achievement and competence in the so-called STEM disciplines is one cornerstone of current educational research and practice. In particular, as early as elementary school, the fostering of an adequate understanding of science is a normative goal of science education. It facilitates students' science learning and enables them to understand the nature and development of scientific knowledge. Based on the relevance of the promotion of young children's understanding of science, a corresponding science intervention was recently developed and successfully evaluated in a first study under highly controlled conditions. The goal of the present study was to investigate the effectiveness of this intervention when implemented in practice. One hundred seventeen third- and fourth-grade students and 10 trained course instructors participated in this study. We applied a randomized block design with waitlist control groups and repeated measures. The results revealed that children assigned to the intervention compared with children assigned to the waitlist control group showed better inquiry-related methodological competencies (a better understanding of the scientific inquiry cycle and experimentation strategies) and a higher need for cognition. The findings point to the successful implementation of the intervention and are compared with the results of the first study.

Keywords: implementation, science intervention, understanding of science, inquiry cycle, elementary school age, randomized controlled trial

Elementary School Children’s Understanding of Science: The Implementation of an Extracurricular Science Intervention

Science and scientific knowledge are an important part of our culture and play an essential role in our everyday lives (Bybee, 1997; OECD, 2016). To understand the fundamental elements of our world and to be able to participate in socioscientific discussions, it is essential to have not only knowledge and skills in STEM (Science, Technology, Engineering, and Mathematics) but also an understanding of the *nature of science* (Driver, Leach, Millar, & Scott, 1996; Duschl, Schweingruber, & Shouse, 2007; OECD, 2016). An understanding of the nature of science (for reasons of better legibility, we refer to this as an *understanding of science* in the following) includes an understanding of “what science is and how it is done” (McComas, 1998, p. 50). Due to the essential relevance of science, most nations have advocated the development of students’ understanding of science as a normative goal of science education as early as elementary school (e.g., European Commission, 2007).

Several approaches have been developed to increase students’ understanding of science (e.g., Bendixen, 2016; Muis, Trevors, & Chevrier, 2016). In this context, extracurricular interventions are one important part of the educational landscape and complement science education in school (e.g., Valla & Williams, 2012). Interventions can offer an effective way to promote students’ understanding of science (e.g., Bendixen, 2016; Elder, 2002). Also, the European Commission (2007) encouraged the importance of science interventions especially for elementary school children when they are in their “curiosity golden age” (p. 12). In order to promote students’ understanding of science across a broad range, it is important to put effective interventions into practice (Lendrum & Wigelsworth, 2013). However, interventions that have considered not only science content knowledge but also a fundamental understanding of science have been rather rare, especially in the context of elementary school education (e.g., Bendixen, 2016; Elder, 2002; Muis et al., 2016). Furthermore, not many interventions have been successfully evaluated when implemented under real-world conditions (see Fixsen, Blase, Metz, & van Dyke, 2013; Spiel, Schober, & Strohmeier, 2016).

To close this gap, the goal of this study was to analyze the effectiveness of a recently developed science intervention for elementary school children when implemented under real-world conditions. The intervention was part of an extracurricular enrichment program and focused on promoting the understanding of science as well as

the need for cognition and epistemic curiosity. The effectiveness of the intervention under standardized conditions was already demonstrated in a randomized controlled study (Schiefer, Golle, Tibus, et al., 2016). In the current study, we applied a randomized block design with waitlist control groups and tested the effectiveness of the intervention with respect to the same outcomes as in the first study but added instruments to measure further central aspects of the understanding of science.

Outcomes of the Science Intervention

Understanding of science

To date, there is no universal view or standard conceptualization of this broad construct, which can be theoretically located at the intersection of philosophy of science, history of science, sociology of science, and psychology of science (McComas, 1998). Lederman's (1992) operational definition has been cited most often. According to him, the understanding of science refers to the epistemology of science, science as a way of knowing, or the values and beliefs inherent to scientific knowledge and its development (Lederman, 1992; Lederman & Zeidler, 1987). The intervention focused on two constructs that are considered essential for the development of a basic understanding of science: (a) inquiry-based methodological competencies and (b) epistemic beliefs (Elder, 2002; Lederman, 1992, 2007; Osborne, 2013).

Inquiry-based methodological competencies. Inquiry-based methods build the basis for the genesis, construction, and development of knowledge in science. An understanding of these methods is an important aspect of the understanding of science (e.g., Dogan & Abd-El-Khalik, 2008; Ryder & Leach, 2000; Zimmerman, 2007). A basic inquiry-based method is the control and systematic combination of variables (Chen & Klahr, 1999; Zimmerman, 2007). This so-called *control of variables strategy (CVS)* is required for the design of unconfounded experiments. It is relevant for the targeted testing of hypotheses and enables causal inferences to be made from experiments (Simon, 1989; Zimmerman, 2007). Beyond strategies such as the CVS, the understanding of the whole process of scientific inquiry (the so-called *scientific inquiry cycle, SIC*; Kuhn, 2002; White & Frederiksen, 1998; White, Frederiksen, & Collins, 2009; Zimmerman 2007) is a central methodological competence in the context of the understanding of science. The SIC includes the consecutive steps of (a) generating hypotheses on the basis of a specific research question, (b) planning and conducting experiments, (c) collecting data, (d)

computing analyses, (e) evaluating evidence, and (f) drawing inferences. Thus, the SIC subsumes all individual components of scientific inquiry under a meta-perspective. The SIC emphasizes a holistic view as the single components of this cyclical and cumulative process build the basis of knowledge acquisition and change (Kuhn & Franklin, 2006; Zimmerman, 2007). It represents the theory-driven deductive approach that is applied in scientific investigations (e.g., White et al., 2009).

Epistemic beliefs. Besides methodological competencies, the development of sophisticated epistemic beliefs plays an essential role in the development of a profound understanding of science (Elby, Macrander, & Hammer, 2016; Lederman, 2007; Osborne, 2013). Epistemic beliefs are subjective beliefs about the *nature of knowledge* (what one believes knowledge is) and the *nature of knowing* (beliefs about the process through which one comes to know) in science (see Elby et al., 2016; Hofer & Pintrich, 1997; Lederman, 2007). In recent decades, one major line of research has focused on identifying dimensions of epistemic beliefs, and a debate has raged on this issue for a long time (e.g., Chinn, Buckland, & Samarapungavan, 2011; Hofer & Pintrich, 1997; Schommer, 1990, 1994). In our studies, we adhered to Conley, Pintrich, Vekiri, and Harrison's (2004) conceptualization of epistemic beliefs, which was based on previous work by Elder (2002) and Hofer and Pintrich (1997). Conley et al.'s (2004) model as well as their respective questionnaire has focused explicitly on elementary school children (fifth graders). They identified four dimensions: *source*, *certainty*, *development*, and *justification of knowledge*. The source dimension addresses beliefs about the knowledge that resides in external authorities. The certainty dimension reflects beliefs about the (un)changeability of knowledge in the natural sciences. The development dimension is associated with beliefs that recognize science as an evolving subject. Finally, the justification dimension refers to the role of experiments and how students evaluate claims (Conley et al., 2004).

Need for Cognition and Epistemic Curiosity

Conducting scientific inquiry requires active thinking and reasoning (Kuhn, 2002). *Need for cognition* (the tendency to engage in and enjoy thinking; Cacioppo & Petty, 1982) and *epistemic curiosity* (the desire for new knowledge; Litman & Spielberger, 2003) might therefore be important in the context of science learning and inquiry. There is evidence that these constructs positively affect problem solving and

decision-making, are related to exploratory behavior, and motivate individuals to learn new things (e.g., Litman, 2008; Litman, Hutchins, & Russon, 2005; Nair & Ramnarayan, 2000; Peltier & Schibrowsky, 1994; Richter & Schmid, 2010). In particular, need for cognition has been considered an *epistemic motive* (Oschatz, 2011), an individual disposition for the willingness to engage in thinking. A high level of need for cognition points to high cognitive motivation (Fleischhauer et al., 2010) and is an important prerequisite for making an effort to examine and solve scientific problems. Need for cognition and epistemic curiosity have been described as stable personality traits but can develop already in childhood in response to environmental factors and can be affected by interventions (e.g., Caspi & Roberts, 2001; Magidson, Roberts, Collado-Rodriguez, & Lejuez, 2014).

Implementing Interventions in the Real World

Based on the requirement and development of science interventions, questions concerning the successful implementation of such programs are a major focus of educational research and practice (Hulleman & Cordray, 2009; Lendrum & Wigelsworth, 2013; McDonald, Keesler, Kauffman, & Schneider, 2006). Putting an intervention into practice offers a great challenge (Lendrum & Humphrey, 2012). For the successful implementation of an intervention, its effectiveness and practicability should be demonstrated at different stages between its development and broad dissemination in practice (Humphrey et al., 2016). It is particularly important to investigate whether an intervention is effective under real-world conditions (Durlak, 1998; Gottfredson et al., 2015), for example, when it is implemented by the staff and resources that are normally available (Dane & Schneider, 1998; Greenberg, Domitrovich, Graczyk, & Zins, 2005). Thus, it is important to investigate factors that affect the implementation of an intervention such as the implementer's characteristics (e.g., education, skills, attitudes, and experiences) as well as implementation fidelity (i.e., Hulleman & Cordray, 2009; Humphrey et al., 2016; Rockoff, 2004; Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). Implementation fidelity is the degree to which an intervention is delivered as intended (Carroll et al., 2007). By understanding and measuring whether an intervention has been implemented with fidelity, researchers gain a better understanding of "how and why an intervention works" (Carroll et al., 2007, p. 1).

The Present Study

The goal of the present study was the practical implementation of a science intervention for elementary school children, which was recently developed by a team of researchers at a university as part of a statewide enrichment program in southwest Germany (*Hector Children's Academy Program, HCAP*). The effectiveness of the intervention was already analyzed in a first effectiveness study with 65 children under standardized conditions: It was held by three program developers from the university who followed the manual strictly (Schiefer, Golle, Tibus, et al., 2016). Specifically, a randomized block design with a treated control group (who participated in a parallel course, a speech training) was used in the first study. Positive effects of the intervention were found on children's epistemic beliefs ($.18 < ES < .61$) and epistemic curiosity ($ES = .34$). The current study involved a larger sample ($N = 117$ elementary school children), and the intervention was offered under real-world conditions by 10 HCAP course instructors—teachers and course instructors with a professional background in the natural sciences—who usually taught the courses in the HCAP. To maximize implementation fidelity, they participated in a mandatory training program and were given a detailed course manual and all required materials. We applied again a randomized block design with pretest and posttest, but used instead of a treated control group—who also dealt in part with scientific topics—a waitlist control group to estimate the treatment effects more ecologically valid and simple. The effectiveness of the intervention was assessed with respect to the same outcomes as in the first study, but the current study used additional instruments that were required to assess central aspects of children's understanding of science (see Schiefer, Golle, & Oschatz, 2016). These were the understanding of the SIC, experimentation strategies, and need for cognition.

The science intervention was the same as in the first study. It was intended to foster children's inquiry-based methodological competencies (experimentation strategies and understanding of the SIC) and epistemic beliefs. Both constructs are essential for the development of a basic understanding of science (Elder, 2002; Lederman, 1992, 2007; Osborne, 2013). The intervention included learning settings that allowed students to participate actively, that is, by means of inquiry-based learning approaches, active experimentation and testing of hypotheses, application of the CVS, and working scientifically according to the SIC. Thus, we expected to find positive intervention effects on inquiry-based methodological competencies. Consequently, we hypothesized that after

participating in the intervention, the children would show a better understanding of the SIC and experimentation strategies than the children in the waitlist control group (Hypothesis 1). Second, the intervention included reflections on epistemological issues (i.e., by means of discussions, science communication, teaching, and critical scrutiny). Specifically, we expected that children in the intervention would develop more sophisticated epistemic beliefs than children in the control group (Hypothesis 2). Finally, the intervention included intellectually challenging elements with the aim of engaging children in critical thinking and reflection on scientific problems and was therefore expected to help eliminate information gaps. These aspects are associated with need for cognition and epistemic curiosity. Consequently, we hypothesized that after participating in the intervention, the children would show a higher level of need for cognition and epistemic curiosity than the control group (Hypothesis 3).

Method

Participants

Students. Data were collected from 117 elementary school children who participated in the HCAP (71.2% male, age: $M = 8.89$, $SD = 0.82$, Grade 2: $N = 9$, Grade 3: $N = 54$, Grade 4: $N = 54$), which provides extracurricular enrichment courses for talented elementary school children. To take part in the program, children have to be nominated by their teachers. At 61 local sites, children can choose from a variety of afternoon courses, which are taught not only by teachers but also by a large number of external course instructors who have different kinds of professional backgrounds (e.g., computer scientists, engineers, natural scientists). Ten local HCAP sites participated in the study. Overall, the participating children were from 68 different schools and 82 classes. The intervention group consisted of 58 children, and the control group consisted of 59 children (see Table 1 for the baseline demographics). Written parental consent was required for the children's participation in the study.

Course instructors. Ten course instructors (two men, eight women, age: $M = 46.40$, $SD = 11.15$) from the respective local sites participated in this study. They underwent a 1-day mandatory preparatory training program about the course. A scientist from the university who developed and taught the course several times conducted this training. The training included a theoretical introduction to the understanding of science, detailed insights into the promotion goals of each course unit, hands-on exercises, as well

as insights into the structure of the individual course units. All course instructors received a detailed course manual that included a comprehensive introduction to the theoretical background of each course unit, the promotion goals, all required materials, worksheets for the children, as well as a detailed schedule for each course unit (including time schedule, goals, execution, and materials). They were instructed to document the implementation fidelity (O'Donnell, 2008) of the single course units (see more details below). Six course instructors had a pedagogical qualification. Five of them had a background in the natural sciences and had already worked scientifically. With the exception of one person, all course instructors had experience teaching elementary school children and had already taught at a local HCAP site (between five and 40 courses, $M = 19.22$, $SD = 11.42$). This constellation of course instructors corresponded to the usual selection of HCAP instructors and made research under “real-world” conditions (see Lendrum & Wigelsworth, 2013) possible. Before the study started, all course instructors gave written consent for their participation.

Table 1

Description of the Sample

Group	Male	Age	Grade 2	Grade 3	Grade 4
Intervention group ($N = 58$)	69.0%	$M = 8.86$ ($SD = 0.82$)	$N = 5$	$N = 27$	$N = 26$
Control group ($N = 59$)	74.6%	$M = 8.92$ ($SD = 0.83$)	$N = 4$	$N = 27$	$N = 28$

Note. The intervention was developed and written for children in Grades 3 and 4. Nevertheless, some children in Grade 2 participated as they were in the process of skipping a grade or were nominated by their teachers because of an extraordinary talent or interest in science that might be comparable to the abilities of children in Grades 3 and 4.

Experimental Design

We investigated the effectiveness of the intervention (see the “description of the intervention” section below for details) by using a randomized block design with waitlist control groups and repeated measures: pretest (T1) and posttest (T2). Data collection was integrated into the first (T1) and last (T2) course sessions of the intervention. Children from both groups participated in these sessions. The children in the intervention group met for 90 min one afternoon per week for 8 weeks. The children in the waitlist control group participated at T1 and T2 (to participate in the data collection) within the respective local sites and received a compacted block course that covered all relevant course elements on the weekend after T2. This practice corresponded to the ethical principles for intervention studies (Emanuel, Wendler, & Grady, 2000) because all participants received the treatment. It also minimized the risk of attrition in the control group between the pretest and posttest. We controlled for the activities of the control group (e.g., participation in other courses or science activities) by administering a questionnaire to parents at T1 and T2.

The respective courses consisted of four to 10 children. If a total of fewer than eight children were registered for a course, we did not split the children into two groups at this local site (the implementation of the course was not possible with fewer than four participants) but cluster randomized this local site and assigned all course participants to either the intervention or the waitlist control group. Nine of the 10 local sites reached the required number of participants. One local site had only five applicants for the course. This group was cluster randomized to the control group.

