Steffen Alexander Tröbst

The Mediation of Effects of Family Socio-economic Status on Fourth- and Sixth-Graders' Science Achievement by Individual Cognitive and Motivational Propensities

2014

Psychologie

The Mediation of Effects of Family Socio-economic Status on Fourth- and Sixth-Graders' Science Achievement by Individual Cognitive and Motivational Propensities

Inaugural-Dissertation
zur Erlangung des Doktorgrades
im Fachbereich Psychologie und Sportwissenschaft
der Westfälischen Wilhelms-Universität in Münster

Vorgelegt von
Steffen Alexander Tröbst
aus Recklinghausen
2014

Dekanin/Dekan:	Prof. Dr. Manfred Holodynski
Erste Gutachterin:	Prof. Dr. Kornelia Möller
Zweiter Gutachter:	Prof. Dr. Manfred Holodynski
Tag der mündlichen Prüfung:	12. Februar 2015

Table of Contents

List of Tables

List of Figures

Acknowledgements

The present analyses relied on data from the initial cross-sectional phase of a longitudinal research project (*Professionswissen von Lehrkräften, verständnisorientierter naturwissenschaftlicher Unterricht und Zielerreichung im Übergang von der Primar- zur Sekundarstufe* – PLUS) on the development and interplay of teachers' professional competence, science instruction, and student outcomes in science in the transition from elementary to secondary education in Germany. The project was funded by the German Research Foundation (DFG) as part of the Research Group Teaching and Learning of Science NWU Essen.

I am thankful to Prof. Dr. Hans Fischer and Prof. Dr. Kornelia Möller for their trust in my ability to work properly with the complex data set of the cross-sectional study of the PLUS-project. Likewise, I am indebted to Dr. Thilo Kleickmann, Prof. Dr. Alexander Kauertz, Dr. Kim Lange, Dr. Anne Ewerhardy, Dr. Annika Ohle, and Katharina Fricke. They devised the instruments, organized the study, and collected the data.

Last but not least, I thank Prof. Dr. Manfred Holodynski for helpful comments on an earlier version of this paper and Prof. Dr. Martin Bonsen for serving as the third member of my supervisory committee.

Abstract

In the wake of the international large-scale assessments of student achievement of the past fifteen years, the German educational system has been found – despite a slightly reverse trend in recent years – to exhibit a comparatively close association between students' social background and academic achievement. From a genuine psychological perspective, a perspective that focuses on processes within individual learners, it appears natural to ask which student characteristics relevant for learning processes mediate the effects of socio-economic status on student achievement. So, the present investigation aimed to complement the extant knowledge base about social gradients in academic achievement by examining the relative contributions of various psychological propensities to the mediation of effects of family socio-economic status on student achievement in the context of a concrete teaching unit in formal education.

For this purpose, the examinations utilized data collected in the initial cross-sectional phase of a longitudinal research project on the development of classroom climate, science instruction and students' interest in science in the transition from elementary to secondary education in Germany. In the cross-sectional phase 60 fourth- and 54 sixth-grade teachers had been asked to provide their classrooms with a series of three 90-minute lessons on the topic of evaporation and condensation. The corresponding student sample consisted of 1326 fourth- and 1354 sixth-graders. Assessments conducted prior to the series of lessons offered the cognitive propensities of students' fluid ability and topic-specific prior knowledge as well as the motivational propensities of students' self-concept of ability in science and individual interest in science as potential mediators of the relation between social background and post-instructional science achievement. Students' social background was operationalized in terms of the International Socio-Economic Index of occupational status, which was coded on the basis of parental reports on current occupation.

Tests and scales were calibrated according to psychometric models of the Rasch family. To handle the problem of missing data, 50 imputed data sets were generated by means of multiple imputation with chained equations. The imputation model included variables from external sources (e.g., local unemployment rate). After computation of the social gradient in post-instructional science achievement, mediation analyses were conducted first separately for each potential mediator. These analyses took the multilevel nature of the data as well as mean differences in dependent variables between grade levels or school types into consideration. Furthermore, it was explored to what extent relations between family socio-economic status, mediators and post-instructional science achievement were moderated by grade level or

school type. Eventually, the findings were summarized in a final multilevel path model containing all relevant mediators.

The cognitive propensity of topic-specific prior knowledge was identified as the central mediator of the relation between family socio-economic status and post-instructional science achievement in the transition from elementary to secondary education. In addition to prior knowledge, only students' self-concept of ability in science appeared to mediate a very small portion of the effect of socio-economic status on achievement in its own right. However, there were two substantial moderations of the mediational pathways between social background and post-instructional achievement. On the one hand, a relevant association between socio-economic status and self-concept of ability in science was observed exclusively for the subsample of fourth-graders. Presumably, contextual changes that accompany the transition from elementary to secondary education and influence students' self-concept of ability – for instance, the occurrence of the big-fish-little-pond effect – overlaid potential top-down effects of family socio-economic status on self-concept of ability. On the other hand, the importance of prior knowledge as a predictor of science achievement tended to be lower for students from the negatively selective school type of Hauptschule than for students from the comprehensive school type of Grundschule. This observation conforms to the conception that the relative importance of prior knowledge as a precursor of academic achievement increases with students' absolute level of competence and the absolute complexity of the content to be acquired.

The paramount relevance of students' topic-specific prior knowledge for the mediation of effects of socio-economic status on post-instructional science achievement in the present examinations accords with evidence pinpointing differential competence development outside of formal schooling as the primary source of social disparities in academic achievement. Moreover, the current analyses uncovered no systematic association between family socio-economic status and individual interest in science, and thus no mediation of effects of social background on post-instructional achievement via individual interest. This absence of social inequalities in students' individual interest in science points towards – at least for students in late childhood and early adolescence – the potential of socially disadvantaged children's emotional eagerness to learn as a resource for the reduction of social disparities in academic achievement.

Introduction

The association between family socio-economic status and students' academic achievement attracts attention from sociology, educational research, and psychology, with each of these academic disciplines contributing their own genuine viewpoints to the investigation of the phenomenon (cf. Maaz, Baumert, & Trautwein, 2009). From a sociological perspective the influence of socio-economic status on student achievement and educational attainment can be viewed as an important mechanism for the reproduction of social inequalities over generations and thus for the stability of the social stratification of societies (Bourdieu, 1983). In educational research there is a special interest in the contribution of the outfit of educational systems to the formation of social disparities in student achievement (Maaz, Trautwein, Lüdtke, & Baumert, 2008). Apparently, this interest is inspired by the perception that the active reduction of social inequalities or at least the fair treatment of students from differing social backgrounds is one of the core responsibilities of contemporary educational systems. Finally, adopting a psychological point of view, family socio-economic status appears as a factor that influences the extent to which students can take advantage of learning opportunities in school (Helmke, 2007). Refining this approach, it is possible to ask which student characteristics directly relevant for learning processes mediate the effects of socio-economic status on academic achievement (Baumert, Watermann, & Schümer, 2003). International large-scale assessments have repeatedly demonstrated that the association between family socio-economic status and student achievement is particularly large in Germany (e.g., Baumert, Stanat, & Watermann, 2006a; Marks, Cresswell, & Ainley, 2006). Therefore, the German educational system appears especially suitable to investigate possible psychological mediators of this association. From a practical point of view, the identification of such mediators might aid the design of learning environments that maximize achievement for all students.

Theoretical Background

Conceptualizations of Socio-economic Status

Despite its omnipresence as a construct in educational research, socio-economic status lacks a unified definition (Bornstein & Bradley, 2003). In very general terms, socio-economic status describes the relative position of an individual or a group within a social hierarchy of wealth, power and prestige (Mueller & Parcel, 1981; Sirin, 2005). Family socio-economic status is usually assessed by parental income, parental education, or parental occupation, or a composite of these features (e.g., Gottfried, 1985). Occasionally, the consideration of such parental features is complemented by the assessment of the availability of certain household items, like books, or computers (Baumert & Maaz, 2006; May, 2006; McLoyd, 1998). Naturally, the aforementioned parental characteristics are not independent from each other, yet plausibly contribute in different ways to the association between family socio-economic status and students' academic achievement (e.g., Hauser & Huang, 1997; Kim & Sherraden, 2011; Yang & Gustafsson, 2004). For instance, it might be expected that parental education is specifically related to parental aspirations for their children's educational success and to parental capabilities to provide their offspring with advice on educational career choices, whereas parental income obviously bears more directly on the material resources available at home.

Sociological research has devised a variety of categorization schemes for an internationally comparable coding of socio-economic status, which all use information on current occupational status. Based on the evaluation of ratings in approximately 60 countries, the Standard International Occupational Prestige Scale (SIOPS) allows ranking individuals according to the socially perceived prestige of their occupation (Treiman, 1977). Apart from that, via integration of information on occupational status, self-employment, and supervisory status, it is possible to group individuals into distinct social classes, commonly known as the EPG classes (Erikson, Goldthorpe, & Portocarero, 1983). In contrast, the International Socio-Economic Index of occupational status (ISEI) allows assigning individuals, based on their current occupation, scores from a continuous scale. This continuous scale has been derived from an optimal scaling procedure that maximized the relevance of occupation as an intervening factor between education and income, thereby putting an emphasis on the financial aspect of socio-economic status (Ganzeboom, de Graaf, Treiman, & de Leeuw, 1992).

The Association Between Socio-economic Status and Academic Achievement

Meta-analytic results indicate a moderate to strong relation between socio-economic status and academic achievement (Sirin, 2005; White, 1982). Though this association figures as a prominent feature of many contemporary educational systems, it appears to be especially pronounced when strict ability tracking is practiced (Schnabel, Alfeld, Eccles, Köller, & Baumert, 2002). Correspondingly, large-scale international assessments demonstrated that the relation between socio-economic status and student achievement is, for various subjects and age levels, comparatively large in Germany, despite a tendency for a slight attenuation of this relation during the past fifteen years (e.g., Ehmke & Baumert, 2007; Ehmke & Jude, 2010; Baumert et al., 2006a). For the domain of science, the most recent national assessments of German fourth-graders revealed once more a comparatively strong association between socio-economic status and student achievement (Bonsen, Frey, & Bos, 2008; Stubbe, Tarelli, & Wendt, 2012). Whether the impact of socio-economic status on academic achievement actually increases or decreases during the years of formal schooling is less clear than the sheer existence of that impact and probably depends upon a range of contextual factors. At least three sets of arguments can be put forward in favor of growing social disparities in academic achievement (Caro & Lehmann, 2009; Maaz et al., 2009).

First, it is possible to argue that the contemporary educational systems of western societies discriminate against children from families of low socio-economic status (Bourdieu & Passeron, 1990). In this vain, it has been shown that teachers tend to hold low expectations for the achievement of children from disadvantaged social backgrounds (Diamond, Randolph, & Spillane, 2004; Helsper, Kramer, Hummrich, & Busse, 2009). These expectations may, in turn, cause teachers to interact with students of low socio-economic status in less stimulating ways than with students of middle or high socio-economic status, thereby acting as self-fulfilling prophecies (McLoyd, 1998; Quay & Jarrett, 1986). With respect to the organizational frame provided by contemporary educational systems, it has to be noted that ability grouping and tracking tend to place children from families of low socio-economic status in learning environments composed of students with low ability and children from families of high socio-economic status in learning environments comprising students with high ability (Baumert, Stanat, & Watermann, 2006b; Baumert, Trautwein & Artelt, 2003; Haller & Davis, 1981; Oakes, 1985). Especially in Germany, where children are tracked into different school types after elementary education, these grouping practices contribute to an amplification of the association between family socio-economic status and student achievement in the course of formal schooling (Becker, Lüdtke, Trautwein, & Baumert, 2006; Maaz et al., 2008). Even after controlling for individual ability, children from privileged family backgrounds are more likely than their socially disadvantaged peers to become enrolled in the academic track of the German educational system (Ehmke & Baumert, 2007; Lehmann, Peek, & Gänsfuß, 1997). Those contributions to social inequalities in academic achievement not attributable justifiably to social disparities in ability prior to tracking are commonly referred to as secondary effects of social origin; in contrast to primary effects of social origin, which result legitimately from social discrepancies in competence prior to ability grouping and appear, at first glance, not necessarily unfair (Bos et al., 2004; Boudon, 1974; Maaz & Nagy, 2009; Maaz, Schroeder, & Gresch, 2010). In summary, students of high socio-economic status tend to profit – in part unduly – from placement in the academic track, the school type of Gymnasium, where they enjoy a more favorable composition of the student body and more challenging learning opportunities than children of low socio-economic status in other school types, most notably the lowest track of secondary education, the school type of Hauptschule (Baumert et al., 2006b).