The courses took place over the summer semester 2015. After course registration, the children within each participating local site were randomly assigned to either the intervention or the control group. Using a random number generator, the randomization was conducted by a neutral person before the intervention began and did not involve any restrictions (e.g., grade level or gender). Participating children, their parents, as well as the course instructors were informed about their group membership after the T1 session. This procedure was chosen to avoid any differences regarding children’s motivation or expectations between the two groups (Torgerson & Torgerson, 2001, 2008). At T1 and T2, trained research assistants administered the questionnaires. They were blind to the group membership of the children. The course instructors stayed in class during data collection.

Description of the Intervention

The intervention—a course titled *Little Researchers—We Work Like Scientists*—was developed, tested, and evaluated by scientists in the field of education sciences, psychology, and science education (Oschatz & Schiefer, in press; Schiefer, Golle, Tibus, et al., 2016). Single course units (see Figure 1) were arranged in such a way that children experienced and applied the cumulative and cyclical process of scientific research in each course unit (Conley et al., 2004; Kuhn, 2002; White et al., 2009). On the basis of theories and derived hypotheses about their research themes, the children conducted experiments, analyzed data, evaluated evidence, presented results, and drew inferences with the goal of generating or revising the theories. This process was intended to allow them to gain insights into scientific work according to the SIC (see White et al., 2009; Zimmerman, 2007). Thereby, they experienced science as a hypothesis-driven and evolving discipline, which is—besides the development of general methodological competencies—also relevant for the development of sophisticated epistemic beliefs. The promotion of more sophisticated epistemic beliefs was reinforced by the explicit integration of conflicting information and results (i.e., due to the use of different materials or different methods of investigation). This method was intended to support the children’s critical thinking about the subjectivity and the boundaries of empirical research (see Kienhues, Bromme, & Stahl, 2008). Further course elements for the fostering of sophisticated epistemic beliefs included reflections about relevant aspects of the epistemology of science by means of discussions, science communication, teaching, and critical scrutiny (set forth by Akerson & Hanuscin, 2007; Blanchard et al., 2010; Conley et al., 2004).

Hands-on activities as well as inquiry-based learning have revealed positive effects on the enhancement of inquiry-based methodological competence as well as epistemic beliefs in past research (Blanchard et al., 2010; Elder, 2002). Therefore these elements were also integrated into the course units, for example, experiments on the human senses, experiments in a student lab for neuroscience, examination of an unknown object—a so-called “black box” (Frank, 2005), and simple physical experiments. The CVS (Chen & Klahr, 1999) was introduced to the children by the use of images with combinations of object characteristics (e.g., the nose, wings, and elevator of an aircraft; according to the material developed by Bullock & Ziegler, 1999). Afterwards, they were given the opportunity to apply the CVS in practice in simple physical experiments (i.e., the children were asked which elements they would have to manipulate in an experiment

to determine whether the weight of a car would have an impact on its speed when driving down a ramp).

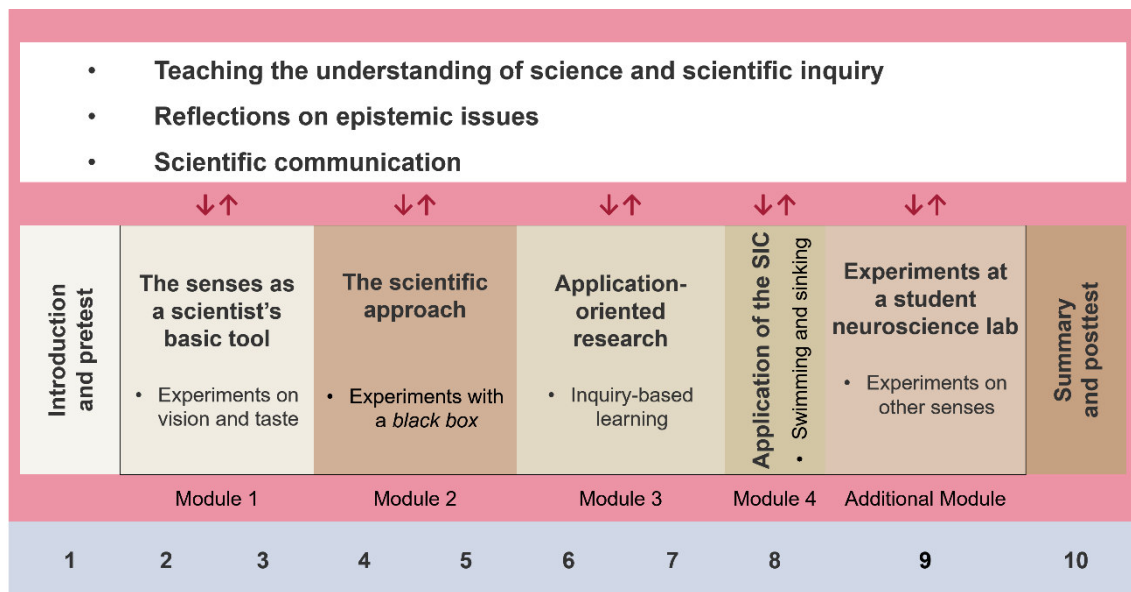


Figure 1. Course concept for the intervention *Little Researchers—We Work Like Scientists*.

Implementation fidelity. Assessments of implementation fidelity strongly depend on the particular intervention (Abry, Hulleman, & Rimm-Kaufman, 2015). In this study, we assessed adherence (i.e., compliance) to the elements in the course manual (Humphrey et al., 2016). For this purpose, all course teachers were instructed to provide written feedback on their implementation of the course units with the use of a self-developed feedback questionnaire. For each course element, the instructors reported whether the element was carried out or not (item: “Was the course element conducted?”; dummy-coded: 0 = *no*, 1 = *yes*). As the relevance of the single course elements differed in their importance for reaching the instructional goals, each course element was additionally weighted by the course developers (0 = element has no relevance for the understanding of science, i.e., getting-to-know you game; 1 = element is associated with a scientific topic, i.e., information about the human senses; 2 = element refers to an implicit understanding of science, i.e., conducting experiments; 3 = element refers to an explicit understanding of science, i.e., reflecting on the results of the conducted experiments). We chose this procedure to adapt the theoretical

guidelines regarding the assessment of implementation fidelity (see Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; Hulleman & Cordray, 2009; McGrew, Bond, Dietzen, & Salyers, 1994). To measure implementation fidelity, the percentage of course elements that were conducted was calculated for each course instructor.

Measures

Children

All administered scales and the corresponding descriptive statistics, Cronbach's alphas, numbers of items, and examples are presented in Table 2.

SIC test. To assess the understanding of the SIC, we used a previously developed and IRT scaled instrument (SIC test¹; see Schiefer, Golle, & Oschatz, 2016). This instrument consisted of 12 items that were scored dichotomously (0 = *not correct*, 1 = *correct*). The tasks required (a) the active reconstruction of the sequences of all steps of the SIC and (b) an understanding of the consecutive next steps of the cycle within a given inquiry process (see Figures 2, 3, and 4 in the Appendix for item examples).

Intelligence. Fluid intelligence was measured with the nonverbal *fluid intelligence* subscale from the *BEFKI-short* (Schroeders, Schipolowski, Zettler, Golle, & Wilhelm, 2016). It consists of 16 items. Within a time limit of 15 min, children had to complete figural patterns (see Figure 5 in the Appendix for an example item). Sum scores were calculated for further analyses.

Experimentation strategies. Experimentation strategies were assessed with six single-choice items with three answer alternatives (one correct, two misconceptions). The items focused on the CVS (Chen & Klahr, 1999; Zimmerman, 2007). As no published (German) test for assessing experimentation strategies exists, we used three items from research projects by Mayer, Sodian, Koerber, and Schwippert (2014) and developed three other items with the same format (following Mayer et al., 2014, and Ehmer, 2008, see Figure 6 in the Appendix for an example item). The items were scored dichotomously. Sum scores were used in further analyses.

¹ The SIC test had been scaled with the Birnbaum measurement model as a two-parameter logistic model for dichotomous items (2PL model; Birnbaum, 1968). The model fit revealed acceptable results (RMSEA = .035; $\chi^2/df = 2.10$, $\chi^2(54) = 113.662$, $p < .001$, CFI = .89, TLI = .86, see Schermelleh-Engel, Moosbrugger, & Müller, 2003, for recommendations for model evaluation). The SIC items showed an acceptable overall marginal EAP reliability of .64.

Table 2

Descriptive Statistics for the Scales: Means, Standard Deviations, Internal Consistencies, Numbers of Items, and Examples

Construct	T1			T2			Number of items and example				
	N	M	SD	α	N	M		SD	α		
Methodological competencies											
SIC test (Schiefer, Golle, & Oschatz, 2016)	IG	56	-0.19 ^b	1.01	.64 ^c	51	0.51 ^b	1.25	12	See Figures 2, 3, and 4	
	CG	53	-0.28 ^b	0.78		47	0.35 ^b	1.00			
Strategies of experimentation (according to Mayer et al., 2014)	IG	56	3.92	1.52	.51	52	4.69	1.23	.81	6	See Figure 6
	CG	57	4.12	1.45		48	4.38	1.43			
Epistemic beliefs (Conley et al., 2004)											
Source of knowledge ^a	IG	56	2.74	0.56	.72	52	3.00	0.55	.70	5	Everybody has to believe what scientists say.
	CG	57	2.79	0.66		49	3.00	0.62			
Certainty of knowledge ^a	IG	57	2.77	0.55	.68	52	2.97	0.51	.73	6	All questions in science have one right answer.
	CG	57	2.71	0.50		49	2.85	0.64			Ideas in science sometimes change.
Development of knowledge	IG	57	3.29	0.48	.69	52	3.31	0.48	.66	6	
	CG	57	3.34	0.50		49	3.34	0.41			
Justification of knowledge	IG	57	3.50	0.34	.60	52	3.44	0.39	.76	9	It is good to have an idea before you start an experiment.
	CG	57	3.43	0.33		49	3.38	0.44			I like to solve tricky tasks.
Need for cognition (Baudson et al., 2012)	IG	56	3.09	0.74	.88	52	3.14	0.77	.93	6	
	CG	57	3.21	0.64		48	2.99	0.80			
Epistemic curiosity (Litman, 2008)	IG	56	3.00 [*]	0.52	.86	52	2.87	0.56	.91	14	I always like to learn something new.
	CG	57	3.18 [*]	0.38		48	2.97	0.53			
Figural intelligence (Schroeders et al., 2016)	IG	57	9.56	3.23	.76	-	-	-	-	16	See Figure 5
	CG	57	9.11	3.37		-	-	-	-		

Note. *N* = Number of participating children, *M* = mean, *SD* = standard deviation, α = Cronbach's alpha. Measurement time points: T1 = February to April 2015, T2 = May to July 2015. IG = intervention group, CG = waitlist control group. SIC = scientific inquiry cycle. *t*-tests for independent samples (IBM SPSS, version 22) were computed to test for significant differences between the IG and the CG at T1.

^aItems were recoded so that higher scores reflected more sophisticated beliefs. ^bEAP score. ^cEAP reliability.

**p* < .05. (Significant differences occurred between the two groups at T1)

Epistemic beliefs. Epistemic beliefs were assessed with a 26-item instrument (Conley et al., 2004; German version by Urhahne & Hopf, 2004). Four subscales reflect the dimensions source, certainty, development, and justification of knowledge. Items were rated on a 4-point Likert scale ranging from 1 (*I completely disagree*) to 4 (*I completely agree*), presented via stars of increasing size. The source and certainty scales were recoded for the analyses. Thus, for each scale, higher scores reflected more sophisticated beliefs. Example items can be found in Table 2.

Need for cognition. Need for cognition was assessed with a short version of an instrument developed by Budson, Strobel, and Preckel (2012). The six items address a student's tendency to engage in and enjoy thinking (see Cacioppo & Petty, 1982). Items were rated on a 4-point Likert scale ranging from 1 (*never*) to 4 (*always*).

Epistemic curiosity. Epistemic curiosity was assessed with an instrument developed by Litman (2008). The items address the desire for knowledge that motivates individuals to learn something new and to solve intellectual problems. Items were rated on a 4-point Likert scale ranging from 1 (*never*) to 4 (*always*).

Parents and Course Instructors

In addition to assessing the variables described above for the children, we assessed a variety of variables that pertained to the parents (i.e., demographics, level of education, socioeconomic status, reason for registering their child in the course) and the course instructors (e.g., demographics, pedagogical experience, professional background, prior knowledge, interest in science, epistemic beliefs, understanding of science).

Statistical Analyses

Possible group differences at T1 were examined with t-tests for independent samples in IBM SPSS (version 22). The effectiveness of the intervention was analyzed with multiple regression analyses in Mplus (Muthén & Muthén, 1998-2012). All analyses used the robust maximum likelihood estimator (MLR), which corrects standard errors for the non-normality of the variables (Muthén & Muthén, 1998-2012). To correct for the clustering of the data (children nested in Hector courses), we used *type = complex* for all analyses (Muthén & Muthén, 1998-2012). For experimentation strategies, need for cognition, epistemic curiosity, and epistemic beliefs, the dependent variables were the respective z-standardized posttest measures. The predictors in these regression models

were group assignment (0 = waitlist control group, 1 = intervention group), and the corresponding z-standardized pretest scores (see Enders & Tofighi, 2007). Due to the standardization of these dependent variables, the regression coefficient of the group variable indicated the standardized intervention effect (effect size, ES) controlled for the corresponding pretest scores. To analyze the effects of the intervention on the SIC performances², latent regression analyses were computed. For the SIC test, effect sizes were calculated by dividing the regression coefficient by the standard deviation of the T2 performances of the SIC. One-tailed tests of significance ($\alpha = .05$) were used in all analyses used for the treatment because we formulated directional hypotheses about the effects of the intervention. To estimate differential intervention effects due to the respective pretest scores, interactions between group assignment and the pretest scores were added to the models. In a further step, we controlled for gender and intelligence in all analyses.

Missing data. Overall, 117 children participated in the study: 114 of them participated at T1, and 101 children participated at T2. Due to illness, three children were not able to participate at T1 but came to T2. In the intervention group (IG), 56 children participated at T1 and 51 at T2 (91.07%). In the control group (CG), 58 children participated at T1 and 50 at T2 (86.21%). There was no differential drop-out between the two groups on any of the instruments used in the present study (see Table 2 for the exact number of participants for each measure). To analyze the intervention effects, we used the full information maximum likelihood approach implemented in Mplus to deal with the missing values (Muthén & Muthén, 1998-2012). All measured variables were taken into account to estimate the model parameters (Schafer & Graham, 2002).

Results

Descriptive Statistics and Bivariate Correlations

In a first step, we analyzed the characteristics and differences between the intervention and control groups at T1 (see Table 2). There were no IQ or gender differences between the intervention and control groups. There were also no differences between the groups in their performances on the SIC test, their understanding of

² Prior to the analyses, children's latent EAP scores at T1 and T2 were estimated by using the item parameters of the 2PL model from the pilot sample (see Schiefer, Golle, & Oschatz, 2016).

experimentation strategies, their epistemic beliefs, or their need for cognition. However, we found differences between the two groups at T1 in epistemic curiosity, $t(111) = -2.19$, $p = .031$, in favor of the control group.

The correlations between all outcome variables are shown in Table 3. At T1, the SIC performances were positively correlated with experimentation strategies and certainty of knowledge and negatively correlated with source of knowledge and epistemic curiosity. Experimentation strategies were positively correlated with the dimensions source, certainty, and development of knowledge as well as fluid intelligence. Beliefs about source of knowledge were positively correlated with beliefs about certainty of knowledge, and development of knowledge was positively correlated with justification of knowledge (see the coding of the items). Epistemic curiosity was positively correlated with development and justification of knowledge as well as need for cognition. Correlations at T2 showed similar patterns with just a few exceptions.