Second, socially underprivileged families' and students' reservations towards and disillusionment with formal education might engender increasing social inequalities in academic achievement in the course of schooling (Guo, 1998). Specifically, as they mature, students of low socio-economic status might commence to realize that their choices and opportunities in life are restricted by their family background. In consequence, they should experience a severe loss of interest in academic activities. Of course, such motivational processes can be construed as a complementary reaction to actual discrimination by educational systems but also as a corollary of experiences of failure due to primary social disparities. In fact, secondary effects of social origin on academic achievement result to a considerable portion from socially disadvantaged parents' and students' decisions to refrain from pursuing advanced educational qualifications (Boudon, 1974; Breen & Goldthorpe, 1997; Ditton & Krüsken, 2006; Goldthorpe, 1996).

Third, differential competence development of students of low and high socio-economic status in their respective home environments is apt to engender an intensification of the relation between social background and academic achievement across the years of formal schooling. In other words, the accumulation of primary effects of social origin is suitable to progressively widen social discrepancies in academic achievement. In agreement with this contention, it has been shown that disparities in achievement between children from socially advantaged and disadvantaged homes widen particularly during summer breaks, while competence development during periods of schooling proceeds in a parallel fashion (Alexander, Entwistle, & Olson, 2001, 2007; Becker, Stanat, Baumert, & Lehmann, 2008). So, in this perspective, teachers and schools do not act as promoters of social inequalities, but formal instruction actually counteracts adverse effects of disadvantaged home environments and diminishes differences in academic achievement between children from different social backgrounds (Alexander et al., 2007; Downey, von Hippel, & Broh, 2004; cf. also Adey, 2007 for similar considerations concerning the equalizing power of compulsory schooling with respect to the development of intelligence).

Despite the aforementioned lines of reasoning in favor of a growth of social disparities in academic achievement during formal education, the empirical evidence with regard to the development of the strength of the association between social background and student achievement is not unequivocal (Maaz et al., 2009). Recent meta-analytic research has found a substantial intensification of the relation between socio-economic status and academic achievement from kindergarten to middle school (Sirin, 2005), in contrast to a preceding meta-analysis revealing a diminishment of the social gradient in academic achievement with students' grade level (White, 1982). Likewise, studies of the German educational system have yielded evidence in favor of a widening of social gaps in academic achievement (Baumert, Nagy, & Lehmann, 2012), of stable social disparities in achievement (Baumert, Köller, & Schnabel, 2000; Ehmke, Hohensee, Siegle, & Prenzel, 2006) as well as of decreasing social inequalities in academic achievement (Caro & Lehmann, 2009).

In summary, the association between family socio-economic status and student achievement constitutes a ubiquitous phenomenon (Sirin, 2005). A number of theoretical considerations as well as meta-analytic results point in unison towards an increase of this association in the course of formal schooling (Caro & Lehmann, 2009; Maaz et al., 2009; Sirin, 2005). In case of the German educational system, the growth of social inequalities appears to be especially spurred by selective allocation of children with varying social backgrounds to different school types in secondary education and subsequent differential development of student achievement between those school types (Baumert et al., 2006b; Maaz et al., 2008; Marks et al., 2006; Trautwein, Köller, Schmitz, & Baumert, 2002; van Ophuysen & Wendt, 2009).

The Concepts of Cultural and Social Capital

In order to explain the association between family socio-economic status and student achievement with improved precision the rather static family characteristics of parental income, education, and occupation, have to be supplemented with conceptualizations of family processes that model the transformations relevant for the relation of socio-economic status and achievement (Parcel, Dufur, & Zito, 2010). In this respect, the concepts of cultural and social capital are of particular prominence (Coleman, 1988; Marks et al., 2006; Portes, 1998; Yang & Gustafsson, 2004). These concepts have developed as extensions of the classical concept of financial capital (Bourdieu, 1983). Generally, the concepts of cultural and social capital denote all cultural and social resources that are apt to increase the behavioral options of individuals and thereby have the potential to foster the socio-economic status of these individuals. In particular, cultural capital encompasses human capital, the possession and use of specific culturally relevant items as well as the successful operation with the symbolic representations of the mainstream culture. The notion of social capital refers to social relationships, like mutual obligations, expectations, and trust, which can be utilized to generate socio-economic status. The different forms of capital can be converted into each other. So, parents can invest their financial capital in the acquisition of cultural resources, e.g. books, for their home. Their children, in turn, can utilize such cultural resources at home to generate educational success and attain advanced qualifications. Eventually, these qualifications – as a form of human capital – can be transformed into financial capital again by exercising well-paid occupations. In this way, access to and transformations of financial, cultural, and social capital contribute to the hierarchical stratification of societies (Portes, 1998).

Field research in the context of international large-scale assessments has demonstrated that the access to cultural capital is related to both family socio-economic status and student achievement (e.g., Ehmke, 2008). Specifically, cultural and social resources of families partly mediate the effects of socio-economic status on student achievement (Baumert et al., 2003; Jungbauer-Gans, 2004, 2006; Maaz & Watermann, 2007). In addition, there is evidence that cultural capital outstrips social capital in terms of its relevance as a mediator of the relation between socio-economic status and academic achievement, with mechanisms associated with cultural capital encompassing students' participation in the legitimate culture as well as parents' deliberate and ambitious support of their offspring's education in the sense of concerted cultivation (Cheadle, 2009; Jaeger, 2011; Lareau, 2003; Marks et al., 2006).

Individual Propensities for Achievement as Mediators of Effects of Socio-economic Status

Conceptualizations of family processes involving the notions of cultural and social capital are well suited to explain the differential participation of social classes in formal education or the acquisition of competencies outside of school settings (Breen & Goldthorpe, 1997; Marks et al., 2006; Winne & Nesbit, 2010). However, models of family processes are less apt to account for disparate academic achievement of children of low and high socio-economic status in given formal learning environments. In order to elucidate the contribution of social background to achievement in specific instructional situations, it is necessary to take individual student characteristics relevant for learning into consideration.

Following this rationale, analyses of German 15-year-olds' reading competency have revealed that psychological characteristics of students significantly explain variance in achievement beyond institutional factors, i.e. the school type under investigation, and family processes (Baumert et al., 2003). Notably, the additional inclusion of psychological characteristics as predictors of achievement in a procedure of stepwise hierarchical regression diminished the effects of family socio-economic status thereby producing evidence for a partial mediation of the influence of social background on academic achievement by those characteristics. The individual student characteristics under consideration encompassed basic cognitive ability, decoding capacity as a proxy for prior knowledge, interest in reading, and metacognitive learning strategies.

Similarly, in case of the domains of math and science, it has been suggested to distinguish between opportunity factors, propensity factors and distal factors as predictors of achievement (Byrnes & Miller, 2007). Within this taxonomy, opportunity factors are defined as factors that relate to the confrontation of students with content to be learned, for instance the participation in specific courses or specific teacher behaviors. Propensity factors are described as factors that are concerned with the capacity and willingness of students to learn specific content once confronted with it. Accordingly, students' fluid intelligence and interest in a given domain qualify as examples of propensity factors. Distal factors, finally, are seen as factors that enable and explain the occurrence of opportunities and the development of propensities. Thus, family socio-economic status and parental expectations can be viewed as instances of distal factors. Using data from the National Education Longitudinal Study of 1988 (NELS:88), this general framework has been tested empirically with regard to the math and science achievement of students in the United States. In agreement with the aforementioned results for reading achievement in Germany, it has been found that propensities, specifically students' prior knowledge, mediated effects of socio-economic status on subsequent math and science achievement.

Apparently, students' psychological characteristics or propensities for learning co-vary with students' socio-economic background, and, in consequence, possibly mediate influences of socio-economic status on academic achievement. From an applied point view, it appears far more manageable and feasible for teachers to adapt instructional practices to the cognitive and motivational capacities of socially underprivileged students than to influence the living conditions and internal processes of their families. As will be shown in detail in the following considerations, in the cognitive realm both fluid intelligence and prior knowledge are promising candidates for a mediating role between social background and achievement, whereas in the motivational arena self-concept of ability and academic interest are potentially of relevance as mediators for the connection of socio-economic status to student achievement.

As students traverse the years of schooling, their general mental ability, their academic achievement, their self-concept of ability and their interest in academic learning undergo characteristic developmental changes. In the past much research has been devoted to the delineation of group-average alterations in these variables during the course of formal education. For instance, students' fluid ability increases from early childhood to early adulthood (Cattel, 1987; Rindermann, 2011), whereas their academic interest displays a substantial decline during the same period (Frenzel, Goetz, Pekrun, & Watt, 2010; Krapp, 1998). However, for the investigation of the mediation of effects of socio-economic status on academic achievement by cognitive and motivational propensities it is not the absolute value of these variables but the strength of the associations between these variables that is of primary relevance for the analyses.

Furthermore, in case of the German educational system, the development of the relation between socio-economic status and academic achievement as well as the development of its mediation by students' cognitive and motivational propensities must be viewed against a background of a comparatively early transition from elementary to secondary education, regularly after fourth grade (LeTendre, Hofer, & Shimizu, 2003). This transition from comprehensive elementary education to secondary education is inextricably intertwined with clearly visible tracking of students into separate school types (Maaz et al., 2008; Pietsch & Stubbe, 2007). Within the taxonomy of school types in German secondary education, the school types of Hauptschule and Gymnasium form, respectively, the negative and positive endpoints of a continuum of learning environments with favorable compositional and institutional features (Baumert et al., 2006b). On the one hand, with respect to the formation of academic achievement and mental ability, the school types of secondary education represent differential learning environments, with the school type of Hauptschule producing the smallest progress and the school type of Gymnasium spawning the largest growth (Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012; van Ophuysen & Wendt, 2009). On the other hand, with regard to the development of self-concept and academic interests, the school type of Hauptschule, in contrast to the school type of Gymnasium, might function as a protective niche for low-achieving students (Baumert et al., 2006b).

Intelligence

The concept of intelligence figures as one of the most prominent constructs of scientific psychology, bearing relevance for almost all its sub-disciplines; interindividual differences in intelligence predict a wide range of life outcomes from such diverse areas as health, education, and occupational success (e.g., Gale, Batty, Tynelius, Deary, & Rasmussen, 2010; Kuncel, Hezlett, & Ones, 2004; O'Toole & Stankov 1992; Schmidt, 2002; Strenze, 2007; see Deary, 2012, for a recent review). In agreement with its outstanding role, the concept of intelligence has a long and complex history in scientific psychology, in which definitions and theories of the structure of intelligence abound (Cianciolo & Sternberg, 2004; Hunt, 2011; Jensen, 1998). Notably, this history is intimately intertwined with the invention, development, and refinement of techniques of factor analysis (Bartholomew, 2004).

Contemporary scholars in the psychometric tradition of intelligence research agree that general intelligence is best captured in terms of the general factor, coined the g-factor, that can be extracted from a diverse collection of mental tests by means of hierarchical factor analysis (e.g., Carroll, 1993; Deary, 2012; Jensen, 1998, 2002; see Sternberg & Grigorenko, 2002, for a discussion on the utility of the concept of g-factor). In particular, a three-layered structure has been found in such analyses. At the bottom of the hierarchy lie first-order factors representing variance at the level of specific tests. These are superseded by second-order factors pertaining to broad domains of cognitive functioning (e.g., verbal or numerical ability). At the top of the hierarchy resides the g-factor of general intelligence. Despite vast agreement on the framework of a three-layered structure with the g-factor at its top, there is a persisting debate concerning the precise number and nature of second-order factors (Carroll, 1993; Deary, 2012). Moreover, it is noteworthy that mental tests differ considerably with regard to their loadings on the g-factor. For instance, tests calling for decontextualized reasoning – such as Raven's Progressive Matrices (Raven, Raven, & Court, 1998) or the Culture Fair Intelligence Test (Cattell, 1957) – typically are found to possess extraordinarily high loadings on the g-factor, whereas tests drawing on domain-specific knowledge, e.g. word knowledge, commonly display moderate loadings on the g-factor. However, it is evident that the definition of intelligence in terms of the g-factor is first and foremost a psychometric definition, i.e. a definition that consists of loadings on first-order factors, second-order factors and the general factor. Thus, a person's general mental ability in terms of the g-factor is neither identical to that person's score on a specific test with a high loading on the g-factor nor to that person's conventional intelligence quotient derived from a specific test battery. Furthermore, it is obvious that verbal descriptions of the concept of intelligence are secondary to its definition. They represent merely attempts to summarize concisely the complex fabric of a hierarchical factor analysis identifying the g-factor (Blair, 2006; Jensen, 1998; see Nisbett et al., 2012, for a de-emphasis of the concept of the g-factor).