Table 3

Correlations between the Dependent Variables at the Pretest (Below Diagonal) and the Posttest (Above Diagonal)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) SIC test		.58***	.06	.27	-.18	.17	.19	.16	-.11
(2) Experimentation strategies	.32**		.08	.16	.10	.30**	.32**	.27**	.45**
(3) Source of knowledge	-.26*	.20*		.60**	.38**	.05	.02	-.10	.22*
(4) Certainty of knowledge	.33**	.35**	.52*		.40**	-.01	-.11	-.06	.23*
(5) Development of knowledge	.00	.27**	.07	.08		.34**	.08	-.01	.18
(6) Justification of knowledge	.13	.18	.04	-.00	.57**		.45**	.28**	.25*
(7) Epistemic curiosity	-.28**	.18	-.06	-.08	.22*	.31**		.66**	.24*
(8) Need for cognition	.08	.09	.15	.04	.15	.24**	.53**		.26**
(9) Fluid intelligence	.20	.27**	-.02	.14	.17	.07	.06	.25**	

Note. SIC = scientific inquiry cycle. Latent EAP (expected a posteriori) scores were used for analyses.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Implementation fidelity. The analyses of the implementation fidelity of the intervention (see Hulleman & Cordray, 2009; Humphrey et al., 2016; O'Donnell, 2008) revealed that most of the course instructors kept to the program and appropriately presented most of the course elements. Table 4 summarizes the fidelity scores for each of the course instructors. Scores ranged between 57% and 98% ($M = 90.63\%$, $SD = 13.86$). We analyzed the intervention effects also without the groups taught by Course Instructors 8 (low fidelity) and 9 (no information about treatment fidelity available). The results remained stable, and there were no additional intervention effects when these local sites were excluded from the analyses. However, due to the small sample size ($N = 10$), it was not possible to include the fidelity scores in the regression analyses (e.g., as mediator).

Table 4

Implementation Fidelity

Course instructor	Course unit (weeks)							Total
	2	3	4	5	6	7	8	
1	87	100	100	77	100	100	100	95
2	87	100	100	100	100	100	100	98
3	100	100	100	100	100	100	100	100
4	100	100	100	92	77	77	100	92
5	87	100	100	92	100	100	93	96
6	100	85	100	92	92	100	93	95
7	100	100	100	62	92	100	93	92
8	20	100	71	77	46	77	7	57
9	<i>na</i>	<i>na</i>	<i>na</i>	<i>na</i>	<i>na</i>	<i>na</i>	<i>na</i>	<i>na</i>
<i>M</i>	85.13	98.13	96.38	86.50	88.38	94.25	85.75	90.63
<i>(SD)</i>	(27.09)	(5.30)	(10.25)	(13.29)	(18.87)	(10.65)	(32.01)	(13.86)

Note. Percentage of course elements implemented by each course instructor and in each course unit. Course Units 1 and 9 are missing because the pretest and posttest were administered during these units. The data from Course Instructor 9 are not available (*na*) because he did not fill out the fidelity questionnaire. Course Instructor 10 is missing as this instructor offered the block course only for the waitlist control group due to cluster randomization.

Table 5

Variables	SIC (T2) ^{c, d}		Experimentation Strategies (T2) ^a		Need for Cognition (T2) ^a		Epistemic Curiosity (T2) ^a	
	B	SE	B	SE	B	SE	B	SE
Model 1								
Treatment ^b	0.53 ^{**}	(0.22)	0.37 [*]	(0.20)	0.25 ^{**}	(.10)	0.00	(.12)
Pretest score ^a	1.21 ^{***}	(0.16)	0.35 ^{***}	(0.07)	0.63 ^{***}	(.04)	0.58 ^{***}	(.07)
Explained variance (R ²)	.95		.16		.40		.33	
Model 2								
Treatment ^b	0.82 ^{***}	(0.24)	0.37 [*]	(0.22)	0.25 ^{**}	(.10)	0.00	(.13)
Pretest score ^a	0.91 ^{***}	(0.23)	0.44 ^{***}	(0.10)	0.56 ^{***}	(.14)	0.46 ^{**}	(.21)
Treatment ^b x Pretest Score ^a	0.58 [*]	(0.23)	-0.17	(0.10)	0.20	(.20)	0.19	(.27)
Explained variance (R ²)	.41		.16		.41		.34	

Note. SIC = scientific inquiry cycle. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention. ^cLatent regression analyses were computed to investigate the intervention effects on the SIC test. ^dTo test the interaction between the latent SIC T1 score and course participation (Model 2), we used *type is complex random*. This analysis does not provide an estimation of the explained variance.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6

Course Effects on Methodological Competencies and Cognitive Motivation: Predicting Children's Posttest Measures (T2) Controlling for Gender and IQ

Variables	SIC (T2) ^{d, e}		Experimentation Strategies (T2) ^a		Need for Cognition (T2) ^a		Epistemic Curiosity (T2) ^a	
	B	SE	B	SE	B	SE	B	SE
Treatment ^b	0.73 ^{**}	(.24)	0.34 [*]	(.20)	0.24 ^{**}	(.09)	-0.05	(.12)
Pretest score ^a	0.81 ^{***}	(.20)	0.37 ^{**}	(.12)	0.53 ^{***}	(.15)	0.42 [*]	(.20)
Treatment ^b x Pretest Score ^a	0.55 [*]	(.23)	-0.13	(.11)	0.12	(.20)	0.18	(.24)
Gender ^c	-0.33	(.31)	-0.07	(.19)	0.01	(.18)	-0.20	(.14)
IQ ^a	0.21 [*]	(.09)	0.15	(.10)	0.11	(.07)	0.17 [*]	(.08)
Explained variance (R ²)			.19		.41		.37	

Note. SIC = scientific inquiry cycle. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention. ^cGender was dummy-coded 0 = girls, 1 = boys. ^dLatent regression analyses were computed to investigate the intervention effects on the SIC test. ^eTo test the interaction between the latent SIC T1 score and course participation, we used *type is complex random*. This analysis does not provide an estimation of the explained variance.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Effects of the Intervention on Inquiry-Based Methodological Competencies

The first hypothesis concerned the intervention's enhancement of children's inquiry-based methodological competencies. Regression models for children's SIC performances (latent) and experimentation strategies (manifest) were used to assess the effectiveness of the program. The predictors consisted of group assignment and the pretest score. The results are presented in Table 5. The findings showed that the children assigned to the intervention exhibited better performances on the SIC test and a better understanding of experimentation strategies than the children assigned to the control group. The children in the intervention scored significantly higher on the posttest measures of the SIC test ($B = 0.53, p = .014; ES = 0.32$) and experimentation strategies ($B = 0.37, p = .033$) than the children in the control group. Thus, the intervention participants exhibited a better understanding of the process of scientific inquiry as well as of the designing of controlled experiments. To investigate whether the intervention effects depended on children's initial methodological knowledge, interactions between group assignment and the pretest scores were additionally included in the regression analyses (see Table 5, Model 2). The analyses revealed a significant positive interaction ($B = 0.58, p = .010$) between course participation and the values of the SIC test at T1. This means that children with higher scores on the SIC test at T1 benefitted more from the intervention than children with lower scores at T1. The results remained stable when we additionally controlled for participants' intelligence and gender (see Table 6). Overall, these findings provide evidence that it was possible to foster inquiry-related methodological competencies among the participants of the intervention program.

Effects of the Intervention on the Development of Epistemic Beliefs

The second hypothesis concerned the intervention's enhancement of epistemic beliefs, which are essential for an adequate understanding of science. Regression models were calculated for each epistemic belief subscale. The results are presented in Table 7. The findings showed that the children assigned to the intervention did not exhibit more sophisticated epistemic beliefs than the children assigned to the control group. None of the four dimensions was positively affected by the intervention. The children in the intervention group did not score significantly higher on the posttest measures of source of knowledge ($B = 0.07, p = .315$), certainty of knowledge ($B = 0.20, p = .102$), development of knowledge ($B = -0.03, p = .440$), or justification of knowledge ($B = 0.07,$

$p = .262$) than the children in the control group. To investigate whether any intervention effects depended on children's initial epistemic belief scores, interactions between the group variable and each pretest score were additionally included in the regression analyses (see Table 7, Model 2). The only significant interaction was between course participation and the justification of knowledge pretest scores ($B = 0.29, p = .042$); that is, children in the intervention group with higher pretest scores benefitted more from the intervention than the children in the intervention group with low pretest scores. The results remained stable when we additionally controlled for participants' intelligence and gender (see Table 8). Overall, these findings revealed that children's epistemic beliefs were not affected by the intervention in this study.

Effects of the Intervention on Need for Cognition and Epistemic Curiosity

Our third hypothesis concerned whether children's need for cognition and epistemic curiosity could be enhanced by the intervention. The results of the regression analyses are presented in Table 5. The children assigned to the intervention scored significantly higher on need for cognition ($B = 0.25, p = .005$) than the children in the waitlist control group. This means that the intervention participants reported a greater tendency to engage in and enjoy thinking. No intervention effect was found for epistemic curiosity ($B < 0.01, p = .494$). After the intervention, the two groups did not differ in their desire for knowledge or in their motivation to learn something new. The analyses did not reveal any significant interactions between course participation and the pretest scores (see Table 5, Model 2). The results remained stable when we additionally controlled for participants' intelligence and gender (see Table 6). Overall, the findings provide evidence that it was possible to foster a need for cognition in the participants of the intervention program.

Table 7

Course Effects on Epistemic Beliefs: Predicting Children's Posttest Measures (T2)

Variables	Source of Knowledge (T2) ^a		Certainty of Knowledge (T2) ^a		Development of Knowledge (T2) ^a		Justification of Knowledge (T2) ^a	
	B	SE	B	SE	B	SE	B	SE
Model 1								
Treatment ^b	0.07	(.15)	0.20	(.16)	-0.03	(.17)	0.07	(.11)
Pretest score ^a	0.54 ^{***}	(.08)	0.47 ^{***}	(.06)	0.48 ^{***}	(.07)	0.29 ^{**}	(.11)
Explained variance (R ²)	.29		.23		.23		.09	
Model 2								
Treatment ^b	0.08	(.15)	0.21	(.16)	-0.02	(.16)	0.07	(.12)
Pretest score ^a	0.65 ^{***}	(.12)	0.52 ^{***}	(.12)	0.42 ^{***}	(.11)	0.13	(.11)
Treatment ^b x Pretest Score ^a	-0.26	(.13)	-0.10	(.18)	0.09	(.13)	0.29 [*]	(.14)
Explained variance (R ²)	.31		.24		.22		.11	

Note. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 8

Course Effects on Epistemic Beliefs: Predicting Children's Posttest Measures (T2) Controlling for Gender and IQ

Variables	Source of Knowledge (T2) ^a		Certainty of Knowledge (T2) ^a		Development of Knowledge (T2) ^a		Justification of Knowledge (T2) ^a	
	B	SE	B	SE	B	SE	B	SE
Treatment ^b	0.06	(.15)	0.18	(.17)	-0.03	(.17)	0.04	(.12)
Pretest score ^a	0.66 ^{***}	(.11)	0.51 ^{***}	(.12)	0.41 ^{***}	(.12)	0.15	(.10)
Treatment ^b x Pretest Score ^a	-0.27 [*]	(.13)	-0.12	(.18)	0.09	(.14)	0.24	(.13)
Gender ^c	-0.02	(.13)	0.01	(.15)	0.08	(.23)	-0.19	(.14)
IQ ^a	0.15 [*]	(.07)	0.19 [*]	(.09)	0.06	(.09)	0.13	(.08)
Explained variance (R ²)	.34		.28		.23		.13	

Note. For the pretest score and the treatment, one-tailed significance levels are reported because we tested directional hypotheses.

^aVariables were z-standardized prior to the analyses. ^bThe treatment was dummy-coded 0 = control group, 1 = intervention. ^cGender was dummy-coded 0 = girls, 1 = boys.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This study tested whether a recently developed and evaluated science intervention for elementary school children (Oschatz & Schiefer, in press; Schiefer, Golle, Tibus, et al., 2016) could be put into practice by nonresearch HCAP course instructors. Treatment effects were assessed by applying a randomized block design with waitlist control groups, a design that is considered the gold standard in educational research (Torgerson & Torgerson, 2013). The target outcomes of the science intervention were children's understanding of science (inquiry-based methodological competencies and epistemic beliefs), epistemic curiosity, and need for cognition.

Putting the Intervention into Practice

The results of our study revealed that it was possible to put the intervention into practice. In the first effectiveness study (Schiefer, Golle, Tibus, et al., 2016), the intervention was conducted by scientists who developed the intervention and strictly adhered to the manual. In the present study, the program was implemented by nonresearch course instructors from the HCAP with a variety of professional backgrounds and pedagogical experience. This shift in instructors presented a great challenge as it was necessary to ensure that they would completely adhere to the intervention as outlined and would not modify it (Humphrey et al., 2016). To reach this goal, the course instructors of the HCAP participated in a mandatory 1-day training program and worked with a detailed course manual. Their self-report of adherence revealed that most of them kept to the program and implemented most of the course elements. Compared with other investigations, the implementation fidelity of this study can be considered quite satisfactory (e.g., Durlak & DuPre, 2008). This formed the basis for further analyzing and interpreting the effects of the intervention. In the following, the intervention effects will be discussed with regard to their educational relevance and—if possible—compared with the results of the first effectiveness study (Schiefer, Golle, Tibus, et al., 2016). In this regard, it must be noted, that it was—for practical reasons—not possible to use identical instruments in the two studies. Therefore, not all outcomes can be compared. Furthermore, the control group was—in contrast to the first study, which used a treated control group—a waitlist control group, which enables to estimate the treatment effects more ecologically valid and simple. However, the effect sizes cannot be compared directly as studies using a waitlist control group are supposed to produce stronger effects

than studies using treated control groups, in particular when the topics overlap to some extent.

Effects of the Intervention on the Intended Outcomes

Inquiry-based methodological competencies

The science intervention positively affected the participants' inquiry-based methodological competencies. The intervention served to enhance children's understanding of the SIC and application of experimentation strategies. Specifically, children with a high prior knowledge of the SIC benefitted from the intervention and developed a deeper understanding of the SIC compared with children with lower prior knowledge. Although mature scientific inquiry does not necessarily proceed in a fixed order (see Pedaste et al., 2015), an understanding of the steps of the SIC offers an effective initial model for scientific investigations and a prerequisite for targeted empirical research (see Kuhn & Franklin, 2006). An early understanding of this deductive hypothesis-driven approach is an important basis for students' later science learning and was successfully promoted by the intervention. Besides an understanding of the SIC, children in the intervention improved their understanding of experimentation strategies. Thus, they understood the relevance of the CVS, which is essential for the drawing of valid inferences from experiments (Chen & Klahr, 1999; Zimmerman, 2007). Our results are in line with previous research that demonstrated that instruction in CVS can lead to a significant improvement in the ability to design simple, unconfounded experiments as early as elementary school (Bullock & Ziegler, 1999; Chen & Klahr, 1999; Klahr & Nigam, 2004).

Epistemic beliefs

Contrary to our expectations, no effects of the intervention were found on children's epistemic beliefs (Conley et al., 2004). In the first effectiveness study, positive intervention effects were shown on the dimensions certainty, development, and justification of knowledge (see Schiefer, Golle, Tibus, et al., 2016). The following factors might explain why these results failed to replicate. First, it has to be noted that in the first study, the scientists who taught the course were experts in the field of epistemic beliefs and were highly familiar with all course elements. Conversely, the HCAP course instructors underwent only a 1-day training, and some of them had no prior knowledge of

epistemic beliefs. Therefore, it might have been difficult for them to internalize all the course elements that were intended to foster sophisticated epistemic beliefs. Moreover, the stimulation of participants' epistemic beliefs offers a great challenge as it should be done by guided reflections of relevant aspects of the epistemology of science or critical scrutiny (Akerson & Hanuscin, 2007). Such elements presumably require the course instructors to embody a sophisticated personal epistemology, and it is more difficult to implement this than it is to conduct specific experiments. Second, there is evidence that teachers' epistemic beliefs play an essential role in promoting students' sophisticated beliefs and influence teachers' teaching behavior (Akerson & Hanuscin, 2007; Buehl & Fives, 2016; Lederman, 1992; Lederman & Zeidler, 1987). Our descriptive analyses revealed that the course instructors' epistemic beliefs differed individually (the standard deviations of the four dimensions by Conley et al., 2004, were spread around the average value in a range of $.32 < SD < .54$). This may have led to different teaching styles in the respective courses. Furthermore, we assumed that even when course instructors embody sophisticated epistemic beliefs, this does not automatically result in teaching behavior that fosters children's epistemic beliefs.