Apart from its investigation as an outcome of formal schooling (Becker et al., 2012; Ceci, 1991), intelligence has received attention in educational research primarily as a prerequisite for efficient learning and as a proxy for prior knowledge (e.g., Hasselhorn & Grube, 1997; Opdenakker & van Damme, 2001). However, educational researchers tend to assess intelligence rather unsophisticatedly, referring to scores of any established test battery or of highly g-loaded subtests of test batteries as intelligence instead of measuring the g-factor of general intelligence by means of a maximally large and diverse set of mental tests (e.g., Becker et al., 2012; Jensen & Weng, 1994; Nisbett et al., 2012). In face of this, the renowned distinction between fluid intelligence and crystallized intelligence has been invoked to reconcile the principle of a g-factor at the apex of a hierarchically structured intelligence construct with the common practice of assessing intelligence by narrow but highly g-loaded tests (Baumert, Lüdtke, Trautwein, & Brunner, 2009; Blair, 2006; Brunner, 2008; Brunner & Kraus, 2010; Brunner, Kraus, & Kunter, 2008; Cattell, 1987). Fluid intelligence can be thought of as the knowledge-independent capacity to process information for problem solving and reasoning. Thus, it can be considered raw mental horse power or pure mental fitness. Accordingly, the development of fluid ability is claimed to be determined primarily by biological factors. On the contrary, the concept of crystallized intelligence captures cognitive functioning that relies on the retrieval of stored knowledge from long-term memory. Therefore, the formation of crystallized intelligence is thought to depend on experience, acculturation and formal education (Blair, 2006; Brunner, 2008; Cattell, 1987; Horn & Noll, 1997). Moreover, there is a unidirectional relation between the two forms of intelligence. The investment of fluid ability in learning opportunities spawns the generation of crystallized intelligence (Cattel, 1987; Kvist & Gustafsson, 2008; see Christensen, Batterham, & Mackinnon, 2013 for contradicting findings). This relation is accompanied by characteristic developmental trajectories for the two forms of intelligence across the life span. Fluid ability rises from birth onwards and declines after a peak in early adulthood. Naturally, crystallized intelligence also increases during childhood and adolescence. In addition, it has the potential to continue its growth beyond early adulthood (Cattell, 1987; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002).

Intelligence and achievement. Considering that the first intelligence tests were devised as tests for school aptitude (Simon & Binet, 1904), intelligence appears as a natural predictor of student achievement. With respect to the thousands of studies documenting the association between intelligence and student achievement, this association can be considered an unquestionable fact (Deary, 2012; Deary, Strand, Smith, & Fernandes, 2007; Jensen, 1998). In general, mental ability enables individuals to acquire new concepts, deduce relations between concepts, and solve unfamiliar problems. These processes form a fundamental part of learning processes in school.

Though general intelligence in terms of the g-factor traditionally has been seen as the best single predictor of academic achievement (Glutting, Watkins, & Youngstrom, 2003; Gottfredson, 2002; Neisser et al., 1996; Rohde & Thompson, 2007), intelligence researchers have begun to construe more fine-grained models of the relation between intelligence and student achievement, focusing on the specific importance as well as on the interplay of broad cognitive abilities such as reasoning, mental speed or short-term memory – in the psychometric tradition of intelligence research these are conventionally located side by side at the level of second-order factors – in the prediction of academic achievement. Overcoming a purely factor-analytic approach to the conceptualization of intelligence, the distinction of basic, i.e. relying primarily on automatic mechanisms, and complex, i.e. involving conscious multiple manipulations of information, cognitive abilities has enabled the formulation of causal hypotheses with regard to the interrelations of second-order cognitive abilities in the vein of information-processing models of cognitive functioning (Vock, Preckel, & Holling, 2011). In particular, it has been assumed that basic cognitive abilities, such as mental speed and short-term memory, represent constraining factors for the execution of complex, higher-order forms of cognitive functioning, such as reasoning or divergent thinking (Campione & Brown, 1978; Deary, 1995; Floyd, 2005; Neubauer, 1997). Accordingly, it has been shown for German secondary school students that the effects of the basic cognitive abilities of mental speed and short-term memory on student achievement are mediated, at least partly, by higher-order cognitive processes (Rindermann & Neubauer, 2004; Vock et al., 2011). In this context, the complex cognitive ability of reasoning, which can be equated roughly with the concept of fluid ability, has been found to be a specifically strong predictor of achievement (Freund & Holling, 2008; Rindermann & Neubauer, 2004; Vock et al., 2011). These results are complemented by the finding that the relevance of the basic cognitive ability of mental speed for academic achievement tends to decrease with students' age (Floyd, Keith, Taub, & McGrew, 2007; Taub, Floyd, Keith, & McGrew, 2008).

The association of the g-factor of intelligence in general and fluid ability in particular with academic achievement varies systematically with the academic subject under consideration: The relation is specifically pronounced for abstract subjects, in which the cumulative acquisition of domain-specific knowledge is based on sequences of hierarchically ordered concepts (cf. Colom & Flores-Mendoza, 2007). Accordingly, correlations of intelligence and fluid ability with math and science achievement have been found to be comparatively large, as opposed, for instance, to correlations between intelligence or fluid ability and achievement in the social sciences (e.g., Deary et al., 2007; Rindermann & Neubauer, 2004; Vock et al., 2011). In the same vein, it has been shown that math achievement in elementary school is closely related to fluid ability whereas early reading achievement is specifically associated with crystallized forms of intelligence, such as word knowledge (Floyd et al., 2007; Taub et al., 2008).

When viewed under a developmental perspective, the association between intelligence and academic achievement reveals a tendency to attenuate in the course of formal schooling; the relation is strongest in elementary school and weakens with each subsequent step in the hierarchy of educational attainment (e.g., Baumert et al., 2009; Jensen, 1980). In agreement with this general trend, a longitudinal study of German elementary school students has demonstrated that the relative importance of intelligence as a predictor of achievement decreases from second to fourth grade while the relevance of domain-specific prior knowledge for the prediction of achievement increases simultaneously (Helmke & Weinert, 1997; Weinert & Helmke, 1995). Educational researchers and intelligence researchers have contributed differing viewpoints to the explanation of the phenomenon of the attenuation of the relation between intelligence and achievement. In particular, educational researchers have focused on the role of the familiarity and specificity of the material to be mastered. When content to be learned is unfamiliar, as is the case for all school subjects at the beginning of formal education, general intelligence is a powerful predictor of knowledge acquisition. However, as soon as some relevant domain-specific prior knowledge has been accumulated, it starts to outstrip intelligence with respect to its importance for the prediction of academic achievement. Eventually, in upper secondary and tertiary education, when content to be learned is highly specific and builds upon extended prior knowledge, the relevance of intelligence for the prediction of achievement is comparatively small. Accordingly, the differential provision of learning opportunities contributes to the dilution of the relation between intelligence and achievement (Baumert et al., 2009). This line of reasoning agrees with the finding that measures of achievement aggregated over a range of academic subjects tend to display larger associations with intelligence than measures of achievement restricted to specific academic subjects (e.g., Colom & Flores-Mendoza, 2007; Deary et al., 2007). Nonetheless, intelligence researchers have stressed the impact of restrictions of variance in general mental ability on the association between intelligence and academic achievement. On the one hand, students with low intelligence are more likely than their peers to drop out early from the system of formal education. On the other hand, practices of tracking, prevalent in secondary education, constrain the variance of mental ability in specific learning groups, thereby naturally reducing the association between intelligence and achievement within those groups (e.g., Jensen, 1980, 1998). In fact, the reduction of the importance of intelligence for academic outcomes due to the formation of homogeneous learning groups might actually be considered a pedagogical intention of ability grouping.

Intelligence and socio-economic status. It has been estimated that family socio-economic status accounts for approximately 20% of the variation in children's intelligence (e.g., Gottfried, Gottfried, Bathurst, Guerin, & Parramore, 2003). As the development of intelligence is shaped by genetic and environmental factors (Plomin & Spinath, 2004; Nisbett et al., 2012), it is possible to identify two pathways that plausibly relate family socio-economic status to children's general mental ability. On the one hand, there is a non-causal pathway that contains parental intelligence as the common source of both family socio-economic status and children's intelligence. On the other hand, there is a causal pathway that features family socio-economic status as the cause of specific home environments and childhood experiences that, in turn, influence the development of children's general mental ability (e.g., Colom & Flores-Mendoza, 2007; Loehlin, 2000). In a broader theoretical perspective, these two pathways represent instantiations of the approaches of social selection and social causation to the phenomenon of the association between socio-economic status and mental ability. According to the approach of social selection, intelligence, largely genetically predefined, determines where individuals settle within the occupational hierarchy. On the opposite, according to the approach of social causation, early socio-economic status restricts the subsequent development of intelligence (cf. Farah & Hackman, 2012).

The assumption of a non-causal relation between family socio-economic status and child intelligence rests on two findings. First, parental intelligence exerts a strong influence on parental educational attainment and occupational success, and thus on family socio-economic status (Gottfredson, 2003; Strenze, 2007). Second, intelligence is heritable, i.e. parental intelligence determines a large portion of children's intelligence genetically (Plomin & Spinath, 2004). In particular, the average intelligence of persons employed in a given occupation increases with the complexity of and the qualifications needed for that occupation, whereas the variance of intelligence of persons active in a given job shrinks according to the complexity of and the qualifications required for that job. Apparently, general intelligence has a threshold function with regard to the attainment of specific positions in the occupational hierarchy, with the most complex – and best paid – jobs requiring the highest minimum intelligence (Cronbach, 1960; Gottfredson, 2003). Moreover, general mental ability has been shown to predict such diverse criteria of occupational success as vocational training success, job performance, career success and income (e.g., Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007; Hülsheger, Maier, & Stumpp, 2007; Kramer, 2009; Ng, Eby, Sorensen, & Feldman, 2005).

The construct of intelligence certainly features as the best researched human trait in behavioral genetics (Plomin & Spinath, 2004). Overall, heritability has been estimated to account at least for half of the variance of individual differences in intelligence (Bouchard &

McGue, 1981; Chipuer, Rovine, & Plomin, 1990; Devlin, Daniels, & Roeder, 1997; Plomin, DeFries, McClearn, & McGuffin, 2001). In a developmental perspective, the heritability of general intelligence in terms of the g-factor has been found to change drastically across the life span (Plomin & Spinath, 2004): It is lowest in infancy, accounting for approximately 20% of the variance in intelligence, increases in middle childhood, explaining 40% of the variance in intelligence, and reaches its peak in adulthood, where it may account for as much as 80% of the variance in general mental ability (Davis, Haworth, & Plomin, 2009; McGue, Bouchard, Iacono, & Lykken, 1993; Plomin, 1986; Spinath, Ronald, Harlaar, Price, & Plomin, 2003). This, at first glance counterintuitive, developmental trend is hypothesized to result from a general tendency to seek and create environments that fit and foster the individual genetic potential, a gene-environment interaction (Plomin & DeFries, 1985). In addition to this, shared environmental influences appear to be relevant for intelligence in childhood, but not for intelligence after adolescence (McGue et al., 1993; Plomin & Spinath, 2004). In this context, findings with regard to a possible gene-environment interaction concerning the heritability of intelligence across the hierarchy of socio-economic status are inconsistent (Asbury, Wachs, & Plomin, 2005; Grant et al., 2010; Hanscombe, Trzaskowski, Haworth, Davis, Dale, & Plomin, 2012; Nisbett et al., 2012; Scarr, 1981).