Need for cognition and epistemic curiosity

As expected, we found positive intervention effects on the development of children's need for cognition (individual differences in the tendency to engage in and enjoy effortful cognitive activity; Cacioppo & Petty, 1982). This is in line with evidence that rather stable personality traits can be affected by interventions (e.g., Caspi & Roberts, 2001; Magidson et al., 2014). However, it is important to note that, given the pre-posttest design of our study, we do not know how long-lasting such effects are. By contrast, we were not able to replicate the effects on epistemic curiosity (Schiefer, Golle, Tibus, et al., 2016) in the present study. Overall, the effects on need for cognition and epistemic beliefs were not quite consistent. As these constructs are closely related to each other (Mussel, 2010), the differences in results that occurred between the constructs as well as between the first and the present study are difficult to interpret and require further research.

Implications

Our results have important implications for educational research and practice. First, the science intervention had positive effects on elementary school children's

inquiry-based methodological competencies and need for cognition. It can be concluded that the intervention was successful overall and positively affected the central elements of an adequate understanding of science as the SIC, experimentation strategies, as well as children's engagement and enjoyment of thinking. Therefore, the main objectives of the intervention were fulfilled. The intervention fostered fundamental aspects of children's understanding of science not only under standardized conditions but also when situated in the real world.

Second, to the best of our knowledge, this is the first study to show that it is possible to successfully foster an understanding of the complete SIC in elementary school children who were nominated to participate in an extracurricular enrichment program. This has educational implications for science learning. Our findings strengthen the recommendation that educators should incorporate scientific inquiry methods into science curricula from Grade 3 on (see education plans; Mullis & Martin, 2015; National Research Council, 1996) at least for the subgroup of talented children with special educational needs (i.e., in the context of enrichment). Under the guidance of a teacher, such children might be able to independently plan, conduct, and interpret experiments (see Colburn, 2000).

Limitations and Future Research

Although our study demonstrated beneficial effects of the science intervention, some limitations should be considered when interpreting the results. First, we used a particular sample of 117 third and fourth graders who were nominated to participate in an enrichment program. Besides high cognitive abilities, the children's interest and motivation were considered in these nominations. As was already shown in the first study (Schiefer, Golle, Tibus, et al., 2016), we found significant intervention effects for this group of students. However, it cannot be presumed that the intervention would have similar effects in other groups of children (e.g., children in their regular school classes). Future research might focus on replicating this study with different samples or on incorporating the selected course elements into the school context (e.g., during the regular school day or school projects for high-ability learners).

Second, our results point to the possibility that our intervention could be implemented successfully by teachers and external course instructors from the HCAP. Due to the small sample size, it was not possible to determine whether the intervention

effects would differ according to implementer characteristics. Further studies with larger samples could explore how characteristics of the course instructors (e.g., expert knowledge in the natural sciences, epistemic beliefs and understanding of science, pedagogical experience) are related to their teaching styles and the learning outcomes of the children. Other methods for assessing implementation fidelity (e.g., class observations, participant self-report; see Humphrey et al., 2016) could be taken into account to gain insights into the “intervention black box” (Abry et al., 2015, p. 1). However, adherence is a suitable and accepted measure in an effectiveness study, and its assessment provides insights into the time-related and organizational practicability of the intervention.

Third, it was challenging to find appropriate instruments for children in Grades 3 and 4. We implemented newly developed or adapted instruments (SIC test, experimentation strategies) or instruments that were originally developed for older children (e.g., epistemic beliefs questionnaire; Conley et al., 2004). It might be fruitful to assess some aspects of the understanding of science (in particular, epistemic beliefs) in young children not only with single-choice questions but with more open, qualitative instruments (e.g., cognitive interviews, scenario-based instruments, or think-aloud protocols; see Mason, 2016, for an overview) to gain better insights into children’s beliefs and possible changes in these beliefs. However, in the context of intervention studies, instruments should be suitable for group-testing situations and economical for large sample sizes. Future research could focus on developing comprehensive new instruments that measure different aspects of young children’s understanding of science.

Finally, no conclusions can be drawn about the long-term effects of our intervention. Therefore, future studies should follow students’ development for a longer period of time and might take into account other outcomes such as their science achievements, school career, or later vocational choices (see Brandwein, 1995; Robertson, Smeets, Lubinski, & Benbow, 2010).

Conclusion

This study investigated the effectiveness of an extracurricular science intervention program when it was put into practice. Trained teachers and course instructors with a professional background in the natural sciences successfully implemented the intervention, even though—compared with the first effectiveness study—some treatment

effects could not be replicated. Future research might focus on the further development of the intervention and its scaling up at all local sites of the HCAP. Investigations of differential effects (e.g., which individual prerequisites of children and course instructors influence the success of the intervention) might contribute to the understanding of children's learning processes and outcomes.

References

- Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation, 36*(3), 320-338.
- Akerson, V. L., & Hanuscin, D. L. (2007). Teaching nature of science through inquiry: Results of a 3-year professional development program. *Journal of Research in Science Teaching, 44*(5), 653-680.
- Baudson, T. G., Strobel, A., & Preckel, F. (2012). *Validierung einer neuen Need for Cognition (NFC)-Skala für Grundschülerinnen und Grundschüler: Struktur, Messinvarianz und Zusammenhänge mit Intelligenz- und Leistungsmaßen [Validation of a new Need for Cognition scale for elementary school children: Structure, measurement invariance, and relations to intelligence and performance measures]*. Vortrag beim 48. Kongress der Deutschen Gesellschaft für Psychologie (23.–27. September 2012), Bielefeld. [Abstract veröffentlicht in: R. Riemann (Hrsg.), 48. Kongress der Deutschen Gesellschaft für Psychologie (S. 56). Lengerich: Pabst Science Publishers.]
- Bendixen, L. D. (2016). Teaching for epistemic change in elementary classrooms. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 281-299). New York: Routledge.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading MA: Addison-Wesley.
- Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education, 94*(4), 577-616.
- Brandwein, P. F. (1995). *Science talent in the young expressed within ecologies of achievement* (Report No. EC 305208). Storrs, CT: National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED402700).
- Buehl, M. M., & Fives, H. (2016). The role of epistemic cognition in teacher learning and praxis. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 247-264). New York: Routledge.

- Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & A. Ziegler (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 38-54). Cambridge: University Press.
- Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Westport: Heinemann.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40-49.
- Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, 12(2), 49-66.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 1098-1120.
- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, 46(3), 141-167.
- Colburn, A. (2000). An inquiry primer. *Science Scope*, 23(6), 42-44.
- Conley, A. M., Pintrich, P. R., Vekiri, I. & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology*, 29, 186-204.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45.
- Dogan, N., & Abd-El-Khalick, F. (2008). Turkish grade 10 students' and science teachers' conceptions of nature of science: A national study. *Journal of Research in Science Teaching*, 45(10), 1083-1112.
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University Press.
- Durlak, J. A. (1998). Why program implementation is important. *Journal of Prevention and Intervention in the community*, 17(2), 5-18.

- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3-4), 327–350.
- Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.
- Ehmer, M. (2008). *Förderung von kognitiven Fähigkeiten beim Experimentieren im Biologieunterricht der 6. Klasse: Eine Untersuchung zur Wirksamkeit von methodischem, epistemologischem und negativem Wissen [Fostering cognitive abilities while experimenting in biology lessons in Grade 6: Investigating the impact of methodological, epistemological, and negative knowledge]*. Kiel: Christian-Albrechts-Universität. http://eldiss.uni-kiel.de/macau/receive/dissertation_diss_00003034 (01.09.2014)
- Elby, A., Macrander, C., & Hammer, D. (2016). Epistemic cognition in science. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 113-127). New York: Routledge.
- Elder, A. (2002). Characterizing fifth grade students' epistemological beliefs in science. In B. K. Hofer & P. R. Pintrich (Eds.), *The psychology of beliefs about knowledge and knowing* (pp. 347-363). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Emanuel, E. J., Wendler, D., & Grady, C. (2000). What makes clinical research ethical? *Jama, 283*(20), 2701-2711.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138.
- European Commission (2007). *Science education now: A renewed pedagogy for the future of Europe. Community Research Expert Group: Vol. 22845*. Luxembourg: EUR-OP.
- Fixsen, D. L., Blase, K., Metz, A., & van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children, 79*(2), 213–230.
- Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to

- personality and intelligence. *Personality and Social Psychology Bulletin*, 36(1), 82-96.
- Frank, A. (2005). Naturwissenschaftliches Arbeiten mit der Blackbox [Scientific work with a blackbox]. *Unterricht Biologie 307/308 Beilage*.
- Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science: The Official Journal of the Society for Prevention Research*, 16(7), 893–926.
- Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). *The study of implementation in school-based preventive interventions: Theory, practice and research*. Washington, DC: Department of Health and Human Services.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice*, 15(4), 198-205.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88-140.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88-110.
- Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature*. Retrieved from Education Endowment Foundation website:
https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf
- Kienhues, D., Bromme, R., & Stahl, E. (2008). Changing epistemological beliefs: The unexpected impact of a short-term intervention. *British Journal of Educational Psychology*, 78(4), 545-565.

- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661-667.
- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Eds.), *Handbook of Childhood Cognitive Development* (pp. 371-393). Oxford: Blackwell.
- Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how). In W. Damon & R. M Lerner (Eds.), *Handbook of Child Psychology* (pp. 517-550), New York: Wiley.
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching, 29*(4), 331-359.
- Lederman, N. G. (2007). Nature of Science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 831-879). New York: Routledge.
- Lederman, N. G., & Zeidler, D. L. (1987). Science teachers' conceptions of the nature of science: Do they really influence teaching behavior? *Science Education, 71*(5), 721-734.
- Lendrum, A., & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education, 38*(5), 635-652.
- Lendrum, A., & Wigelsworth, M. (2013). The evaluation of school-based social and emotional learning interventions: Current issues and future directions. *Psychology of Education Review, 37*, 70-76.
- Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences, 44*(7), 1585-1595.
- Litman, J., Hutchins, T., & Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition and Emotion, 19*(4), 559-582.
- Litman, J. A., & Spielberg, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment, 80*, 75-86.
- Magidson, J. F., Roberts, B. W., Collado-Rodriguez, A., & Lejuez, C. W. (2014). Theory-driven intervention for changing personality: Expectancy value theory,

- behavioral activation, and conscientiousness. *Developmental Psychology*, 50(5), 1442-1450.
- Mason, L. (2016). Psychological perspectives on measuring epistemic cognition. . In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 375-392). New York: Routledge.
- Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, 29, 43-55.
- McComas, W. F. (1998). The Principal Elements of the Nature of Science. In W.F. McComas (Hrsg.), *The Nature of Science in Science Education. Rationales and Strategies* (53-70). Dordrecht: Kluwer.
- McDonald, S. K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher*, 35(3), 15-24.
- McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology*, 62(4), 670-678.
- Muis, K. R., Trevors, G., & Chevrier, M. (2016). Epistemic climate for epistemic change. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 331-359). New York: Routledge.
- Mullis, I. V.S. & Martin, M. O. (Eds.). (2015). *TIMSS 2015 Assessment Frameworks*. Boston: Timss & Pirls.
- Mussel, P. (2010). Epistemic curiosity and related constructs: Lacking evidence of discriminant validity. *Personality and Individual Differences*, 49(5), 506-510.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.
- Nair, K. U., & Ramnarayan, S. (2000). Individual differences in need for cognition and complex problem solving. *Journal of Research in Personality*, 34(3), 305-328.
- National Research Council (1996). *National science education standards*. Washington, DC: The National Academies Press.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33-84.


- OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265-279.
- Oschatz, K. (2011). *Intuition und fachliches Lernen. Zum Verhältnis von epistemischen Überzeugungen und Alltagsphantasien [Intuition and subject-related learning. The relation between epistemic beliefs and everyday life imagination]*. Wiesbaden: VS Verlag.
- Oschatz, K. & Schiefer, J. (in press). Über Wissenschaft nachdenken in der Grundschule: Ein Hector Core Course [Thinking about science in elementary school: A Hector Core Course]. In U. Trautwein & M. Hasselhorn (Eds.), *Begabungen und Talente (Reihe Tests & Trends, Band 15)*. Göttingen: Hogrefe.
- Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47-61.
- Peltier, J. W., & Schibrowsky, J. A. (1994). Need for Cognition, Advertisement Viewing Time and Memory for Advertising Stimuli. *Advances in Consumer Research*, 21(1), 244-250.
- Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning*, 5(1), 47-65.
- Robertson, K. F., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science*, 19(6), 346-351.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252.
- Ryder, J., & Leach, J. (2000). Interpreting experimental data: the views of upper secondary school and university science students. *International Journal of Science Education*, 22(10), 1069-1084.

- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020-1049.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.
- Schiefer, J., Golle, J., Tibus, M., Herbein, E., Hoehne, V., Trautwein, U., & Oschatz, K. (2016). Fostering epistemic beliefs, epistemic curiosity, and investigative interests in elementary school children – A randomized STEM intervention study. *Manuscript submitted for publication*.
- Schiefer, J., Golle, J., & Oschatz, K. (2016). Scientific reasoning in elementary school children – Assessment of the inquiry cycle. *Manuscript submitted for publication*.
- Schommer, M. (1994). Synthesizing epistemological belief research: Tentative understandings and provocative confusions. *Educational Psychology Review*, 6(4), 293-319.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82(3), 498.
- Schroeders, U., Schipolowski, S., Zettler, I., Golle, J., & Wilhelm, O. (2016). Do the smart get smarter? Development of fluid and crystallized intelligence in 3rd grade. *Intelligence*, 59, 84-95.
- Simon, H. A. (1989). The scientist as problem solver. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 375-398). Hillsdale, NJ: Lawrence Erlbaum.
- Spiel, C., Schober, B., & Strohmeier, D. (2016). Implementing intervention research into public policy—the “I3-Approach”. *Prevention Science*, 1–10.
- Torgerson, C. J., & Torgerson, D. J. (2013). *Randomised trials in education: An introductory handbook*. London: EEF.
- Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences: An introduction*. Houndmills: Palgrave Macmillan.

- Torgerson, C. J., & Torgerson, D. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Educational Studies*, 49(3), 316-328.
- Urhahne, D., & Hopf, M. (2004). Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien [Epistemological beliefs in science and their relations to motivation, self-concept, and learning strategies]. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 71-87.
- Valla, J. M., & Williams, W. M. (2012). Increasing achievement and higher-education representation of under-represented groups in science, technology, engineering, and mathematics fields: A review of current K-12 intervention programs. *Journal of Women and Minorities in Science and Engineering*, 18(1), 21-53.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3-118.
- White, B., Frederiksen, J., & Collins, A. (2009). The interplay of scientific inquiry and metacognition: More than a marriage of convenience. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 175-205). New York: Routledge.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172-223.

Appendix

Tom has a new pet, an aardvark. He doesn't know much about aardvarks and wants to find out more about them. How should Tom proceed best if he wants to work like a scientist?



These are his working steps. They are not in the correct order.

He is considering an attempt to find out if the aardvark can detect sweet smells better than his dog.

He gathers information and reads in an animal encyclopedia that aardvarks have a good nose for sweet smells. He wants to check this.

He hides a small piece of cake in the garden and lets the dog and the aardvark look for the cake 10 times alternately.

He has evidence that the aardvark has a keener sense of smell for sweet smells than his dog.

He calculates how long each animal needed to find the cake. The aardvark was faster than the dog.

He wants to find out more about aardvarks' good noses and gathers information about whether aardvarks have a good nose for musty smells, too.

He is interested in aardvarks and wants to find out more about them.

→ **Put the working steps in the right order. The first step is already predetermined.**
 → **Glue the cards only if you are sure about the order you want to put them in.**

Figure 2. Example item from the SIC test (sorting task Part 1). The items required the sorting of the single steps of the SIC via printed labels. The respective research issue was introduced to the children in a short paragraph. The six single inquiry steps were read aloud to the children in a random order by the test instructors.

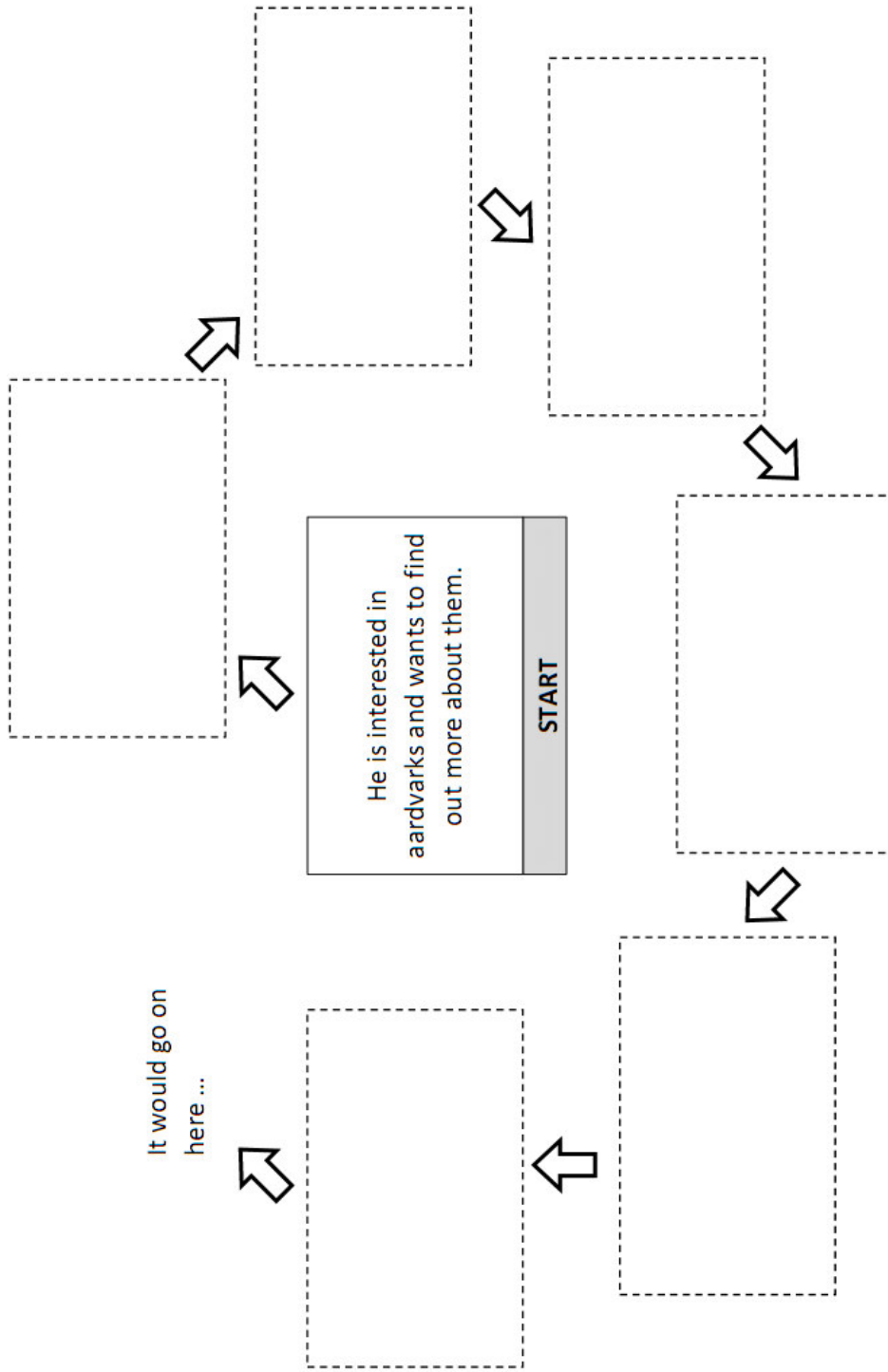


Figure 3. Example item from the SIC test (sorting task Part 2). Printed labels that contained the six inquiry steps were given to the children. They had to put the steps in the right order by sticking them in the questionnaire. The starting point—finding a research question—was given to the children. Only completely accurate solutions counted as correct because partial solutions did not indicate an understanding of the entire SIC.


<p>Mr. Abendstern is a famous scientist and he knows exactly how a scientist has to work.</p>	
<p>Now it's up to you. Please answer the following questions about Mr. Abendstern's working steps.</p> <p>The working steps for Mr. Abendstern are not in the correct order.</p> <p>Mark the <u>best</u> answer with a cross.</p>	
<p>Mr. Abendstern is interested in what causes tooth decay and wants to find out more about it. What is his next step?</p>	
<p><input type="checkbox"/> He plans an experiment to find out if chocolate causes tooth decay.</p>	
<p><input type="checkbox"/> He evaluates results. He observes whether people who only ate only apples for one week had worse teeth than those who ate only chocolate.</p>	
<p><input type="checkbox"/> He establishes a hypothesis: he believes that chocolate causes tooth decay.</p>	

Figure 4. Example item (single-choice item from the SIC test). Children were asked to select the respective next step in the SIC. The questions referred to each of the steps of the cycle and were offered in a random order. The response options referred either to the next step in the inquiry cycle (correct answer) or to two randomly selected other steps (wrong answers).

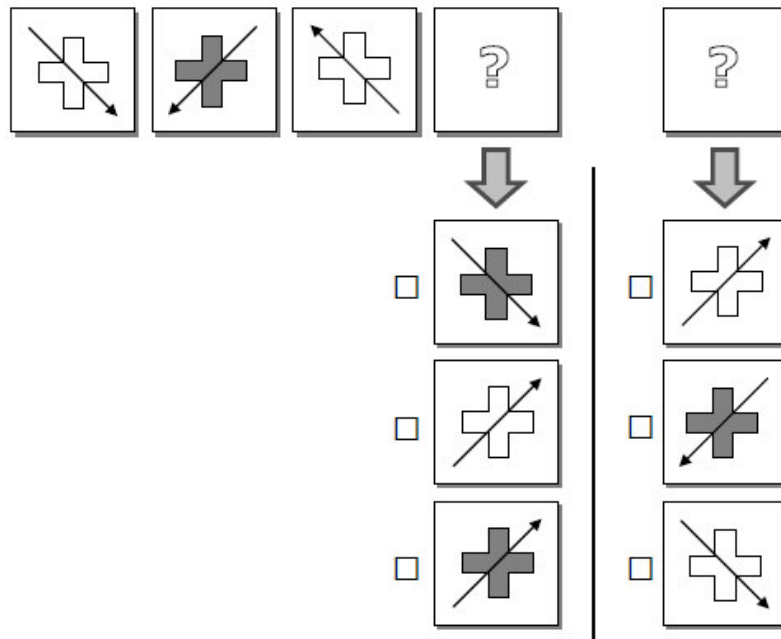
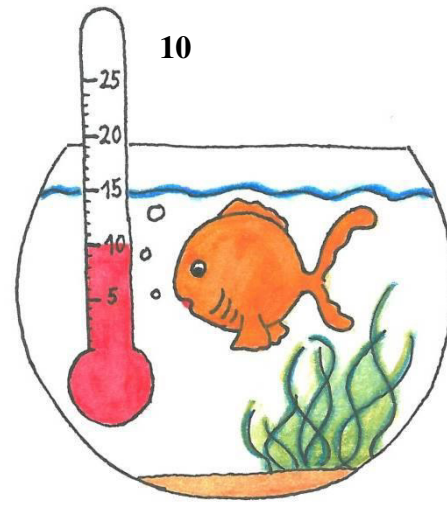
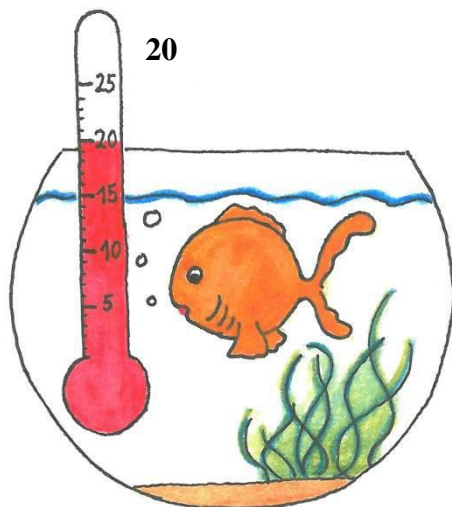


Figure 5. Example item from the BEFKI fluid intelligence test (Schroeders et al., 2016).

Eva is given two identical goldfish. She conducts an experiment about the respiration of the fish.

To do this, she uses two equally sized aquariums. In one of them, the water temperature is 20 degrees Celsius; in the other one, 10 degrees Celsius. She puts a goldfish and a plant in each aquarium.

She observes how often the fish breathe per minute. She recognizes this according to the movement of the gills of the fish.



**Why does Eva do this experiment?
Select the best answer.**

- (A) Because she wants to find out something about the respiration of the fish and believes that the plants bring oxygen into the aquarium.
- (B) Because she assumes that the frequency of the breathing of the fish depends on the water temperature.
- (C) Because she wants to know how often the fish move their gills per minute on average.

Figure 6. Example item for assessing experimentation strategies (authors' own item, modeled after Ehmer, 2008, and Mayer et al., 2014).

General Discussion

5 General Discussion

Promoting students' understanding of science is a normative goal of science education (EC, 2007; OECD, 2016) and a central element of scientific literacy (Jenkins, 1994). Promoting students' understanding of science as early as elementary school is supposed to support their natural curiosity, to lay an important foundation for their science learning and their later understanding of socioscientific issues (Jones, Wheeler, & Centurino, 2015; OECD, 2016). So far, the promotion of elementary school children's science competencies (i.e., in the context of interventions) has often focused on teaching scientific content knowledge (e.g., Andrés et al., 2010). There is a lack of interventions that have aimed to at fostering very fundamental aspects of the understanding of science such as general science methods and epistemic beliefs.

The present dissertation aimed to close this gap by addressing the central questions of how young children's understanding of science can be fostered effectively. To this end, an intervention was developed as part of an extracurricular enrichment program for gifted children. The effectiveness and practicability of the intervention was tested at different stages between its development and its broad dissemination into practice (Humphrey et al., 2016). Due to a lack of paper-and-pencil instruments for assessing elementary school children's understanding of science, a new instrument was previously developed and implemented within the scope of the effectiveness studies.

The discussion is structured in the following way: First, the findings of the three empirical studies are summarized and located within the broader research context. The discussion of these findings revolves around two major topics: (a) the measurement of elementary school children's understanding of science and (b) the effectiveness and implementation of the science intervention. Second, the strengths and limitations of the present dissertation are described. Based on the limitations, the consequential needs for future research are derived. In the final chapter, general implications for future research as well as educational policy and practice are discussed.

5.1. Discussion of General Findings

5.1.1. Measurement of the understanding of science

The starting point for the development of a new instrument was the identified lack of adequate paper-and-pencil tests for assessing elementary school children's understanding of science and the need of such an instrument for the evaluation of the newly developed intervention. The new instrument (SIC test) focused on the measurement of children's understanding of the whole process of scientific inquiry. This process can be located within inquiry-based methodological competencies. The SIC builds the basis for the genesis, construction, and development of knowledge in science and is therefore an important aspect of the understanding of science (Dogan & Abd-El-Khalik, 2008; Ryder & Leach, 2000; Zimmerman, 2007). In the process of the development of the SIC test, a special emphasis was placed on the assessment of quality criteria as well as quality standards (Downing, 2006).

The reliability of the SIC test was moderate but comparable to the reliabilities of similar tests developed for this age group (e.g., scientific thinking scale by Koerber et al., 2015, or Mayer et al., 2014). This indicates that the instrument can distinguish between different competence levels and can be used for assessing the understanding of SIC in 8 to 10 year old children. Furthermore, analyses of item difficulties revealed no ground and ceiling effects: The items were neither too easy nor too difficult for the children. Thus, the constructed items corresponded to the abilities of third and fourth graders.

Conclusions regarding the construct validity of the SIC test were derived from the analyses of the item structure as well as the investigation of relations to related constructs (see Moosbrugger & Kelava, 2008). The results of the confirmatory factor analyses pointed to the postulated one-dimensional factor structure of the instrument. Therefore, the ability to solve the SIC items can be explained by one underlying (latent) factor. This corresponds to our expectation that the understanding of the steps of the SIC can be considered as a closely interrelated and holistic process. In addition, the investigated relations to other constructs were in line with our theoretically derived expectations and provided further evidence for the construct validity. Cognitive abilities, in particular, (e.g., text comprehension, fluid intelligence) contributed to the understanding of the steps of the SIC and be considered as an important basis for the SIC tasks. However, cognitive

abilities were separable from the SIC, which points to the discriminant validity of the scale (Kline, 2015). It is particularly important at elementary school level—when children’s reading abilities are partly limited—to ensure that a specific competence as the SIC can be measured as a separate construct distinct from children’s reading skills. The convergent validity was indicated by the positive relations between SIC performance and experimentation strategies as well as sophisticated epistemic beliefs about the uncertainty of knowledge. On the one hand, this indicates a relation between the whole SIC and a specific strategy (CVS) within this inquiry process. Both require some kind of planning, a hypotheses-driven approach, and farsighted thinking in the research process. On the other hand, the positive relations between the SIC test and epistemic beliefs (in particular to beliefs about the uncertainty of knowledge) point to the relation of inquiry-based methodological competencies and sophisticated epistemic beliefs (beliefs about science as a changing and reversible discipline). Both are considered to be relevant components of an adequate understanding of science.

Taken together, the results showed that it was possible to assess elementary school children’s understanding of the SIC with a paper-and-pencil test. Specifically, the SIC test enabled to assess the understanding of the complete process of the steps of scientific inquiry. This aspect of the understanding of science has not yet been considered in existing tests. Therefore, the SIC fills a gap in existing tests and makes an important contribution to the assessment of elementary school children’s understanding of science.

5.1.2. Effectiveness and implementation of the intervention

Previous studies have demonstrated that certain aspects of the understanding of science can be promoted in school as well as in extracurricular contexts. However, hardly any attempts have been made to identify effective strategies to foster very fundamental aspects of the understanding of science (e.g., general science methods or epistemic beliefs), in particular in elementary school children below Grade level 5 (e.g., Bendixen, 2016; Conley et al., 2004; Elder, 2002; Muis et al., 2016). To fill this gap, an extracurricular 10-week intervention for third and fourth graders was developed and evaluated by two empirical studies within this dissertation.

Study 2 showed that the intervention was successful at enhancing children’s epistemic beliefs and epistemic curiosity. In this first effectiveness study, three scientists who developed the intervention and strictly adhered to the manual conducted the

intervention. Epistemic beliefs were enhanced in the dimensions certainty, development, and justification of knowledge (Conley et al., 2004). Hardly any previous studies had demonstrated positive intervention effects on the epistemic beliefs of children below Grade 5. The results support the idea that it is possible to consider such beliefs in young children for comprehensive science learning, although the specific sample (participants of an enrichment program) should be kept in mind. However, as part of this enrichment program, the intervention can be considered effective.