Home environments of families of low socio-economic status constitute a syndrome of risk factors for children's health, cognitive maturation, socio-emotional development and academic outcomes (Bradley & Corwyn, 2002; Evans, 2004; McLoyd, 1998). Among the correlates of poverty, processes related to early health and nutrition, stressful living conditions and lack of cognitive stimulation appear as particularly relevant for the growth of general mental ability (Bradley & Corwyn, 2002; Farah & Hackman, 2012). Infants born by mothers of low socio-economic status are more likely than children from non-disadvantaged families to suffer from certain deficiencies, such as low birth weight or being small for gestational age (Nepomnyaschy, 2009). In this context, the fact that smoking and drinking during pregnancy – with the fetal alcohol syndrome as its most severe consequence – occurs comparatively often for women of low socio-economic status appears particularly relevant for prenatal cognitive development (May & Gossage, 2001; Paarlberg, Vingerhoets, Passchier, Heinen, Dekker, & van Geijn, 1999; Rahu, Rahu, Pullmann, & Allik, 2010). Disparities between children of low and high socio-economic status are also prevalent in the field of nutrition. As, relative to the calories contained, fresh food is more expensive than pre-fabricated food (Drewnowski & Specter, 2004), and mothers of low socio-economic status often lack the knowledge to cook meals from fresh ingredients (Northstone & Emmett, 2005), children from families of low socio-economic status consume a comparatively high number of fast food meals; in turn, this is correlated with decelerated cognitive development (von Stumm, 2012).

Furthermore, children of low socio-economic status, especially children living in deep poverty, have to face various physical and psychological stressors in their home environments (Conger et al., 1993; Evans & English, 2002; Evans, 2004; Linver, Brooks-Gunn, & Kohen, 2002; McLoyd, 1998; Turner & Avison, 2003). Socially disadvantaged families tend to live closer to heavily trafficked streets, industrial facilities and waste dumps than non-poor families (Bullard & Wright, 1993; Evans, 2004; Massey, 1994). Thus, children from adverse social backgrounds are chronically exposed to various toxins and forms of pollution (Evans, 2004). For instance, children of low socio-economic status display higher blood concentrations of lead than children of high socio-economic status (Brody et al., 1994; McLoyd, 1998). Likewise, children of low socio-economic status tend to suffer from chronic exposure to noise hampering healthy cognitive development (Evans, 2001; Haines, Stansfeld, Head, & Job, 2002). Moreover, they inhabit crowded homes (Evans, 2004; Myers, Baer, & Choi, 1996). This is often accompanied by the absence of clearly designated play spaces within their homes as well as by a lack of access to parks and nature (Newson & Newson, 1976; Sherman, 1994). Under a psychological perspective, socially disadvantaged children are at risk to experience stressful life events (Dubow, Tisak, Causey, Hryshko, & Reid, 1991; Liaw & Brooks-Gunn, 1994). Apart from that, the adverse living conditions of families of low socio-economic status exert decisive indirect influences on children's development (McLoyd, 1998): Socio-economic strain and tension is apt to cause parental dysphoria and depression (McLoyd, 1990). This, in turn, negatively affects parenting behavior, with children of low socio-economic status experiencing harsher, more punitive and more inconsistent parenting than middle- or upper-class children (Duncan, Brooks-Gunn, & Klebanov, 1994; Elder, 1974). Additionally, the dysphoric parents of socially disadvantaged families tend to lack responsiveness to their children's needs (Magnusson & Duncan, 2002; McLoed & Shanahan, 1993). In very general terms, the psychological home environments of children of low socio-economic status are characterized by a relative deprivation of warmth, stability, routines and structure (Brody & Flor, 1997; Dodge, Pettit, & Bates, 1994; Evans, 2004; Jensen, James, Boyce, & Hartnett, 1983).

Last but not least, children growing up in poverty receive less cognitive stimulation at home than their more affluent peers (Bradley & Corwyn, 2002; Evans, 2004; McLoyd, 1998). Particularly, children of low socio-economic status own comparatively few books or no books at all (Newson & Newson, 1977; Sherman, 1994). In fact, self-reports on the quantity of books in a given household serve as a proxy variable for family socio-economic status in large-scale assessments of student achievement (e.g., Baumert & Maaz, 2006). In general, the home environments of socially disadvantaged children appear to be relatively impoverished with respect to the availability of age-appropriate learning resources such as toys teaching shape, color and size (Duncan et al., 1994; Smith, Brooks-Gunn, & Klebanov, 1997). Not surprisingly, parents in families of low socio-economic status spend less time reading to their children than parents in families of high socio-economic status (Coley, 2002). Similarly, amount and quality of parental speech towards children varies with socio-economic status, socially disadvantaged parents talking less but in a more directive fashion with their children (Hart & Risley, 1995; Hoff, Laursen, & Tardiff, 2002). This accords with the classical notion that parents' experiences at the work place determine their parenting style and thus the provision of cognitive stimulation at home (Kohn & Schooler, 1982; Parcel & Menaghan, 1990) Naturally, the outlined disparities in cognitive stimulation are prone to engender social differences in cognitive development as demonstrated by longitudinal studies showing the negative influences of within-family decreases in socio-economic status on the children's home environments and general mental ability (Dubow & Ippolito, 1994; Garrett, Ng'andu, & Ferron, 1994).

Recently, researchers have begun to delineate the traces that socio-economic status leaves in the developing human brain (Farah & Hackman, 2012). For instance, low family socio-economic status has been found to adversely affect the size of the hippocampus, a brain region susceptible to stress and relevant for memory processes (Noble, Houston, Kan, & Sowell, 2012). In the light of such findings, obviously, socially disadvantaged children's early health status and nutrition as well as the aforementioned physical and psychological stressors have the potential to impact the maturation of the brain and thereby the development of general mental ability.

Summary. In intelligence research, general intelligence is commonly defined as the g-factor at the top of a three-layered hierarchal factor analysis of a diverse array of mental tests (Deary, 2012). The idea of an abstract reasoning capacity, of raw mental horse power, as a prerequisite for learning is captured well by the concept of fluid intelligence, which is identified by some researchers as a second-order factor (Carroll, 1993; Cattell, 1987). However, despite high g-loadings of tests of fluid ability, the concepts of fluid ability or decontextualized reasoning capacity should not be equated precipitately with the g-factor of intelligence (Jensen, 1998; but see Brunner, 2008, for a different position).

Almost by definition, intelligence appears as a natural precursor of student learning. In this context, fluid ability has been shown to be a particularly powerful predictor of academic achievement (Freund & Holling, 2008; Rindermann & Neubauer, 2004; Vock et al., 2011). Moreover, the predictive power of intelligence and fluid ability for achievement is moderated by the domain under consideration. It tends to be specifically strong for hierarchically and logically structured domains, i.e. the association of intelligence and fluid ability with achievement is closer for math and science than for languages and social sciences (Colom & Flores-Mendoza, 2007; Deary et al., 2007; Rindermann & Neubauer, 2004). Apart from that, the relation between mental ability and academic achievement underlies a specific developmental trend: It declines across the years of formal schooling (e.g., Baumert et al., 2009; Weinert & Helmke, 1998). This decline is accompanied by a rise of the importance of domain-specific prior knowledge as a predictor of academic achievement. Thus, the attrition of the relation between intelligence and achievement has been attributed to the accumulation of effects of the differential provision and utilization of learning opportunities during the course of schooling (Baumert et al., 2009). Alternatively, the reduction of the association between intelligence and academic achievement has been explained by restrictions of variance in intelligence in higher grades, either due to dropout of low-ability students or due to practices of ability grouping (Jensen, 1980, 1998).

Both non-causal and causal influences can be suspected to contribute to the relation between family socio-economic status and children's intelligence. To a large degree, parental intelligence determines family socio-economic status (Gottfredson, 2003; Strenze, 2007). Moreover, intelligence is highly heritable (Plomin & Spinath, 2004). Thus, children of low and high socio-economic status might simply inherit their parents' intelligence. With respect to a causal contribution to the association between social background and children's intelligence, it has to be noted that family socio-economic status restricts poor children's home environments and living conditions (Bradley & Corwyn, 2002; Evans, 2004; McLoyd, 1998). Specifically, impoverished health and nutrition (Nepomnyaschy, 2009; von Stumm, 2012), environmental stressors such as pollution and noise (Brody et al., 1994; Evans, 2001) and a lack of cognitive stimulation at home (McLoyd, 1998; Sherman, 1994) appear as causes for the relation between socio-economic status and children's intelligence. Under the assumption that the negative effects of impoverished home environments and living conditions accumulate over time, it is plausible that the association between socio-economic status and children's general mental ability intensifies with students' age. The increase in heritability of intelligence across the life span points into the same direction (Davis et al., 2009).

Prior Knowledge

In educational psychology, prior knowledge figures as the central precursor of successful learning (e.g., Alexander, Kulikowich, & Jetton, 1994; Ausubel, 1968; Dochy, Segers, & Buehl, 1999; Krause & Stark, 2006; Murphy & Alexander, 2002). As a result of its general importance and its usage in a wide range of contexts, the provision of a comprehensive definition of the construct of prior knowledge is more complex than it may seem at first glance. The numerous terms used in the literature under the umbrella of prior knowledge – ranging from archival memory to personal knowledge – are testament to this fact (Dochy & Alexander, 1995; Dochy et al., 1999). Certainly, to be identified as prior knowledge, a particular piece of information has to be stored in the knowledge base of a learner and has to be available before the start of a given learning task (Dochy, 1994; Dochy et al., 1999). In order capture the various manifestations of prior knowledge, it has been suggested to resort to a number of key dimensions for the description of its forms of appearances (Dochy & Alexander, 1995; Krause & Stark, 2006).

In accordance with the conventions of cognitive science, prior knowledge can exist as declarative knowledge, procedural knowledge or conditional knowledge. The realm of declarative knowledge covers the knowledge of facts, symbols and concepts of a particular field, i.e. it is concerned with knowing that (Ryle, 1949). The area of procedural knowledge includes knowledge about actions, manipulations and procedures, i.e. it pertains to knowing how (Ryle, 1949), whereas the section of conditional knowledge comprises the knowledge about suitable occasions for the application of specific knowledge, i.e. the knowing when and where (Paris, Lipson, & Wixson, 1983). In addition, it is possible to characterize instantiations of prior knowledge according to the degree of their accessibility; explicit prior knowledge is the object of conscious processing, while tacit prior knowledge resides outside conscious awareness. Furthermore, it has been suggested to contrast conceptual prior knowledge with metacognitive prior knowledge (Dochy & Alexander, 1995). The former is concisely described as enclosing the knowledge of ideas and concepts. Here, by now classical findings on the domain specificity of learning, thinking and problem solving advocate to distinguish between domain-transcending, domain-specific and topic-specific prior knowledge (Alexander, 1992; Alexander, Schallert, & Hare, 1991; Chi, Glaser, & Rees, 1982; Dochy & Alexander, 1995; Glaser, 1984; Krause & Stark, 2006; Murphy & Alexander, 2002). On the contrary, metacognitive prior knowledge entails individuals' knowledge about their own cognition and capacities to regulate that cognition (Garner, 1987). Apart from that, prior knowledge can vary with respect to its size and integration, with greater and more integrated prior knowledge enabling more successful learning, and thus possibly engendering Matthew effects in knowledge acquisition (Stanovich, 1986; Walberg & Tsai, 1983). Last but not least, prior knowledge can be judged according to its agreement with the accepted state of knowledge in a given field of study, i.e. it can be categorized as false or correct. This aspect bears particular relevance for the field of science education, where it emerges in the form of learners' misconceptions about natural phenomena and scientific models (Duit, 1999).

In the domain of science the development of conceptual knowledge is conventionally characterized as conceptual change (diSessa, 2006; Duit & Treagust, 2003; Posner, Strike, Hewson, & Gertzog, 1982; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001). Naturally, in the absence of formal science instruction, even very young children start to form conceptions about and explanations for the physical and biological phenomena they encounter in their daily lives (e.g., Driver, 1989; Vosniadou & Brewer, 1992). In many instances these naïve conceptions do not conform to scientifically acceptable explanations, which have to be propagated in formal science instruction. Thus, for the area of science instruction, the notion of prior knowledge is often identified with the naïve and alternative conceptions that students bring to the classroom and that might hamper the acquisition of scientifically more adequate conceptions (Duit & Treagust, 2003; Wandersee, Mintzes, & Novak, 1994). The idea that students bring conceptions of their own to the classroom or spontaneously devise conceptions of their own in the classroom, which call for careful consideration and possibly for restructuring, lies at the heart of science instruction following the principles of conceptual change (Limón, 2001; Scott, Asoko, & Driver, 1992; Vosniadou et al., 2001). In this context, conceptual development as defined by the framework of conceptual change theory is decidedly not identical to the accumulation of factual knowledge but entails the acquisition of a deep-level understanding of scientifically acceptable explanations for natural phenomena (Harlen, 1998).

For instance, in order to form a scientifically acceptable conception of evaporation and condensation, students have to acknowledge that matter can alter its state, i.e. that it can exist in the solid, fluid or gaseous state, and that matter is always conserved. Specifically, to achieve a sound understanding of the evaporation and condensation of water, students have to grasp that air permanently contains water in the gaseous state (Johnson, 1998; Strunk, 1999; Tytler, 2000). Common naïve conceptions of the process of evaporation include the ideas that water simply disappears, that water is absorbed by objects and that the sun as an active agent transfers evaporated water to another place (Bar & Galili, 1994; Osborne & Cosgrove, 1983; Russell, Harlen, & Watt, 1989; Tytler, 2000). Furthermore, some students believe that evaporation necessitates the supply of external heat (Costu & Ayas, 2005). The idea that evaporated water rises into the air without altering its state represents a prevalent partially correct conception of the process of evaporation (Russell et al., 1989). In order to explain the process of condensation, students resort to the naïve conceptions that water penetrates surfaces (e.g. a beverage can), that objects sweat and that low temperature turns into condensed water (Johnson, 1998; Osborne & Cosgrove, 1983; Tytler, Peterson, & Prain, 2006). Another popular conception is that evaporated water splits up into hydrogen and oxygen and that, complementary, condensed water results from a combination of hydrogen and oxygen (Driver, Squires, Rushworth, & Wood-Robinson, 1994; Osborne & Cosgrove, 1983). Apparently, this idea is informed by encounters with the concepts of atoms and molecules. Accordingly, it is a good example of a misconception that is induced by media or instruction (Duit & Häußler, 1997).

In the wake of international large-scale assessments of student achievement, the conceptualization and operationalization of science achievement – and thus of prior knowledge relevant for subsequent learning of science – in terms of competences within the framework of competence models has gained popularity in educational research (e.g., Klieme, Hartig, & Rauch, 2008; Walter, Senkbeil, Rost, Carstensen, & Prenzel, 2006). Commonly, such competence models set up a grid of a content dimension and a dimension of cognitive operations on that content that underlies the formulation of tasks for the measurement of science competence. In particular, here the content dimension may cover various disciplines of science, e.g. biology, physics, chemistry and earth science as well as subordinate basic concepts specific to each discipline; the dimension of cognitive activities may include such operations as reproduction of knowledge, application of knowledge and reasoning (Kauertz, Fischer, Mayer, Sumfleth, & Walpuski, 2010; Kleickmann, Brehl, Saß, Prenzel, & Köller, 2012). In addition, reminiscent of Bybee's (1997) influential vision of scientific literacy, it is prevalent among competence models for science to distinguish an area of factual and conceptual knowledge, on the one hand, from an area of knowledge of scientific practices and processes, on the other hand (Kauertz et al., 2010; PISA-Konsortium Deutschland, 2007). Furthermore, it is common practice in large-scale assessments of achievement to assign students in a process of proficiency scaling to competence levels. Typically, such competence levels range from rudimentary knowledge of science or the ability to recall simple facts to reflective thinking about science and proficient use of scientific knowledge to solve problems (Kleickmann et al., 2012).

Prior knowledge and achievement. The importance of prior knowledge for subsequent learning constitutes one of the cornerstones of conceptualizations of learning in terms of cognitive psychology (Alexander, 1996; Ausubel, 1968). Students' prior knowledge affects all stages of information processing pertinent to learning (Renkl, 1996, 2008, 2011). In the initial phase of the learning process, prior knowledge helps students to focus on crucial information, select relevant content and avoid distractions (Mandl, Gruber, & Renkl, 1993; Renkl, 2009). Moreover, prior knowledge provides a framework for parsing information into meaningful chunks, thereby freeing cognitive resources for the processing and manipulation of information essential for actual learning (Mayer, 1997; Renkl, 1996, 2011). In this vein, as experts in a given domain differ from novices with respect to the size of their domain-specific knowledge base, empirical research on the nature of expertise has highlighted that pronounced prior knowledge engenders deep-level processing of learning content whereas deficient prior knowledge coincides with the use of surface-level encoding strategies (Alexander, Murphy, Woods, Duhon, & Parker, 1997; Chi, Feltovich, & Glaser, 1981). Finally, prior knowledge represents the basis for the formation of manifold connections between existing knowledge structures and new content. This ensures long-term retention and facilitated retrieval of newly learned information (Mayer, 2009; Renkl, 1996; 2009). In other words, it constitutes the basis for the construction of new knowledge within the individual learner.

In the context of science education, students' prior knowledge in form of conceptions about the physical world can both foster and hinder further learning (Sandoval, 1995). Some of the conceptions held by students prior to instruction contain already elements of scientifically acceptable conceptions so that formal science instruction can straightforwardly ensue in enriching these pre-conceptions (Vosniadou, 1994). Obviously, in this case, prior knowledge facilitates the acquisition of scientifically adequate conceptions. On the contrary, some of the conceptions brought by students to the science classroom stand in open opposition to scientifically acceptable conceptions, i.e. the content that has to be acquired. Accordingly, these conceptions have to be described as naïve conceptions or misconceptions (Wandersee et al., 1994, Wodzinski, 1996). When such misconceptions are not properly addressed in formal science instruction, it is likely that pre-instructional misconceptions and post-instructional scientific conceptions are represented in parallel in students' knowledge bases (Duit, 1999; Vosniadou & Ioannides, 1998), thereby rendering newly acquired scientific conceptions inert knowledge constrained to the context of instruction in school (Renkl, Mandl, & Gruber, 1996). So, here, the general recommendation of educational psychology to activate prior knowledge and the implications of domain-specific approaches to instruction coincide in their consequences for successful teaching (Krause & Stark, 2006; Duit & Treagust, 2003). In agreement with the aforementioned considerations about the role of prior knowledge in basic individual learning processes, field studies of teaching effectiveness and student achievement have regularly identified students' prior knowledge as a central predictor of academic achievement (e.g., Corno, Cronbach, & Kupermintz, 2002; Dochy, 1992; Hattie, 2009; Jones & Byrnes, 2006).

In the course of formal schooling the importance of prior knowledge as a precursor of student achievement steadily increases whereas the role of fluid ability for knowledge acquisition experiences a corresponding decline (Helmke & Weinert, 1997; Weinert & Helmke, 1995, 1998). From a psychological perspective, this finding is often seen as a result of the accumulation of effects of fluid ability in students' prior knowledge, i.e. over time students with higher fluid ability seek more demanding learning opportunities, and participate more successful in these learning opportunities, than their less talented peers (Renkl & Stern, 1994). So, the growing relevance of prior knowledge for academic achievement might be considered a corollary of the gradual conversion of fluid ability into crystallized intelligence (Cattell, 1987). From a content-oriented point of view it is often argued that, as content to be acquired late in the school career is particularly specific and complex, meaningful learning requires relative large and elaborated pre-existing knowledge structures; this idea fits well with the conception of cumulative competence development (Baumert et al., 2009; Schneider, Körkel, & Weinert, 1989). Furthermore, the content-oriented view entails the notion that domains vary with regard to the hierarchical structure of their content, and thus with regard to the preponderance of prior knowledge as a precursor of academic achievement (Renkl, 1996). More specifically, it can be argued that the strength of the predictive dominance of prior knowledge depends not only on the domain under consideration but also on preferred modes of learning and prevalent forms of teaching. Consider, for instance, the difference between rote learning of specialized vocabulary for diverse topics in advanced foreign language instruction and learning the mathematical operations of addition and multiplication in elementary education. In case of the former, a vocabulary test on a new topic offers students a fresh start implying a comparatively weak association between prior knowledge and achievement. On the contrary, in early mathematics education, the operation of multiplication is introduced as repeated addition. Therefore, obviously, when students fail to attain a sound understanding of the operation of addition, their further progress with respect to the acquisition of the operation of multiplication is blocked. However, both the psychological view and the contend-oriented perspective on the processes supplanting the increasing importance of prior knowledge as a precursor of achievement in the course of formal schooling accord with the rationale for the occurrence of Matthew effects in academic achievement (Stanovich, 1986).

Prior knowledge and socio-economic status. Obviously, the acquisition of academically relevant prior knowledge depends on children's involvement in appropriate learning opportunities (Baumert et al., 2009). Thus, many of the processes in home environments – especially cognitive stimulation at home – suspected to participate in the relation between socio-economic status and intelligence development possess immediate relevance for the formation of prior knowledge (Hess & Holloway, 1984; Korenman, Miller, & Sjaastad, 1995; Lee & Croninger, 1994). Probably, those mechanisms even bear more directly on the development of academically relevant prior knowledge than on the genesis of general mental ability. (It may also be difficult to discriminate these two propensities in very young children.) Apparently, for instance, noise and crowdedness hamper the utilization of learning opportunities with respect to the acquisition of prior knowledge (Evans, 2004; Haines et al., 2002; Myers et al., 1996).

The definition of socio-economic status encompassing elements of both income and education, families of low socio-economic status are relatively deprived of the resources necessary to provide their children with age-appropriate learning opportunities (Bradley & Corwyn, 2002; Mercy & Steelman, 1982; Scarr & Weinberg, 1978). On the one hand, considering the costs that are implicated in purchasing books, visiting a zoo or playing a musical instrument, the limited financial capabilities of families with low socio-economic status constrain the learning experiences they can offer to their children (Bradley, Corwyn, Burchinal, McAdoo, & Garcia Coll, 2001; Brooks-Gunn, Klebanov, & Liaw, 1995; Entwistle, Alexander, & Olson, 1994). On the other hand parents of low socio-economic status may lack certain knowledge relevant for the provision of challenging learning opportunities (Ehmke, 2008). However, as the provision of cognitive stimulation at home in families of high socio-economic status centers around verbal activities, it can be surmised that for young children social discrepancies in academically relevant prior knowledge are particularly pronounced in the verbal domain (Hart & Risley, 1995; Hoff et al., 2002; Mercy & Steelman, 1982; Shonkoff & Phillips, 2000). Finally, socially privileged parents have higher aspirations for their offspring's educational attainment than socially disadvantaged parents (Adams, 1998; Stocké, 2009; Wentzel, 1998). In accordance with that, parents of high socio-economic status are more likely than parents of low socio-economic status to enroll their children in non-compulsory preschool education (Anders, 2013; Kratzmann & Schneider, 2009).

Unsurprisingly, social disparities in parental provision of cognitive stimulation at home as well as in children's enrolment in preschool education result in social disparities in prior knowledge and school readiness (Autorengruppe Bildungsberichterstattung, 2012; Böhm & Kuhn, Coley, 2002; Elsässer, 1998; Kratzmann & Schneider, 2009; Niklas, Segerer, Schmiedeler, & Schneider, 2012). It is often argued that, in addition to this initial disparity at the beginning of schooling, processes within schools and classrooms contribute to the emergence of cumulatively increasing differences in academic outcomes between students of low and high socio-economic status (Bradley & McCorwyn, 2002; McLoyd, 1998). (Note that, as competencies acquired earlier in the school career figure as prior knowledge for learning activities later in the school career, here basically the same arguments as mentioned above with respect to the general development of social discrepancies in student achievement apply.) Teachers tend to perceive poor students less positively than their more affluent peers. They also may hold low achievement expectations for socially disadvantaged children. Particularly, teachers with a middle-class upbringing may mistake speech, dress and deportment characteristic for students of low socio-economic status or ethnic minority status as indicators of missing competence (Alexander, Entwistle, & Thompson, 1987). In turn, the differential perception of students of low and high socio-economic status may incline teachers to provide their students differentially with positive attention, learning opportunities and reinforcement (e.g., Alexander & Schofield, 2008; Helsper et al., 2009). With the good – but nonetheless mislead – intent to protect children of low socio-economic status from future failure, at the end of elementary education teachers may even refrain from the referral of socially disadvantaged children to academically challenging tracks of secondary education (Ditton, Krüsken, & Schauenberg, 2005; Nölle, Hörstermann, Krolak-Schwerdt, & Gräsel, 2009; Stubbe & Bos, 2008). In a similar vein, a misfit between the unconscious habitus of working-class children and the norms of the allegedly middle-class institution of school has been postulated as the cause of social disparities in student achievement (Bourdieu & Passeron, 1990). The increased incidence of externalizing behavior and conduct problems in children of low socio-economic status lends further credibility to the hypothesis of differential teacher behavior as a source of social discrepancies in academic achievement (Achenbach et al., 1990; Duncan et al., 1994; McLoyd, Ceballo, & Mangelsdorf, 1996). Moreover, it helps to understand why socio-economic deprivation in early childhood has more detrimental effects on cognitive development than socio-economic deprivation in middle or late childhood: Social disparities in school readiness and behavior at the beginning of formal schooling could shape teacher's affective reaction to and treatment of children of low socio-economic status for their entire subsequent school career (Duncan, Brooks-Gunn, Yeung, & Smith, 1998; McLoyd, 1998). However, the arguments in favor of a growth of social disparities in student achievement due to interactions in school and teacher behavior are opposed by findings of growing social inequalities in student achievement during the summer break, i.e. by findings corroborating the equalizing power of formal schooling (Alexander et al., 2001, 2007; Becker et al., 2008; see also Maaz et al., 2009).