Study 3 broadened the research context, as further questions regarding the real-world implementation of the intervention were added. Thus, the intervention was now offered within the frame of the regular course program of the HCAP and administered by different course instructors who normally conduct similar courses at the respective local sites. Therefore, the sample was approximately twice as large as in the first study. The results revealed positive effects of the intervention on children's understanding of the SIC, experimentation strategies, and need for cognition. However, intervention effects on epistemic beliefs and epistemic curiosity could not be replicated. The failed replication of those effects raised further questions, namely if those results might be due to a limited implementation fidelity or might be explained by the characteristics of the course instructors. Analyses of implementation fidelity (self-reports of adherence) revealed that most of the course instructors kept to the program and were able to work with the provided materials. However, implementation varied between the different instructors, and because their teaching behavior was not observed or videotaped, there was no way to know for certain what they really *did* in class and *how* exactly they implemented the course elements. However, the second effectiveness study, which was conducted under the prevailing real-world conditions, was intended to maximize the standardization of data collection and implementation fidelity as the teachers underwent a mandatory 1-day training given by a course developer who was an expert on the scientific grounding and practical implementation of the program. Teachers were also provided with a written manual as well as the complete course materials.

A closer look to the characteristics of the course instructors revealed that they had different professional backgrounds and different levels of prior knowledge with regard to the construct of the understanding of science. This raised questions about the relevance of the understanding of science (e.g., epistemic beliefs) of the implementers. As teachers must demonstrate an adequate understanding of science in order to be able to foster

children's understanding of science (Akerson & Hanuscin, 2007; Lederman, 1992; Muis et al., 2016), it is unclear whether the failed replication was due to limitations to the sophistication of the course instructors or to possible difficulties in implementing the course elements that were intended to foster children's epistemic beliefs.

Furthermore, it must be noted that—for practical reasons—it was not possible to use identical instruments in Studies 2 and 3. This could have aided the comparison of effect sizes and the investigation of the potential loss of power between the two studies. In addition, in contrast to Study 2, the control group in Study 3 was not treated but was instead a waitlist control group, which enabled us to estimate the treatment effects in a more ecologically valid and simple manner. However, the effect sizes could not be compared directly as studies using a waitlist control group are supposed to produce stronger effects than studies using treated control groups when the topics are at least partly similar.

Taken together, the results revealed that it was possible to promote fundamental aspects of elementary school children's understanding of science by the extracurricular intervention and that the intervention could—with some limitations—be successfully put into practice.

5.1.3. Strengths and limitations of the present dissertation

Some general strengths and limitations of the present dissertation should be considered when interpreting its results.

First of all, one strength of this dissertation is that an effective intervention was developed and implemented into practice under real-world conditions. We delivered the intervention *from science to service* (Humphrey et al., 2016), as we followed the recommended steps in the process of the development, evaluation, and implementation of the intervention (Humphrey et al., 2016). These steps began with a sound theoretical conceptualization of an entire intervention program and its instructional design principles, followed by a first study under highly controlled conditions and a second effectiveness study in which the intervention was put into practice. Thus, this dissertation is an example for use-inspired basic research that directly links educational research and practice.

Second, in the whole process of developing and implementing the intervention, different research traditions (natural science education, psychology, education science) were combined fruitfully. Within this dissertation, the different research traditions mesh

with one another and go hand in hand to ensure high-quality research. This includes a theoretically grounded conceptualization of an intervention, psychometric expertise and advanced research methodology.

Another important strength of this dissertation was its use of strong research designs. In the effectiveness studies, we conducted randomized controlled field trials (RCFTs), which are considered the gold standard in educational research (Torgerson & Torgerson, 2013). RCFTs aim at evaluating educational interventions under realistic conditions. They provide the advantage that causal inferences can be drawn from conducting an experiment. It enables researchers to attribute changes in outcomes of interest to a specific intervention rather than to the many other possible causes of human behavior and performance (Towne & Hilton, 2004). In the context of field trials, this is especially challenging because in practice, it is not always easy to randomly assign participants to specific conditions (e.g., because children do not have time on certain days or do not want to participate in a particular course). However, it was possible to successfully meet this challenge in the present dissertation possible, for example by precise planning and providing detailed information about the necessity of RCFTs to all involved persons (e.g., parents, course instructors, directors of the HCAP). Nevertheless, such research is a very complex and time-consuming matter and therefore leads to rather small sample sizes.

A further strength of this dissertation was the use of state-of-the-art methods for data analyses. In Study 1, this included elaborate IRT modeling to scale the test, which enabled a precise estimation of student's understanding of science (see Embretson & Reise, 2013). In Studies 2 and 3, multiple regression analyses were used to estimate intervention effects while controlling for the baseline measures and certain covariates such as gender and intelligence. This increases the power of a study and enables an estimation of the average intervention effects independent from confounding variables. All analyses used the robust maximum likelihood estimator (MLR), which corrects the standard errors for the non-normality of the variables (Muthén & Muthén, 1998-2012). To account for the hierarchical clustering of the data (children nested in classes and HCAP courses), a design-based correction of the standard errors was applied, which is implemented in Mplus (Muthén & Muthén, 1998-2012). Missing data were accounted for by applying full information maximum likelihood (FIML) procedures (Schafer & Graham, 2002).

Although the results of this dissertation contribute significantly to questions about the measurement and promotion of elementary school children's understanding of science, some limitations should be kept in mind, which lead to subsequent directions for future research.

Regarding the SIC, the newly developed instrument measured the understanding of the SIC in a valid way, because the explored relations to cognitive abilities, experimentation strategies, and epistemic beliefs were in line with our expectations. However, only the most relevant validation instruments could be used in the present study due to time constraints within the school context. To get a broader picture of the validity of the SIC test, further investigation is needed, in particular regarding its criterion validity (e.g., in predicting practical experimentation competencies) and construct validity. Therefore, it might be promising to investigate whether the SIC test performance can predict students' practical experimentation skills (e.g., a targeted approach with respect to hands-on activities). Exploring relations between the SIC test and other constructs (e.g., problem solving, spatial abilities, see Klahr, 2000; Mayer et al., 2014), or the existing scientific reasoning test by Koerber et al., 2015 (which was not published yet when we conducted our study) can further determine construct validity and contribute to the theoretical embedding of the test.

Furthermore, the SIC test showed an acceptable but rather low reliability. The reliability of an instrument is essential in educational research as it is a prerequisite for precise measures of students' abilities. Thus, future research might want to improve the reliability of the scale, for example by constructing additional items. As a result, the SIC test could not solely be used for research purposes, but even for single case diagnostics (e.g., for the selection of participants for science enrichment programs).

In the effectiveness studies, we aimed at fostering central aspects of students' understanding of science (e.g., their understanding of the SIC). Due to the lack of instruments assessing student's understanding of the SIC, a new instrument was developed in the first study of this dissertation. Thus, the instrument that was, *inter alia*, used to evaluate the intervention, was developed within the same research group that developed the intervention. This might point to "teaching to the test" effects (Longo, 2010). However, none of the test items were used to teach the course. Nevertheless, the similarities between the test items and the intervention content may have contributed to an overestimation of the effect sizes.

Because the intervention was implemented as part of an enrichment program, a very specific sample was used in the studies (children who were nominated to participate in an enrichment program for gifted children). There were very good reasons for choosing this target group (e.g., the educational relevance of the promotion of talented children in the STEM domains), however, this limits the generalizability of the findings. Although the children in the HCAP did not appear to be gifted according to classical giftedness criteria (an IQ greater than two standard deviations above the mean of the sample; e.g., Terman, 1925), the results are still not directly transferable to a group of children with average IQs or to samples of younger or older children. Thus, there is a need for further research to explore if the intervention effects other children in similar ways.

Next, implementation fidelity is considered a very important factor in the context of intervention studies (Humphrey et al., 2016). A low implementation fidelity might be one possible reason for the failed replication of some effects in the second effectiveness study of this dissertation. We were only able to assess the adherence of the course instructors to the manual. However, this only provides a limited understanding of *what* the course instructors actually did and *how well* the components were implemented. Therefore, it might be important in future research to measure implementation fidelity with extended measures (e.g., quality of deliverance, participant responsiveness, including behavioral observations or video-taping methods, e.g. in the context of a multimedia lab; see O'Donnell, 2008; Humphrey et al., 2016). Moreover, including the fidelity measures in the statistical analyses (as mediators or moderators in regression analyses; see Carroll et al., 2007) can contribute to the understanding of the relevance of fidelity for children's learning outcomes.

Lastly, in this dissertation, questionnaires were used to assess intervention effects. Although paper-and-pencil tests are required in the context of group assessments and provide reliable and valid measures of the understanding of science at least to some extent, it might be fruitful to assess the understanding of science with additional methods as scenario-based interviews or think-aloud protocols (see Mason, 2016). This might allow a thorough insight into the intervention effects and a qualitative assessment of the development of children's understanding of science due to the intervention (see Mason, 2016).

5.2. Implications and Future Directions

5.2.1. Implications for future research

The results of this dissertation revealed that the developed instrument could measure the understanding of the SIC and that the intervention was—with minor restrictions—effective within the described samples. With the implementation under real-world conditions, a first step was made towards scaling up. The implications of this dissertation for future research aim to extend the findings of the present dissertation. In the following, the results are discussed with respect to the measurement and the promotion of elementary school children's understanding of science.

Implications for the Measurement of Children's Understanding of Science

Regarding the SIC, the results of the first study showed that the newly developed instrument successfully assessed children's SIC competencies. The instrument was designed to measure the understanding of the complete SIC. Nevertheless, it might be fruitful to get more insight into the dependencies of the single steps of the inquiry cycle as well as the underlying cognitive processes of the sorting tasks (e.g., see Figures 2 and 3 in Study 1). For scaling reasons, the answers in the SIC were scored dichotomously (correct, incorrect), although the active problem solving and sorting of six inquiry steps were required. The analyses of partial solutions and correct intermediate steps can provide more insight into children's understanding of the SIC and can be used to explore which steps of the SIC are more easy or more difficult for them when considering the process as a whole. This is an important prerequisite for the targeted promotion of inquiry-based learning. In addition, other methods could be used to analyze how children solve the sorting tasks. Tablet computers (e.g., iPads, see Young, 2014) could be used to administer these tasks, and then information about the duration of the sorting of the single steps as well as the targeted approach (e.g., How often do students correct their solutions and which steps do they adjust more often?) could easily be captured. By adding eye-tracking measurements (e.g., Duchowski, 2007) or think-aloud protocols (e.g., Nielsen, Clemmensen, & Yssing, 2002), further insight can be gained into the cognitive processing and strategies involved in the tasks.

In addition, the present dissertation provided evidence that the SIC test can be successfully applied to measure the understanding of the SIC in elementary school children of Grades 3 and 4. Future research should examine if the SIC could also be applied to different target groups, such as children of Grades 5 and 6, for instance. If the SIC would be applicable in broader age groups, (e.g., in children from Grades 3 to 6), children's development with regard to their competencies in solving the SIC tasks could be described using longitudinal research designs.

In sum, future research investigating the SIC test might want to focus on a deeper understanding of the underlying processes as well as on an extended application of the instrument. Expanding the perspective beyond the SIC test, future research could focus on the combination of different measurement approaches (quantitative and qualitative) to go beyond the borders of the respective conceptual frameworks described in Chapter 1.2.2. This might provide more insight into the interplay of the different aspects of the wide-ranging construct of the understanding of science.

Implications for the Promoting of Children's Understanding of Science

The present dissertation provides support that the newly developed intervention successfully fostered children's understanding of science (i.e., the understanding of the SIC and epistemic beliefs). However, future research is needed to extend these findings.

First, future research might explore the mechanisms through which science interventions (specifically, the promotion of the understanding of science in elementary school children) work. Although there have previously been detailed phases of the conceptualization and planning of the specific course elements with regard to children's outcomes, no inferences can be drawn about the importance of the individual elements and their possible interplay. Future studies might include intermediate surveys (to determine the duration up to the first intervention effect) or might subtract specific aspects of the intervention (e.g., phases of abstraction or the communication of results) to identify effective elements and effective instructional design principles. This might be realized by randomized controlled studies with parallel treatment groups (differing in the *intensity* of the treatment). Understanding the mechanisms of the intervention is an important prerequisite for its further development (e.g., by strengthening relevant aspects in the manual). The long-term goal should be to increase the demonstrated intervention effects (effect sizes).

Second, further research is needed to investigate if and how the intervention could be applied in different contexts (especially in different target groups and held by different instructors). In our study, we investigated main effects of the intervention within an extracurricular enrichment program on children's understanding of science. Future studies with larger samples (e.g., consecutive data from all 61 local sites of the HCAP) might investigate differential intervention effects, which provides a better insight *for whom* the intervention works and *by whom* it can be offered. Potential moderators that might be explored include participants' characteristics (e.g., intelligence, prior knowledge, or sophistication of epistemic beliefs of the children) as well as characteristics of the course instructors (e.g., prior knowledge, sophistication of epistemic beliefs, teaching quality, pedagogical experiences, or professional background). This will provide a more fine-grained insight into the intervention effects and the determination of the optimal target groups (e.g., Which children benefit the most from the intervention? Which characteristics of course instructors are required for optimal learning outcomes?). More knowledge about relevant characteristics of the course instructors (e.g., Which relevance has their understanding of science?) is necessary to develop a targeted training for course instructors (for example with regard to their understanding of science, see Abd-El-Khalick & Lederman, 2000; Brownlee, Schraw, Walker, & Ryan, 2016; Buehl & Fives, 2016). Future research could examine the effectiveness of such a training and detect effective methods for promoting course instructors' understanding of science and the impact of such a program on teaching quality and students' science learning (Brownlee et al., 2016).

Increasing knowledge about how the intervention works and for which students, future research might also investigate additional steps between effectiveness studies and the area-wide dissemination of interventions (Humphrey et al., 2016). The long-term goal should be scaling up the intervention to a wider audience without the loss of its effectiveness (Humphrey et al., 2016). Scaling up research can broaden the setting in which the intervention is conducted. It might be promising to transfer our results in everyday school life (e.g., into working groups during the regular school day or science lessons) to contribute to a widespread promotion for all students. As the effectiveness and practicability of the program has thus far been tested with small groups of children who participated in an enrichment program, adjustments with respect to the size as well as the characteristics of the target group will be required and will need evaluation.

Third, the results of the effectiveness studies provide evidence for short-term effects of our intervention. Future research might focus on investigating long-term effects. It might be promising to conduct follow-up studies and to follow children's development for a longer period of time. This might enable researchers to detect effects of an intervention on students' academic performance at the secondary school level or even across transitions to higher education or to professional life. By doing so, researchers could examine the attainment of the postulated goal of science interventions: to prevent a decline in students' interest in science subjects (Krapp, 1998; Pratt, 2007), to support their science learning (Leibham et al., 2013), and to lay a basis for later academic choices (Brandwein, 1995; Maltese & Tai, 2010; Metz, 2008). In this regard, it might also be promising to measure additional outcome variables in the context of future evaluations. It can be assumed that further aspects of the understanding of science (e.g., an understanding of the creative, social, and communicative aspects of science; see Ertl, 2010, 2013; McComas, 1998) might be promoted by the intervention.

Lastly, this dissertation shows that it is possible to foster and to measure children's understanding of the SIC, experimentation strategies, and epistemic beliefs, within a carefully developed and implemented intervention. Although those outcomes are central elements of the understanding of science, there might be further important aspects of the construct of the understanding of science that can be affected by interventions. Due to the great relevance of the promotion of students' understanding of science with effective methods, our intervention might serve as an example for the development of further interventions focusing on other aspects of the understanding of science (e.g., aspects revolving around the history of science or further science methods such as multivariate thinking, see Kuhn, Iordanou, Pease, & Wirkala, 2008; McComas, 1998).

In sum, future research on the promotion of elementary school children's understanding of science should take a longitudinal and fine-grained look at characteristics and shifts in students' as well as teachers' understanding of science and how they interact dynamically within different contexts (see Elby et al., 2016).

5.2.2. Implications for educational policy and practice

The present dissertation contributed to answering central questions regarding the measurement and promotion of elementary school children's understanding of science. Implications for future research were discussed. In this final section, implications of the current results for educational policy and practice are derived.

First, the results pointed to the effectiveness of a recently developed and evaluated science intervention as part of a statewide enrichment program for elementary school children in the German state of Baden-Württemberg. The findings indicate that it can be useful and beneficial to fund and support such programs. In this regard, the results of this dissertation also demonstrate the benefit of a close interrelation of educational research and practice. This is a prerequisite for the implementation of effective programs in routine practice. Meanwhile, the intervention is part of the regular HCAP program as one of the so-called *Hector Core Courses (HCCs)*; see Oschatz & Schiefer, in press). The HCCs are courses in the HCAP that were specifically developed to meet the needs of children with high cognitive abilities and have been tested with regard to their effectiveness and practicability. They build an important component of the quality assurance of this enrichment program and contribute to gifted education and to the promotion of our potential future STEM leaders (NSB, 2010). The materials that were developed in this dissertation (e.g., course manual) are in continuous practical use, and the further training that was developed for the course instructors will be part of another continuous education concept in this program. In the development and evaluation of the science intervention (which was one of the first HCCs), the applied procedure has the character of a pilot test and will be transferred to further HCCs in the STEM disciplines.

Second, the results of this dissertation demonstrate that it is possible to promote the understanding of science at elementary school level. This corresponds to educational policy, which emphasized the development of an early understanding of science (EC, 2007; NRC, 1996; Wendt et al., 2016). However, the latest results of the TIMSS revealed that only 7.6% of the fourth graders in Germany reached the highest competence level in science, and 21.6% did not even reach an intermediate benchmark (Wendt et al., 2016). We found evidence that children in Grades 3 and 4 who participated in an extracurricular enrichment program could—with some limitations—benefit from the targeted promotion of epistemic beliefs and inquiry-based learning approaches. This points towards an advanced understanding of science learning as stated in the education standards, namely

basic knowledge and skills related to scientific inquiry (Martin et al., 2015; Wendt et al., 2016). This indicates that extracurricular learning environments such as the courses at the HCAP complement the educational landscape significantly.

Third, our results provide evidence on how the understanding of science can be promoted in a feasible way at elementary school level. As stated above, the transferability to other samples is still unknown and the effectiveness of the single elements of the intervention and their possible interplay needs further research. However, overall the results point to the effectiveness of the selected methods and the instructional design principles (e.g., inquiry learning, scientific work according to the SIC, hands-on activities combined with reflections on epistemic issues, visit of a student lab) which might be adopted for science learning at school. The results of the assessment of children's understanding of science in regular school classes provide evidence that elementary school children were able to solve tasks with regard to the SIC and the design of controlled experiments, at least from Grade 3 onwards. This strengthens the utility of a comprehensive fostering of student's understanding of science. Under the guidance of a teacher, children might be able to plan, conduct, and interpret experiments independently, apply inquiry-based learning approaches, and think about how science works (see Colburn, 2000; Duschl, 2008).

Finally, the current results can be directly embedded into the discussion of the red-hot PISA results. In the foreword, the Secretary-General of the OECD, Angel Gurría, emphasized the ubiquitous importance of the understanding of science:

More important, science is not only the domain of scientists. In the context of massive information flows and rapid change, everyone now needs to be able to *think like a scientist*: to be able to weigh evidence and come to a conclusion; to understand that scientific *truth* may change over time, as new discoveries are made, and as humans develop a greater understanding of natural forces and of technology's capacities and limitations. (OECD, 2016, p. 2)

This quotation refers to the starting point of this dissertation. There is still a long way to go toward the ideal of all students and citizens embodying an adequate understanding of science. However, the results of this dissertation point in the right direction.

References

- Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: a critical review of the literature. *International Journal of Science Education*, 22(7), 665-701.
- Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, 36(3), 320–338.
- Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, 22(2), 227–257.
- Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121(2), 219-245.
- Adey, P., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. *Cognition and Instruction*, 11(1), 1-29.
- Aebli, H. (1980). Denken: das Ordnen des Tuns, Bd. 1: *Kognitive Aspekte der Handlungstheorie [Cognitive aspects of the action theory]*. Stuttgart: Klett-Cotta.
- Akerson, V.L., & Hanuscin, D.L. (2007). Teaching Nature of Science through Inquiry: Results of a 3-Year Professional Development Program. *Journal of Research in Science Teaching*, 44(5), 653-680.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- Andrés, S. H., Steffen, K., & Ben, J. (2010). *TALIS 2008 technical report*. Paris: OECD, Teaching and Learning International Survey.
- Banister, P. (2011). *Qualitative methods in psychology: A research guide*. McGraw-Hill Education (UK).
- Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, 333, 975-978.
- Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemporary Educational Psychology*, 42, 141-158.

- Bell, R. L., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction. *Science Teacher*, 72, 30–33.
- Benbow, C. P., Lubinski, D., & Sanjani, H. E. (1999). Our future leaders in science: Who are they? Can we identify them early? In N. Colangelo & S. G. Assouline (Eds.), *Talent development III*, (pp. 59-70). Scottsdale, AZ: Great Potential Press.
- Bendixen, L. D. (2016). Teaching for epistemic change in elementary classrooms. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 281-299). New York: Routledge.
- Bildungsplan Grundschule (2004). Ministerium für Kultur, Jugend und Sport Baden-Württemberg & Landesinstitut für Erziehung und Unterricht Stuttgart.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley.
- Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, 94(4), 577-616.
- Blase, K. A., van Dyke, M., Fixsen, D. L., & Bailey, F. W. (2012). Key concepts, themes, and evidence for practitioners in educational psychology. In B. Kelly & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (pp. 13–34). Cambridge, New York: Cambridge University Press.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Social Science & Medicine*, 75(12), 2299-2306.
- Brandwein, P. F. (1995). *Science talent in the young expressed within ecologies of achievement* (Report No. EC 305208). Storrs, CT: National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED402700).
- Brownlee, J. L., Schraw, G., Walker, S., & Ryan, M. (2016). Changes in preservice teachers' personal epistemologies. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 300-317). New York: Routledge.

- Buehl, M. M., & Alexander, P. A. (2005). Motivation and performance differences in students' domain-specific epistemological belief profiles. *American Educational Research Journal*, 42(4), 697-726.
- Buehl, M. M., Alexander, P. A., & Murphy, P. K. (2002). Beliefs about schooled knowledge: Domain specific or domain general? *Contemporary Educational Psychology*, 27(3), 415-449.
- Buehl, M. M., & Fives, H. (2016). The role of epistemic cognition in teacher learning and praxis. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 247-264). New York: Routledge.
- Bullock, M., Sodian, B., & Koerber, S. (2009). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood: Findings from the Munich longitudinal study* (pp. 173-198). Mahwah, NY: Lawrence Erlbaum.
- Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & A. Ziegler (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 38-54). Cambridge: University Press.
- Bundesministerium für Bildung und Forschung (Hrsg.) (2013). *Perspektive MINT. Wegweiser für MINT-Förderung und Karrieren in Mathematik, Informatik, Naturwissenschaften und Technik [Perspective MINT. Guide for promotion and careers in mathematics, informatics, natural sciences, and technology]*. Berlin: BMBF Bildung.
- Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Westport: Heinemann.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131.
- Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). An experiment is when you try it and see if it works. A study of junior high school students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11, 514-529.

- Carnevale, A. P., Smith, N., & Melton, M. (2011). *STEM: Science Technology Engineering Mathematics*. Washington DC: Georgetown University Center on Education and the Workforce.
- Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*(4), 1686-1706.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*(1), 40-49.
- Cavagnetto, A. R. (2010). Argument to Foster Scientific Literacy: A Review of Argument Interventions in K–12 Science Contexts. *Review of Educational Research, 80*(3), 336-371.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098-1120.
- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46*(3), 141-167.
- Chopyak, J., & Levesque, P. (2002). Public participation in science and technology decision making: Trends for the future. *Technology in Society, 24*(1), 155-166.
- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mountain View, CA: Mayfield Publishing Company.
- Colburn, A. (2000). An inquiry primer. *Science Scope, 23*(6), 42-44.
- Conley, A. M., Pintrich, P. R., Vekiri, I. & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology, 29*, 186-204.
- Cotabish, A., Dailey, D., Robinson, A., & Hughes, G. (2013). The effects of a STEM intervention on elementary students' science knowledge and skills. *School Science and Mathematics, 113*(5), 215-226.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23-45.

- Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, 8, 1-44.
- De Jong, T. (2006). Technological advances in inquiry learning. *Science*, 312, 532 f.
- Deng, F., Chen, D. T., Tsai, C. C., & Chai, C. S. (2011). Students' views of the nature of science: A critical review of research. *Science Education*, 95(6), 961-999.
- DiSessa, A. A., Elby, A., & Hammer, D. (2003). J's epistemological stance and strategies. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 237-290). Mahwah, NJ: Lawrence Erlbaum Associates.
- Dogan, N., & Abd-El-Khalick, F. (2008). Turkish grade 10 students' and science teachers' conceptions of nature of science: A national study. *Journal of Research in Science Teaching*, 45(10), 1083-1112.
- Dorsen, J., Carlson, B., & Goodyear, L. (2006). *Connecting informal STEM experiences to career choices: Identifying the pathway*. Waltham: ITEST Learning Resource Center.
- Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna. *Handbook of test development* (pp. 3-25). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University Press.
- Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). London: Springer Science & Business Media.
- Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. *Complex information processing: The impact of Herbert A. Simon*, 109-143.
- Durant, J. R. (1993). What is scientific literacy? In J. R. Durant & J. Gregory (Eds.), *Science and culture in Europe* (pp. 129-137). London: Science Museum.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327–350.

- Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 32, 268–291.
- Duschl, R. (1990). *Restructuring science education: the importance of theories and their development*. New York: Teachers College Press.
- Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39-72.
- Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.
- Elby, A., Macrander, C., & Hammer, D. (2016). Epistemic cognition in science. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 113-127). New York: Routledge.
- Elder, A. (2002). Characterizing fifth grade students' epistemological beliefs in science. In B. K. Hofer & P. R. Pintrich (Eds.), *The psychology of beliefs about knowledge and knowing* (pp. 347-363). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. New York: Psychology Press.
- Erdosne Toth, E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, 18(4), 423-459.
- Ertl, D. (2013). Sechs Kernaspekte zur Natur der Naturwissenschaft [Six key aspects of the nature of science]. *Plus Lucis* 1-2(13)
- Ertl, D. (2010). The Nature of Science. *Plus Lucis* 1-2(10).
- European Commission (2007). *Science education now: A renewed pedagogy for the future of Europe*. *Community Research Expert Group*: Vol. 22845. Luxembourg: EUR-OP.
- Fixsen, D. L., Blase, K., Metz, A., & van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children*, 79(2), 213–230.
- Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to

- personality and intelligence. *Personality and Social Psychology Bulletin*, 36(1), 82-96.
- Flick, L. B. (1993). The meanings of hands-on science. *Journal of Science Teacher Education*, 4(1), 1-8.
- Frank, A. (2005). Naturwissenschaftliches Arbeiten mit der Blackbox [Scientific work with a blackbox]. *Unterricht Biologie 307/308 Beilage*.
- Fraser, M. W., & Galinsky, M. J. (2010). Steps in intervention research: Designing and developing social programs. *Research on Social Work Practice*, 20, 459-466.
- Golle, J., Herbein, E., Hasselhorn, M. & Trautwein, U. (in press). Begabungs- und Talentförderung in der Grundschule durch Enrichment: Das Beispiel der Hector-Kinderakademien [Promotion of giftedness and talent at elementary school using enrichment: The Hector Children's Academy Program]. In U. Trautwein & M. Hasselhorn (Hrsg.), *Begabungen und Talente* (Reihe Tests & Trends, Band 15). Göttingen: Hogrefe.
- Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science: The Official Journal of the Society for Prevention Research*, 16(7), 893–926.
- Grandy, R., & Duschl, R. A. (2007). Reconsidering the character and role of inquiry in school science: Analysis of a conference. *Science and Education*, 16(2), 141-166.
- Greenberg, M. T. (2010). School-based prevention: Current status and future challenges. *Effective Education*, 2, 27-52.
- Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). *The study of implementation in school-based preventive interventions: Theory, practice and research*. Washington, DC: Department of Health and Human Services.
- Greene, J. A., Sandoval, W. A., & Bråten, I. (Eds.). (2016). *Handbook of epistemic cognition*. New York: Routledge.
- Greene, J. A., & Yu, S. B. (2014). Modeling and measuring epistemic cognition: A qualitative re-investigation. *Contemporary Educational Psychology*, 39(1), 12-28.

- Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 169-190). Mahwah, NJ: Laurence Erlbaum Associates.
- Heller, K. A., Mönks, F. J., Subotnik, R., & Sternberg, R. J. (Eds.). (2000). *International handbook of giftedness and talent*. Oxford/Amsterdam: Pergamon Press/ Elsevier Science.
- Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology*, 25(4), 378-405.
- Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist*, 39(1), 43-55.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88-140.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources.
- Höttecke, D. (2010). Forschend-entdeckender Physikunterricht. Ein Überblick zu Hintergründen, Chancen und Umsetzungsmöglichkeiten entsprechender Unterrichtskonzeptionen [Inquiry-based physics lesson. An overview about the background, opportunities and possibilities of implementation of teaching concepts]. *Unterricht Physik*, 19, 4-12.
- Höttecke, D. (2001). Die Vorstellungen von Schülern und Schülerinnen von der „Natur der Naturwissenschaften“ [Students' conceptions about the „nature of science“]. *Zeitschrift für Didaktik der Naturwissenschaften*, 7(1), 7-23.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88–110.
- Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326, 1410-1412.
- Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature*. Retrieved from Education Endowment Foundation website:

https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf

- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). New York: Psychology Press.
- Jenkins, E. W. (1994). Scientific literacy. *The international encyclopedia of education*, 9, 5345-5350.
- Jones, L. R., Wheeler, G., & Centurino, V. A. (2015). TIMSS 2015 science framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 Assessment Frameworks* (pp. 29-60). Boston: Timss & Pirls.
- Kahn, J. H., & Scott, N. A. (1997). Predictors of research productivity and science-related career goals among counseling psychology doctoral students. *The Counseling Psychologist*, 25(1), 38-67.
- Kardash, C. M., & Howell, K. L. (2000). Effects of epistemological beliefs and topic-specific beliefs on undergraduates' cognitive and strategic processing of dual-positional text. *Journal of Educational Psychology*, 92(3), 524-535.
- Keeley, P. (2009). Elementary science education in the K-12 system. *Science and Children*, 46(9), 8-9.
- Kerr, B., & Robinson Kurpius, S. E. (2004). Encouraging talented girls in math and science: Effects of a guidance intervention. *High Ability Studies*, 15(1), 85-102.
- King, P. M., & Kitchener, K. S. (1994). *Developing Reflective Judgment*. San Francisco: Jossey-Bass.
- Kitchener, R. F. (2011). Personal epistemology and philosophical epistemology: The view of a philosopher. In J. Elen et al. (Eds.), *Links between beliefs and cognitive flexibility* (pp. 79-103). Dordrecht: Springer.
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1-48.
- Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, 44(1), 183-203.

- Kline, P. (2015). *A handbook of test construction (Psychology revivals): Introduction to psychometric design*. London: Routledge.
- Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development, 86*(1), 327-336.
- Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology, 64*(3), 141-152.
- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht [Development and promotion of interests in school lessons]. *Psychologie in Erziehung und Unterricht, 44*(3), 185-201.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*(5), 810-824.
- Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.
- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 371-393). Oxford: Blackwell.
- Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178-181.
- Kuhn, D., Amsel, E., O'Loughlin, M., Schauble, L., Leadbeater, B., & Yotive, W. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.
- Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science, 16*(11), 866-870.
- Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how). In W. Damon & R. M Lerner (Eds.), *Handbook of child psychology* (pp. 517-550), New York: Wiley.
- Kuhn, D., Garcia-Mila, M., Zohar, A., Andersen, C., White, S. H., Klahr, D., & Carver, S. M. (1995). Strategies of knowledge acquisition. *Monographs of the society for research in child development, 60*(4), 1-128.
- Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development, 23*(4), 435-451.