In face of the assumption of Matthew effects in knowledge acquisition, it appears sensible to presume that the relation between family socio-economic status and students' prior knowledge intensifies across the course of formal schooling: Initially relatively small differences in academic abilities between children of low and high socio-economic status should display the tendency to intensify due to the general importance of prior knowledge for learning (Alexander, 1996; Ceci & Papierno, 2005; Stanovich, 1986; Weinert & Hany, 2003). This speculation is also corroborated by the finding of increasing social disparities in academic achievement when schools are closed (Alexander et al., 2001, 2007; Becker et al., 2008). It is very plausible that social gaps in prior knowledge relevant for future academic learning widen across consecutive summer breaks. Likewise, underprivileged parents' low involvement in their children's school life might contribute to increasing social disparities in necessary prior knowledge (Benveniste, Carnoy, & Rothstein, 2003; Evans, 2004; Lee & Croninger, 1994). Moreover, students of low socio-economic status tend to attend schools that lack material resources, that employ teachers with limited qualifications and that serve communities of underprivileged and low-ability students (Baumert et al., 2006b; Ingersoll, 1999; McLoyd, 1998). Practices of tracking, especially when they take social background beyond ability into consideration, are prone to amplify this tendency. That is specifically true for the German educational system (Baumert & Schümer, 2001; Ditton, 2005); several longitudinal studies of student achievement have pointed out that differential learning gains for students from different social backgrounds occur between the different school types of secondary education but not within specific school types of secondary education (Becker, 2009; Baumert et al., 2006b; Köller & Baumert, 2001; Maaz et al., 2009). In other words, the differential provision of institutional learning opportunities, in particular high-quality learning opportunities, is apt to amplify social discrepancies in prior knowledge for future learning.

Summary. What is prior knowledge? Generally, with respect to a given learning task, the nature of relevant prior knowledge can be charted along the dimensions of its state, i.e. declarative, procedural or conditional, its conscious accessibility, i.e. explicit or tacit, its object, i.e. conceptual or metacognitive, and its domain specificity, i.e. domain-transcending, domain-specific or topic-specific (Dochy & Alexander, 1995; Krause & Stark, 2006). In the domain of science, students' tend to bring naïve and alternative conceptions about natural phenomena to the classroom that often do not conform to scientifically acceptable views, which represent the goal of formal science instruction (Duit & Häußler, 1997; Duit & Treagust, 2003). These conceptions constitute a domain-specific form of prior knowledge characteristic for the domain of science. Finally, in large-scale assessments of student achievement, which have generated much of the contemporary awareness for social disparities in academic achievement, science achievement, and thus prior knowledge for subsequent science learning, is conceptualized in terms of science competence by means of competence models (e.g., Kleickmann et al., 2012; Walter et al., 2006).

The activation of prior knowledge fosters learning by influencing the cognitive processes of selection, organization, integration and recall of relevant information (Mayer, 1997; Renkl, 1996). Correspondingly, students' prior knowledge in form of naïve or alternative conceptions about natural phenomena constitutes the basis for learning in the domain of science, as a starting point for either straightforward enrichment or necessary revision (Vosniadou, 1994; Wandersee et al., 1994). In sum, prior knowledge constitutes a primary precursor of learning in school (Hattie, 2009; Jones & Byrnes, 2006). Complementary to the diminishing importance of general mental ability as a predictor of student achievement, the relevance of prior knowledge as a resource for successful learning increases across the years of formal schooling (Baumert et al., 2009; Weinert & Helmke, 1995, 1998).

The acquisition of academically relevant prior knowledge depending on adequate learning opportunities, social disparities in students' prior knowledge arise from underprivileged parents' restricted capacities to provide their children with such opportunities (Baumert et al., 2009; Baumert et al., 2003). Specifically, parents of low socio-economic status may lack knowledge for the provision of appropriate learning opportunities (Ehmke, 2008). They also have comparatively low aspirations for their offspring's educational attainment, partly due to fear of their potential school failure (Ditton & Krüsken, 2009). Furthermore, obviously, socially disadvantaged parents' strained financial resources hamper the capability to purchase age-appropriate toys and learning materials for their children (Bradley, Corwyn, Burchinal, et al., 2001; Entwistle et al., 1994). Though it can be suspected with some justification that processes within classrooms, for example teachers' differential expectations for achievement and corresponding selective attention, contribute to the formation of social discrepancies in academically relevant prior knowledge (Helsper et al., 2009; McLoyd, 1998), findings highlighting the growth of social disparities during the summer break and parallel development of socially diverse students during the school year oppose a contribution of classroom processes to the development of social gradients in prior knowledge (Alexander et al., 2001, 2007; Becker et al., 2008). However, the generally assumed principle of Matthew effects (Ceci & Papierno, 2005; Weinert & Hany, 2003), the summer break phenomenon (Alexander et al., 2001, 2007; Becker et al., 2008) and the implications of ability grouping for the quality of learning opportunities offered to underprivileged students (Baumert et al., 2006b; Maaz et al., 2009) in unison point towards an intensification of the relation between family socio-economic status and academically relevant prior knowledge across the years of formal schooling.

Self-concept of Ability

In a broad perspective, the construct of self-concept of ability belongs to those motivational constructs that are concerned with a person's answer to the question "Am I able to do this task?" (Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006; see also Pintrich, 2003; Pintrich, Marx, & Boyle, 1993). Accordingly, it resembles both in content and in theoretical functionality self-efficacy (Bandura, 1977, 1997) and beliefs about ability in expectancy-value theory (Eccles-Parsons et al., 1983). To a lesser degree, it is related to attributions for success and failure (Weiner, 1985, 2005), beliefs about the nature of intelligence and ability (Dweck, 2002) and control beliefs (Connell, 1985; Skinner, 1995). Although early conceptualisations of self-concept considered it a monolithic construct (Coopersmith, 1967; Rosenberg, 1965), subsequent thought has stressed the domain-specificity of self-concepts (Harter, 1982; Marsh, 1990b; Shavelson, Hubner, & Stanton, 1976). Thus, contemporary theories of self-concept assume a hierarchical structure in which a global self-concept, also referred to as self-worth or self-esteem, supersedes more domain-specific self-concepts, e.g. self-concept of math ability (Harter, 2006; Marsh, Xu, & Martin, 2012). The conceptual core of these domain-specific self-concepts is formed by self-perceptions and self-evaluations of ability in those domains (Harter, 1982; Marsh, 1990b; Wigfield & Cambria, 2010; Wigfield et al., 2006). In particular, these evaluations are influenced by feedback from peers, parents and teachers as well as by intrapersonal perception of relative strengths (Marsh, 1986; Marsh et al., 2012; Shavelson et al., 1976).

Despite sharing the principle of a hierarchically structured self-concept, research traditions relevant to academic self-concept tend to differ in scope and focus of their investigations. Studies in the framework of the Marsh-Shavelson model of self-concept (Marsh & Shavelson, 1985; Shavelson et al., 1976) in educational psychology have concentrated on the dimensionality of the academic portion of self-concept and its complex relations to a variety of academic outcomes, most notably academic achievement. For instance, with regard to the structure of the academic portion of self-concept, it has been found that mathematical self-concept and verbal self-concept are almost uncorrelated though they are both related to global academic self-concept and though academic achievement in both domains is associated as well (Marsh & Hau, 2004). This observation inspired the idea that both interpersonal comparisons with the abilities of others and intrapersonal comparisons of capacities across different domains instigate the formation of domain-specific self-concepts (Marsh, 1986; Marsh et. al, 2012). Accordingly, high academic achievement in a particular domain has been shown to predict heightened self-concept in that domain, as a result of interpersonal comparisons, but comparatively lower self-concept in other domains, as a result of intrapersonal comparisons (Marsh, 2007; Marsh & Craven, 1997; Möller, Pohlmann, Köller, & Marsh, 2009; Möller, Retelsdorf, Köller, & Marsh, 2012). In addition to these processes, average ability in a given classroom is known to influence students' self-concept: Students in classrooms containing primarily students of low ability receive an additional augmentation of their self-concept as they can compare their achievement to an environment of generally low achievement. In contrast, students in classrooms consisting predominately of students of high ability experience an extra weight for the strength of their self-concept as they are forced to view their achievement against a background of generally high achievement. Commonly, this phenomenon is labelled the big-fish-little-pond effect (Chiu, 2012; Marsh & Hau, 2003; Marsh et al., 2008; Seaton, Marsh, & Craven, 2009).

Naturally, research in the tradition of developmental psychology has examined the self with a wider lens, aiming to understand the conditions for the formation of a coherent self-representation from early infancy to adulthood and exploring the consequences of maladaptive processes in the development of the self for psychopathology, such as depression or eating disorders (Harter, 1999, 2006). In this developmental tradition self-concept is viewed as both a cognitive and a social construction (Harter, 2006). Accordingly, children's increasing cognitive capabilities as well as social interactions with parents, caregivers, teachers and peers contribute to the formation of self-concept (e.g., Bretherton, & Munholland, 1999; Case, 1992; Higgins, 1991). Due to limitations of cognitive capacity, children in the preschool years have difficulties to integrate various domain-specific self-concepts of differing valence into a coherent global self-concept. In other words, it is difficult for young children to recognize that they have specific strengths and weaknesses (Fischer, Hand, Watson, Van Parys, & Tucker, 1984; Harter, 1986, 1990). Moreover, in adaptive interactions during the preschool years parents and caregivers tend to unconditionally support children's efforts to master the developmental tasks of infancy (Bretherton, & Munholland, 1999). As a consequence of this, when children enter elementary school, their self-concepts have a propensity to be holistic and unrealistically positive (Guay, Marsh, & Boivin, 2003; Harter, 1999; Harter & Pike, 1984). However, this form of self-concept must not necessarily be seen as deficient. Rather, it can be thought of as offering an unlimited supply of confidence for grasping persistently the daunting developmental tasks of early childhood (Harter, 2006). If a young toddler would judge realistically the current state of his or her abilities, would he or she ever dare to face the challenge of developing the capabilities of an adolescent, let alone an adult? In this vein, children appear to be limited in their capacity to distinguish between ability and effort all the way through elementary school (Stipek & MacIver, 1989). Nevertheless, the years of elementary school are a time that confronts children with more and more opportunities for comparison of their current achievement with their previous achievement or with the achievement of other children. Similarly, parents and teachers provide more and more feedback concerning the quality of children's achievement. This new abundance of possibilities for comparisons and feedback occurs during a time of rapid growth of cognitive capabilities so that children start relatively quickly to align their academic self-concept with external sources of information (Lipowsky, Kastens, Lotz, & Faust, 2011; Wigfield & Karpathian, 1991). By the end of elementary school students have become much more realistic with regard to their self-concept of ability (Helmke, 1999). In agreement with the outlined developmental principles, research in the framework of the Marsh-Shavelson model has established that the differentiation of the academic self-concept into distinct domain-specific self-concepts increases with students' age (Marsh & Ayotte, 2003; Marsh, Craven, & Debus, 1991, 1998). Moreover, academic self-concept has been found to decline steadily from early childhood, the start of formal schooling, to early adolescence (Alexander & Entwisle, 1988; Filipp, 2006; Guay et al., 2003; Marsh, 1989; Nagy, Watt, Eccles, Trautwein, Lüdtke, & Baumert, 2010; Wigfield et al., 1997).