- Kuhn, D., & Park, S. H. (2005). Epistemological understanding and the development of intellectual values. *International Journal of Educational Research*, 43(3), 111-124.
- Kuhn, D., Pennington, N., & Leadbeater, B. (1983). Adult thinking in developmental perspective. *Life-span Development and Behavior*, 5, 158-195.
- Kuhn, D., & Weinstock, M. (2002). *What is epistemological thinking and why does it matter?* Lawrence Erlbaum Associates Publishers.
- Kultusministerkonferenz (2009). Empfehlung der Kultusministerkonferenz zur Stärkung der mathematisch-naturwissenschaftlich-technischen Bildung.
[Recommendation of the ministerial conference to strengthen the mathematical scientific and technical education] Beschluss der Kultusministerkonferenz vom 07.05.2009.
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2009/2009_05_07-Empf-MINT.pdf (03.11.2016).
- Kunter, M., & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. Professionelle Kompetenz von Lehrkräften [The model of teaching quality COACTIV: A multi-criterial analysis. Professional competence of teachers]. *Ergebnisse des Forschungsprogramms COACTIV*, 85-113.
- Lapan, R. T., Shaughnessy, P., & Boggs, K. (1996). Efficacy expectations and vocational interests as mediators between sex and choice of math/science college majors: A longitudinal study. *Journal of Vocational Behavior*, 49(3), 277-291.
- Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, 84(1), 71-94.
- Lawson, A. E. (2005). What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching*, 42(6), 716.
- Lazonder, A. W., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction*, 22(6), 458-464.
- Lederman, N. G. (2007). Nature of Science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 831-879). New York: Routledge.

- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29(4), 331-359.
- Lederman, N. G., Wade, P., & Bell, R. L. (1998). Assessing understanding of the nature of science: A historical perspective. In W. F. McComas. *The nature of science in science education* (pp. 331-350). Springer Netherlands.
- Lederman, N. G., & Zeidler, D. L. (1987). Science teachers' conceptions of the nature of science: Do they really influence teaching behavior? *Science Education*, 71(5), 721-734.
- Leibham, M. B., Alexander, J. M., & Johnson, K. E. (2013). Science Interests in Preschool Boys and Girls: Relations to Later Self-Concept and Science Achievement. *Science Education*, 97(4), 574-593.
- Lendrum, A., & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, 38(5), 635-652.
- Lendrum, A., & Wigelsworth, M. (2013). The evaluation of school-based social and emotional learning interventions: Current issues and future directions. *Psychology of Education Review*, 37, 70-76.
- Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, 44(7), 1585-1595.
- Litman, J., Hutchins, T., & Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition and Emotion*, 19(4), 559-582.
- Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment*, 80, 75-86.
- Longo, C. (2010). Fostering creativity or teaching to the test? Implications of state testing on the delivery of science instruction. *The Clearing House*, 83(2), 54-57.
- Maltese, A. V., & Tai, R. H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education*, 32(5), 669-685.
- Mansfield, A. F., & Clinchy, B. M. (2002). Toward the integration of objectivity and subjectivity: Epistemological development from 10 to 16. *New Ideas in Psychology*, 20(2), 225-262.
- Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (Eds.). (2015). *TIMSS 2015 International Results in Science*. Boston: Timss & Pirls.

- Mason, L. (2016). Psychological perspectives on measuring epistemic cognition. . In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 375-392). New York: Routledge.
- Mason, L. (2003). Personal epistemologies and intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 199-236). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction, 21*(1), 137-151.
- Mason, L., & Bromme, R. (2010). Situating and relating epistemological beliefs into metacognition: studies on beliefs about knowledge and knowing. *Metacognition and Learning, 5*(1), 1-6.
- Matthews, M. R. (1994). *Science teaching: The role of history and philosophy of science*. New York: Routledge.
- Mayer, D. (2011). *Die Modellierung des wissenschaftlichen Denkens im Grundschulalter [Modeling scientific thinking in elementary school age]*. Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophie an der Ludwig-Maximilians-Universität München.
- Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction, 29*, 43-55.
- McComas, W. F. (1998). The Principal Elements of the Nature of Science. In W.F. McComas (Ed.), *The Nature of Science in Science Education. Rationales and Strategies* (pp. 53-70). Dordrecht: Kluwer.
- McComas, W. F., Almazroa, H., & Clough, M. P. (1998). The nature of science in science education: An introduction. *Science and Education, 7*(6), 511-532.
- Metz, K. E. (2011). Disentangling robust developmental constraints from the instructionally mutable: Young children's epistemic reasoning about a study of their own design. *The Journal of the Learning Sciences, 20*(1), 50-110.
- Metz, K. E. (2008). Narrowing the gulf between the practices of science and the elementary school science classroom. *The Elementary School Journal, 109*(2), 138-161.

- Meyer, K., & Carlisle, R. (1996). Children as experimenters. *International Journal of Science Education*, 18(2), 231-248.
- Minner, D. D., Levy, A. J., & Century, J. (2009). Inquiry-based science instruction - what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47, 474-496.
- Montgomery, D. E. (1992). Young children's theory of knowing: The development of a folk epistemology. *Developmental Review*, 12(4), 410-430.
- Moosbrugger, H., & Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). [Quality requirements for a psychological test (test quality criteria)] In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 7-26). Berlin Heidelberg: Springer.
- Morris, B. J., Croker, S., Masnick, A. M., & Zimmerman, C. (2012). The emergence of scientific reasoning. In Kloos, H., Morris, B. J., & Amaral, J. L. (Eds), *Trends in Cognitive Development*, (pp. 61-82). Rijeka, Croatia: InTech.Croatia, InTech Publishing.
- Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-general and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, 18(1), 3-54.
- Muis, K. R., Trevors, G., & Chevrier, M. (2016). Epistemic climate for epistemic change. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 331-359). New York: Routledge.
- Mullis, I. V.S. & Martin, M. O. (Eds.). (2015). *TIMSS 2015 Assessment Frameworks*. Boston: Timss & Pirls.
- Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus user's guide*. Seventh edition. Los Angeles, CA: Muthén & Muthén.
- Nair, K. U., & Ramnarayan, S. (2000). Individual differences in need for cognition and complex problem solving. *Journal of Research in Personality*, 34(3), 305-328.
- National Research Council (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. Washington, DC: The National Academies Press.

-
- National Research Council (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academy Press.
- National Research Council (Ed.). (1996). *National science education standards*. Washington, D.C.: National Academy Press.
- National Science Board (2010). *Preparing the next generation of STEM innovators: Identifying and developing our nation's human capital* (NSB-10-33). Retrieved from <http://www.nsf.gov/nsb/publications/2010/nsb1033> (22.03.2016)
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, 39(4), 374–396.
- Nielsen, J., Clemmensen, T., & Yssing, C. (2002). Getting access to what goes on in people's heads?: Reflections on the think-aloud technique. In *Proceedings of the second Nordic conference on Human-computer interaction* (pp. 101-110). ACM.
- Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education*, 30(15), 1977-1999.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33-84.
- OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.
- OECD. (2011). *Education at a glance, 2011: OECD indicators*. Paris: OECD.
- Oschatz, K. (2011). *Intuition und fachliches Lernen. Zum Verhältnis von epistemischen Überzeugungen und Alltagsphantasien*. Wiesbaden: VS Verlag.
- Oschatz, K. & Schiefer, J. (in press). Über Wissenschaft nachdenken in der Grundschule: Ein Hector Core Course [Thinking about science in elementary school: A Hector Core Course]. In U. Trautwein & M. Hasselhorn (Eds.), *Begabungen und Talente* (Reihe Tests & Trends, Band 15). Göttingen: Hogrefe.
- Oschatz, K., Schiefer, J., & Trautwein, U. (2015). *Hector Kids II. „Kleine Forscher – Wir arbeiten wie Wissenschaftler“*. Ein Trainingsprogramm zur Förderung des Wissenschaftsverständnisses von besonders begabten Kindern im Grundschulalter. [„Little Researchers – We work like Scientists“]. A training

- program for the promotion of the understanding of science in gifted elementary school children*]. Unpublished course manual, University of Tübingen.
- Patrick, H., Mantzicopoulos, P., & Samarapungavan, A. (2009). Motivation for learning science in kindergarten: Is there a gender gap and does integrated inquiry and literacy instruction make a difference. *Journal of Research in Science Teaching*, 46(2), 166-191.
- Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47-61.
- Peltier, J. W., & Schibrowsky, J. A. (1994). Need for Cognition, Advertisement Viewing Time and Memory for Advertising Stimuli. *Advances in Consumer Research*, 21(1), 244-250.
- Piaget, J. (1966). *Psychologie der Intelligenz [Psychology of intelligence]*. Freiburg i.Br.: Olten.
- Pintrich, P. R. (2002). Future challenges and directions for theory and research on personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *The psychology of beliefs about knowledge and knowing* (pp. 389-414). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Popper, K. R. (1935). *Logik der Forschung: zur Erkenntnistheorie der moderner Naturwissenschaft [Logic of research: Epistemology of modern science]*. Heidelberg: Verlag Julius Springer.
- Pratt, H. (2007). Science education's 'overlooked ingredient': Why the path to global competitiveness begins in elementary school. *Education Week*, 27(7), 26-32.
- Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning*, 5(1), 47-65.
- Rinn, A. N. (2006). Effects of a summer program on the social self-concepts of gifted adolescents. *Prufrock Journal*, 17(2), 65-75.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252.
- Ryder, J., & Leach, J. (2000). Interpreting experimental data: the views of upper secondary school and university science students. *International Journal of Science Education*, 22(10), 1069-1084.

- Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, 96(3), 488-526.
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020-1049.
- Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, 89(4), 634-656.
- Sawyer, R. K. (2008). Optimising learning: Implications of learning sciences research. In OECD (Ed.), *Innovating to learn, learning to innovate* (pp. 45-66). Paris: OECD/CERI.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32(1), 102-119.
- Schiefer, J., Golle, J., Tibus, M., Herbein, E., Hoehne, V., Trautwein, U., & Oschatz, K. (2016). Fostering epistemic beliefs, epistemic curiosity, and investigative interests in elementary school children – A randomized STEM intervention study. *Manuscript submitted for publication*.
- Schommer, M. (1994). Synthesizing epistemological belief research: Tentative understandings and provocative confusions. *Educational Psychology Review*, 6(4), 293-319.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82(3), 498-504.
- Schommer-Aikins, M. (2002). An evolving theoretical framework for an epistemological belief system. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 105-118). Mahwah, NJ: Laurence Erlbaum Associates.
- Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, 9(6), 523-538.

- Shipman, H. L. (2004). Inquiry learning in college classrooms. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science* (pp. 357-387). Dordrecht: The Netherlands: Kluwer.
- Simon, H. A. (1989). The scientist as problem solver. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 375-398). Hillsdale, NJ: Lawrence Erlbaum.
- Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching*, 40(5), 510-528.
- Smith, C., Maclin, D., Houghton, C. & Hennessey, M. G. (2000). Sixth grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction* 18(3), 349-422.
- Sodian, B., Jonen, A., Thoermer, C., & Kircher, E. (2006). Die Natur der Naturwissenschaften verstehen: Implementierung wissenschaftstheoretischen Unterrichts in der Grundschule. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule* (pp. 147-160). Münster: Waxmann Verlag.
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62(4), 753-766.
- Spiel, C., Schober, B., & Strohmeier, D. (2016). Implementing intervention research into public policy—the “I3-Approach”. *Prevention Science*, 1–10.
- Stake, J. E., & Mares, K. R. (2005). Evaluating the impact of science-enrichment programs on adolescents' science motivation and confidence: The splashdown effect. *Journal of Research in Science Teaching*, 42(4), 359-375.
- Stake, J. E., & Mares, K. R. (2001). Science enrichment programs for gifted high school girls and boys: Predictors of program impact on science confidence and motivation. *Journal of Research in Science Teaching*, 38(10), 1065-1088.
- Stapf, A. (2003). *Hochbegabte Kinder: Persönlichkeit, Entwicklung, Förderung*. [Gifted children: personality, development, promotion.] München: Beck.
- Sternberg, R. J. (2005). *Conceptions of giftedness*. New York: Cambridge University Press.

- Subotnik, R. F., Olszewski-Kubilius, P. & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, 12(1), 3-54.
- Terman, L. M. (1925). *Mental and physical traits of a thousand gifted children*. Genetic studies of genius, Vol. 1. Stanford, CA: Stanford University Press.
- Torgerson, C. J., & Torgerson, D. J. (2013). *Randomised trials in education: An introductory handbook*. London: EEF.
- Towne, L., & Hilton, M. (Eds.). (2004). *Implementing randomized field trials in education: Report of a workshop*. Committee on Research in Education, National Research Council. National Academy of Science. Washington, D.C.: The National Academies Press.
- Trautwein, U., & Lüdtke, O. (2008). Die Erfassung wissenschaftsbezogener Überzeugungen in der gymnasialen Oberstufe und im Studium: Validierung des Fragebogens zur Erfassung des Entwicklungsniveaus epistemologischer Überzeugungen (FREE) [Assessing science-related beliefs at upper secondary and college level: A validation study of the FREE questionnaire]. *Zeitschrift für Pädagogische Psychologie*, 22(34), 277-291.
- Trautwein, U., & Lüdtke, O. (2007). Epistemological beliefs, school achievement, and college major: A large-scale longitudinal study on the impact of certainty beliefs. *Contemporary Educational Psychology*, 32(3), 348-366.
- Tsui, L. (2007). Effective strategies to increase diversity in STEM fields: A review of the research literature. *The Journal of Negro Education*, 76(4), 555-581.
- Valla, J. M., & Williams, W. M. (2012). Increasing achievement and higher-education representation of under-represented groups in science, technology, engineering, and mathematics fields: A review of current K-12 intervention programs. *Journal of Women and Minorities in Science and Engineering*, 18(1), 21-53.
- Veenstra, C. P., Padró, F. F., & Furst-Bowe, J. A. (Eds.). (2012). *Advancing the STEM Agenda: Quality improvement supports STEM: Selected papers from the 2011 Advancing the STEM Agenda in Education, the Workplace and Society Conference at the University of Wisconsin-Stout, July, 2011*. Milwaukee: ASQ Quality Press.
- Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to

- STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology*, 102(4), 860-871.
- Waldmann, M. R., & Weinert, F. E. (1990). *Intelligenz und Denken: Perspektiven der Hochbegabungsforschung [Intelligence and thinking: Perspectives on research on giftedness]*. Göttingen: Hogrefe.
- Wellman, H. M. (1990). *The child's theory of mind* (Vol. 37). Cambridge, MA: MIT press.
- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K., Kasper, D. (Eds.). (2016). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [International comparison of elementary school children's mathematical and science competencies]*. Münster: Waxmann.
- White, B., & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, 40(4), 211-223.
- White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3-118.
- White, B., Frederiksen, J., & Collins, A. (2009). The interplay of scientific inquiry and metacognition: More than a marriage of convenience. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 175–205). New York, NY: Routledge.
- Wilhelm, P., & Beishuizen, J. J. (2003). Content effects in self-directed inductive learning. *Learning and Instruction*, 13(4), 381-402.
- Wilkening, F., & Sodian, B. (2005). Scientific reasoning in young children: Introduction. *Swiss Journal of Psychology*, 64(3), 137-139.
- Wood, P., Kitchener, K., & Jensen, L. (2002). Considerations in the design and evaluation of a paper-and-pencil measure of epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 277-294). Mahwah, NJ: Erlbaum.
- Xue, Y., & Larson, R. C. (2015). STEM Crisis or STEM Surplus: Yes and Yes. *Monthly Lab. Rev.*, 138, 1-14.
- Young, J. (2014). iPolicy: Exploring and evaluating the use of iPads in a social welfare policy course. *Journal of Technology in Human Services*, 32, 39-53.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172-223.