Self-concept of ability and achievement. In research on the relation between self-concept and academic achievement scientific thought traditionally has revolved around the issue of the causal, in longitudinal studies operationalized as the temporal, order of the two factors, spawning two antagonistic views, the skill-enhancement approach and the skill-development approach (Calsyn & Kenny, 1977; Skaalvik, 1997). According to the skill-enhancement approach prior self-concept has an impact on subsequent achievement. Viewed in the light of expectancy-value models of achievement motivation, self-concept of ability is intimately intertwined with expectancy of success for a given task, thereby determining the probability of engaging in specific learning behaviors (Eccles & Wigfield, 1995; Eccles, Wigfield, Harold, & Blumenfeld, 1993; Wigfield et al., 2006). Moreover, it is plausible that a positive self-concept of ability hampers dysfunctional cognitions and behaviors associated with test anxiety (Bandalos, Yates, & Thorndike-Christ, 1995; Krampen, 1988; Urhahne, Chao, Florineth, Luttenberger, & Paechter, 2011). According to the skill-development approach prior achievement causes subsequent self-concept. In this perspective, achievement functions as a basis for intrapersonal and interpersonal comparisons of performance that inform the formation of subsequent self-concept (Möller & Köller, 2004; Pohlmann & Möller, 2009). Furthermore, teachers, parents and peers provide relevant feedback on competence that is based, at least, partially on actual achievement (Marsh et al., 2012). Thus, teacher ratings and teacher-assigned grades have sometimes been hypothesized to be better predictors of subsequent self-concept than standardized test scores (Hansford & Hattie, 1982; Marsh, 1990a).

In a synthesis of previous research contemporary theorizing construes the association between self-concept and academic achievement as reciprocal, with prior achievement influencing subsequent self-concept and prior self-concept influencing subsequent achievement (Guay et al., 2003; Huang, 2011; Marsh, 1990a; Marsh & Craven, 2006; Möller et al., 2011; Pinxten, de Fraine, van Damme, & d'Haenens, 2010). The strength of this relation is moderated by the domain-specificity of the self-concept assessed (Huang, 2011), i.e. the association between domain-specific self-concept and domain-specific achievement is stronger than the association between general academic self-concept and domain-specific achievement (Marsh & O'Mara, 2008; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006). In this context, cross-sectional correlations between domain-specific self-concepts and corresponding domain-specific achievement have been found to increase steadily from early on over the course of formal schooling, attaining an asymptotic level approximately in fifth grade (Denissen, Zarrett, & Eccles, 2007; Eccles et al., 1993; Fredricks & Eccles, 2002; Wigfield et al., 1997).

However, with regard to a clear developmental trajectory of the relative preponderance of processes of skill-enhancement or skill-development in the course of formal schooling, the available empirical evidence paints a complex and, when considered in detail, partially inconsistent picture (Marsh & Martin, 2011). Early in elementary school, i.e. in first grade, processes of skill-enhancement appear to take precedence over processes of skill-development (Bossaert, Doumen, Buyse, & Verschueren, 2011; Kammermeyer & Martschinke, 2006). Apparently, the overly optimistic self-concepts of early childhood continue to operate as a stimulating source of confidence for some time in elementary school before social comparisons and feedback from teachers relatively quickly start to induce a dominance of processes of skill-development (Harter, 2006; Kammermeyer & Martschinke, 2006; Lipowsky et al., 2011). In accordance with that, for the elementary school years from second to fourth grade, a relative preponderance, or even exclusive prevalence, of processes of skill-development has been found (Chapman & Tunmer, 1997; Helmke & van Aken, 1995; Muijs, 1997; Skaalvik & Hagtvegt, 1990; Skaalvik & Valas, 1999). The years of secondary education, however, appear to be accompanied by a progressive increase of the importance of processes of skill-enhancement (Chen, Yeh, Hwang, & Lin, 2013; Guay et al., 2003; Skaalvik & Hagtvegt, 1990). In agreement with this nonlinear pattern of developmental changes, available meta-analytic research has established either no systematic impact or a small negative influence of students' age at study inception on the relevance of processes of skill-enhancement, i.e. the size of the effect of prior self-concept on subsequent achievement (Huang, 2011; Valentine, DuBois, & Cooper, 2004). Similarly, meta-analytic research has also indentified a small negative influence of students' age at study inception on the importance of processes of skill-development, i.e. the size of the effect of prior achievement on subsequent self-concept (Huang, 2011).

Apart from individual developmental trends such as children's growing cognitive capabilities for coordination of feedback and self-representations (Wigfield & Karpathian, 1991), changes in the composition of reference groups are likely to contribute to an alteration of the predictive strength of self-concept for subsequent academic achievement in the transition from elementary to secondary education. Generally, the occurrence of school transitions tends to diminish the association between self-concept and subsequent achievement (Valentine, et al., 2004). In the transition to secondary education, the German educational system applies a particularly explicit form of ability grouping between different school types. Thus, due to the big-fish-little-pond effect, students of comparatively low ability that transfer to the school type of Hauptschule experience a significant boost of their academic self-concept. In contrast, students of relatively high ability that transfer to the school type of Gymnasium undergo a significant decline of their academic self-concept (Marsh & Hau, 2003; Valtin & Wagner, 2004; Watermann, Klingebiel, & Kurtz, 2010). As a consequence of this, self-concepts of students from the school types of Hauptschule and Gymnasium tend to resemble each other, despite apparent disparities in ability, in absolute strength, engendering, in comparison to the situation in elementary education, a reduction of the association between academic self-concept and subsequent academic achievement in secondary education (Kreppold, 2012).

Self-concept of ability and socio-economic status. An impact of family socio-economic status on children's domain-specific self-concepts can be construed both along an indirect and a direct pathway. On the one hand, developmental considerations and correlational evidence suggest that socio-economic status has an influence on general evaluations of the self (Harter, 2006; Twenge & Campbell, 2002). Thus, insofar as global self-concept, in turn, contributes to the formation of domain-specific self-concepts, it can be assumed that social background indirectly affects the development of domain-specific self-concepts. On the other hand, parental behaviors directly relevant for the formation of domain-specific self-concepts, such as providing feedback on academic achievement and school-related activities, are associated with family socio-economic status as well (Evans, 2004). In this vein, the impact of family socio-economic status on domain-specific self-concepts might be viewed, in a wide sense, as a form of intergenerational transmission of academic values and attitudes within families (Gniewosz & Noack, 2012). Finally, as the allocation of students to different tracks of German secondary education is intertwined in various ways with students' socio-economic status, the big-fish-little-pond effect appears as a plausible moderator of the association between social background and domain-specific self-concepts after the end of comprehensive education in elementary school (Maaz et al., 2009; Maaz et al., 2008).

Parent-child interactions characterized by approval, warmth and responsiveness represent a cornerstone of the development of an adaptive self-concept in early infancy and beyond (Bretherton, & Munholland, 1999; Coopersmith, 1967; Feiring & Taska, 1996). In this context, the co-construction of narratives of the child's past in conversation is of particular importance for the formation of autobiographical memory and the emergence of a positive representation of the self (Bretherton, 1993; Harter, 2006; Nelson, 1993; Snow, 1990). Successful realization of such parenting behavior is inversely related to family socio-economic status (Evans, 2004; McLoyd, 1998). Specifically, low socio-economic status has been found to be associated with comparatively harsh, punitive, unresponsive and authoritarian parenting (Bradley, Corwyn, McAdoo, & Garcia Coll, 2001; Dodge et al., 1994; Magnusson & Duncan, 2002). Likewise, parental speech in families of low socio-economic status has been shown to be impoverished, with regard to its amount, its responsiveness and its feasibility to initiate and sustain parent-child conversations (Hart & Risley, 1995; Hoff, 2003; Hoff et al., 2002). Moreover, children from families of low socio-economic status are more likely to be confronted with familial violence or temporary referral to foster care than their more affluent counterparts (Emery & Laumann-Billings, 1998; Evans, 2004; Rutter, 1981). Therefore, it appears plausible that children with an unfavorable social background enter elementary school with a less developed and a less positive overall self-concept than children with high family socio-economic status. This disparity in general self-concept is likely to have an impact on the formation of domain-specific academic self-concepts.

In addition to the preceding developmental considerations meta-analytic research has revealed that global self-concept is associated with socio-economic status across the entire life span. During the years of formal schooling the magnitude of this association increases constantly (Twenge & Campbell, 2002). However, the probability of the occurrence of top-down effects from global self-concept on domain-specific academic self-concepts appears to depend on the configuration of the learning environment: Investigating seventh-grade students from the school type of Gymnasium in East and West Germany, Trautwein, Lüdtke, Köller and Baumert (2006) have found top-down effects from global self-concept on academic self-concepts to be more pronounced in the ego-protective learning environments of West Germany than in the meritocratic learning environments of East Germany.

Parents, at least in western societies, are known to overestimate their children's academic abilities (Miller, Manhal, & Mee, 1991; Pezdek, Berry, & Renno, 2002; Stevenson & Stigler, 1992). In particular, parents tend to view their children's academic capabilities more favorably than teachers (Chamorro-Premuzic, Arteche, Furnham, & Trickot, 2009; Korat, 2011). Thus, with respect to the development of domain-specific academic self-concepts, parental feedback may serve as an antidote to unrelenting evaluations from teachers, supporting the formation and preservation of positive academic self-concepts (Gniewosz, 2010). Notably, parents of low socio-economic status appear to be less interested and less involved in activities related to their children's formal education than parents of high socio-economic status (Baker & Stevenson, 1986; Dumont, Trautwein, & Lüdtke, 2012; Evans, 2004; Stevenson & Baker, 1987). In extrapolation, this renders it plausible that parents of low socio-economic status provide positively biased evaluations of achievement only sparingly to their children, thereby exposing the development of their children's academic self-concepts to an almost exclusive dependence on other sources of information, such as teachers. The gravity of this potential cause for a disparate development of academic self-concepts might be even aggravated when teachers provide negatively biased feedback to students of low socio-economic status as it looms, for instance, in teacher recommendations for track allocation in German secondary education (Ditton et al., 2005; Schneider, 2011; Wong, 1980; see Jussim & Eccles, 1995; Karing, Matthäi, & Artelt, 2011, for evidence in favor of unbiased teacher perceptions of student ability).

Though the considerations outlined above point towards a cumulative growth of social disparities in domain-specific self-concepts, and thus towards an intensification of the association between family socio-economic status and domain-specific self-concepts in the course of formal schooling, in case of the German school system a pronounced big-fish-little-pond effect in secondary education, created by between-school ability grouping, tends to induce a disentanglement of socio-economic status and domain-specific self-concepts after fourth grade (Marsh et al., 2008; Marsh et al., 2012; Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006). Here a variant of the rationale depicted for the relation between domain-specific self-concept and academic achievement applies: Students of low socio-economic status are predominately referred to the lowest track of secondary education, partially due to low academic achievement, partially due to parental preferences and institutional practices of allocation (Maaz et al., 2009; Maaz et al., 2008). Within their new frame of reference they experience an enhancement of their academic self-concepts (Marsh & Hau, 2003; Marsh et al., 2012). In contrast, students of high socio-economic status are primarily enrolled in the highest track of secondary education (Maaz et al., 2009; Maaz et al., 2008). They suffer from a depression of their academic self-concepts due to their frame of reference (Marsh & Hau, 2003; Marsh et al., 2012). Accordingly, it has been suggested to view schools of the lowest track of German secondary education, i.e. of the school type of Hauptschule, as a self-worth protecting niche for students of low academic achievement (Baumert et al., 2006b; Köller & Baumert, 2001). In essence, students from the lowest and the highest track of German secondary education tend to display similar academic self-concepts despite apparent differences in average socio-economic status (Baumert et al., 2006b; Trautwein, Lüdtke, Marsh, et al., 2006). This causes an attrition of the effects of socio-economic status on domain-specific self-concepts in German secondary education as compared to the state of affairs in elementary education.

Summary. In the context of learning in school, domain-specific self-concepts of ability represent important motivational constructs, both as outcomes in their own right and as precursors of domain-specific academic achievement (Marsh et al., 2010). At the beginning of formal education academic self-concepts are rather undifferentiated and overly optimistic (Guay et al., 2003; Harter, 1999). Throughout the years of elementary school the positivity of academic self-concepts tends to decline. Thereby self-concepts of ability are rendered progressively more realistic (Helmke, 1999).

Due to a preponderance of processes of skill-development from second grade onwards to the end of elementary school, domain-specific self-concepts figure as weak predictors of domain-specific academic achievement in middle and late childhood (Helmke & van Aken, 1995; Kammermeyer & Martschinke, 2006). With an increasing relevance of processes of skill-enhancement in secondary education, however, domain-specific self-concepts tend to become comparatively important precursors of domain-specific academic achievement in adolescence (Chen et al., 2013). Similarly, under the assumption of a cumulative growth of the effects of maladaptive parental behavior on the formation of global self-concept and domain-specific self-concepts, the association between family socio-economic status and domain-specific self-concepts is prone to intensify with students' age, and thus also across the transition from elementary to secondary education (Evans, 2004; McLoyd, 1998). In addition, theoretically, this relation should be moderated by characteristics of learning environments (Trautwein et al., 2006). It should be specifically pronounced in ego-protective learning environments, such as the comprehensive school type of Grundschule in elementary education and the school type of Hauptschule, which has a remedial mission, in secondary education.

However, these general developmental trends in the transition from elementary to secondary education are overshadowed by the consequences of a strong big-fish-little-pond effect induced by between-school ability grouping in German secondary education. Creating student groups of similar domain-specific self-concept that differ widely with regard to domain-specific academic achievement and family socio-economic status (Baumert et al., 2006b; Maaz et al., 2008; Trautwein, Lüdtke, Marsh, et al., 2006), the big-fish-little-pond effect diminishes the associations of academic self-concepts with these two variables in secondary education (Kreppold, 2012). This implies a reduction of the relevance of domain-specific self-concepts as mediators between family socio-economic status and academic achievement.

Individual Interest

What is interest? From a very general point of view, interest belongs to that class of motivational constructs that capture students' eagerness and willingness ("Do I want to do this task?") to engage in specific tasks and to devote themselves to specific academic domains (Wigfield et al., 2006). In particular, with regard to its overall emotional tone and its behavioral consequences, interest is related to the concepts of intrinsic motivation (Deci & Ryan, 1985, 2002; Gottfried, Fleming, & Gottfried, 2001) and mastery orientation towards learning (e.g., Ames, 1992; Blumenfeld, 1992); it also possesses partial overlap with the construct of task value from expectancy-value theory (Wigfield & Cambria, 2010; Wigfield et al., 2006). However, the distinguishing feature of interest is its content-specificity (Schiefele, Krapp, Prenzel, Heiland, & Kasten, 1983). Interests are always directed towards specific objects, activities, topics or academic subjects (Krapp, 2002a).

Resorting to the categorization of different forms of interest according to their permanentness and stability, similar to the distinction between states and traits, two conceptualizations of interest have evolved, namely situational and individual or personal interest (e.g., Hidi & Renninger, 2006; Krapp, 2002b; Krapp, Hidi, & Renninger, 1992; Schiefele, 2009). On the one hand, situational interest is seen as a transient psychological state of positive emotion, focused attention, improved cognitive functioning and perseverance (Hidi, 2000), in its most extreme forms blurring the border between self and activity in experiences of flow (Csikszentmihalyi, 1975). For instance, it has been investigated intensively which features of texts elicit temporary states of situational interest (Schiefele, 1996). On the other hand, individual interest is conceptualized as a stable disposition to occupy oneself with a specific topic or object, e.g. an academic discipline (Renninger, Ewen, & Lasher, 2002; Schiefele, 2009). The dispositional construct of individual interest consists of several subcomponents, most notably a feeling-related component and a value-related component. The former covers positive emotions associated with a specific object of interest, such as enjoyment of engagement with a specific academic discipline, whereas the latter comprises evaluations of the importance of a specific object of interest for the self (Krapp, 2002a; Schiefele, 2001). Especially its value-related component underscores that individual interest is conceptualized as an integral part of personality constituting a stable relation of a person to a specific object of interest.

The person-object theory of interest unifies these two forms of interest in a common framework (Krapp, 1998, 2002a, 2002b). Specifically, the person-object theory of interest construes the arousal, or actualization, of situational interest as resulting from characteristics of both the individual person and the specific situation. With regard to learning in school, for instance, a student may experience a state of situational interest in a comparatively tedious science lesson due to a pre-existing pronounced individual interest in science. In other words, he or she actualizes his or her dispositional interest in a given situation. On the contrary, a science lesson that features exciting experiments and points out the relevance of science content for students' personal lives may initiate and maintain situational interest in students initially not possessing a specific individual interest in science (Boekaerts, 1999; Häußler & Hoffmann, 2002; Mitchell, 1993). Apart from the short-term influence of individual interest on the emergence of episodes of situational interest, the person-object theory of interest also postulates a long-term impact of experiences of situational interest on the formation of individual interest: The repeated experience of transient episodes of situational interest for a specific object is thought to transform into an enduring individual interest for that object, e.g. the academic domain of science (Krapp, 1998, 2002b). This internalization of interest is conceptualized to traverse multiple stages from a situational interest aroused primarily by external stimuli to a stable individual interest integrated firmly into the self (Hidi & Renninger, 2006; Mitchell, 1993).

Students' individual interest in academic activities and school subjects undergoes a substantial decline in the course of formal schooling (e.g., Frenzel et al., 2010; Gottfried et al., 2001; Krapp, 1998). This decline is particularly profound for the domain of science (Hoffmann, Häußler, & Lehrke, 1998; Hoffmann, Krapp, Renninger, & Baumert, 1998). Moreover, the decrease of interest in school subjects appears to be accompanied by qualitative changes in the internal structure of the construct of individual interest: At younger ages the affective tone of the feeling-related component dominates individual interest, whereas later on the cognitive orientation of the value-related component takes precedence (Frenzel, Pekrun, Dicke, & Goetz, 2012; Krapp, 2002b; Todt, 1990). A variety of developmental mechanisms and contextual factors have been forwarded as plausible explanations for the group-average decline of individual interest in academic domains across the years of schooling. For instance, the group-average decrease of academic interest might be understood as a natural corollary of the onset of puberty (Crouter, Manke, & McHale, 1995) or an inevitable reduction of the numerous interests of early childhood (Krapp, 2002b; Todt, 1990). In addition, specifically in the transition from elementary to secondary education, changes in school and classroom environments, such as an increasing performance orientation or ability grouping, might contribute to a decline of individual interest in academic activities (Eccles & Roeser, 2011; Frenzel et al., 2010; Wigfield et al., 2006).

The construct of interest can be linked to the frameworks of self-determination theory and modern expectancy-value models of achievement motivation (Daniels, 2008; Wigfield & Cambria, 2010). Interest-driven behaviors – such as reading about a topic one is interested in – are self-intentional, i.e. they are performed independently of external incentives (Prenzel, Lankes, & Minsel, 2000; Schiefele, 1996). In particular, the execution of interest-driven behaviors stands in complacent congruence with the developing self (Krapp, 1998, 2002b). Thus, the person-object theory of interest and self-determination theory congenially interpreting individual human development as a dynamic strive for an increasing integration of the self, the construct of interest resembles the concept of intrinsic motivation (Krapp, 2002b; Ryan & Deci, 2009). In agreement with these considerations, emotional experiences related to the fulfillment of the three basic needs postulated in self-determination theory, the need for competence, the need for autonomy and the need for social relatedness, are seen as essential for the generation of interest (Deci & Ryan, 2000; Krapp, 2005). Unsurprisingly, conceptualizations of individual interest in specific school subjects and conceptualizations of academic intrinsic motivation for specific school subjects display a vast overlap in item content (e.g., Frenzel et al., 2010; Gottfried, 1985; Gottfried et al., 2001).

According to contemporary expectancy-value models of achievement motivation the probability or intensity with which a student engages in a specific learning task depends on the interaction of the expectancy of success for that task with the task value ascribed to that task (Eccles-Parsons et al., 1983; Wigfield & Cambria, 2010; Wigfield et al., 2006). The ascription of task value, in turn, decomposes into the attribution of attainment value, interest value, utility value, i.e. the usefulness of the task for future goals, and cost (Wigfield & Eccles, 1992). Attainment value is concerned with the importance that is placed upon doing well in a given task (Eccles-Parsons et al., 1983). It depends on the centrality of certain tasks for the self (Wigfield & Cambria, 2010). Interest value captures the enjoyment and positive affect that can be gleaned from engaging in a given task (Wigfield & Cambria, 2010). Thus, the concepts of attainment value and interest value can be roughly equated with the value-related and feeling-related component, respectively, of individual interest.

Nevertheless, the construct of interest differs in important ways from intrinsic motivation, attainment value, and intrinsic value (see Daniels, 2008; Wigfield & Cambria, 2010). First of all, interest is conceptualized as a decidedly content-specific construct. So, for instance, the particularly pronounced group-average decline in individual interest for the school subject of physics has been explained by characteristics of the content to be learned, such as abstractness and limited relevance for everyday life (Hoffmann, Häußler, et al., 1998). Secondly, the person-object theory of interest clearly distinguishes and integrates the isomorphic concepts of temporary situational interest as a state and enduring individual interest as a disposition in a common framework. In contrast, focusing on the explanation of engagement in specific tasks, expectancy-value models remain comparatively vague with regard to the nature and relevance of dispositional forms of attainment value and intrinsic value (Eccles & Wigfield, 2002). Similarly, intrinsic motivation is described primarily as a state resulting from the fulfillment of basic needs, whereas conceptualizations of enduring intrinsic motivational orientation are comparatively rare (Amabile, Hill, Hennessey, & Tighe, 1994; Gottfried, 1985; Gottfried et al., 2001; see also Deci & Ryan, 1985, for the concept of dispositional causality orientations as precursors of differential intrinsic motivation).

Interest and achievement. Interest is positively related to learning and academic achievement (Schiefele, 1996, 1998, 2009; Schiefele, Krapp, & Winteler, 1992; see Ainley & Ainley, 2011; Lavonen & Laaksonen, 2009 for recent findings in the domain of science). In the context of contemporary expectancy-value theory of motivation, individual interest can be thought of as a dispositional precursor of task value. Precisely, individual interest is likely to influence attainment value and interest value for specific tasks in that domain, because of its value-related component and feeling-related component, respectively (see also Wigfield & Cambria, 2010; Wigfield et al., 2006). Thus, with task value determining in interaction with expectancy for success choices in favor of specific academic behaviors and persistence in specific academic tasks, expectancy-value theory can be interpreted to predict an indirect effect of individual interest on academic performance (cf. Wigfield et al., 2006). Note, however, that in empirical research attainment value and interest value are often subsumed under the more general construct of global task value (e.g., Eccles et al., 1993; Soric & Palekcic, 2009).

Within the framework of the person-object theory of interest, domain-specific knowledge acquisition is an inherent companion of domain-specific interest. As the establishment and maturation of an individual interest is accompanied by devotion to the corresponding object of interest, the integration of the object of interest into the self is inevitably tied to knowledge growth (Krapp, 2002b). In agreement with this, it has been shown that intraindividual consistency between individual interests and knowledge in various domains increases over the life span (Reeve & Hakel, 2000). It has even been suggested to view extended domain-specific knowledge as an integral component and indicator of individual interest in that domain (Renninger, 2000; Renninger et al., 2002; see Schiefele, 2009, for critical discussion). In this respect, an individual interest in a given domain that does not entail a corresponding growth in domain-specific knowledge has to be considered an interest that has cooled off, a henceforth gradually fading component of personality once deemed important to the self (Krapp, 2002b).

On a short time scale, usually exemplified in text-based studies, interest exerts an enhancing influence on learning and achievement by means of positive affective states, increased persistence and focused attention (Ainley, Hidi, & Berndorff, 2002; Hidi, 1995; Hidi, Renninger, & Krapp, 2004; Schiefele, 1996). In addition, in text-based learning, interest appears to be especially related to deep-level text comprehension, as indicated by inferences from text, conceptual understanding and transfer (Andre & Windschitl, 2003; Kunz, Drewniak, Hatalak, & Schön, 1992; Schiefele, 1990, 2009; Schiefele & Krapp, 1996). On a longer time scale, individual interest has been found to affect academic achievement via a mastery orientation towards learning (Brett, Wilkins, Long, & Wang, 2012), effective self-regulation (Metallidou & Vlachou, 2010), application of advanced learning strategies (Soric & Palekcic, 2009), time spent on homework (Singh, Granville, & Dika, 2002) and course selection (Köller, Baumert, & Schnabel, 2001).

With regard to the development of the association between individual interest and academic achievement in a given domain, thoughtful consideration of the ramifications of the person-object theory of interest, of the consequences of ability grouping and of growth of personal independence in adolescence leads to the prediction that the relation between interest and achievement increases in the transition from elementary to secondary education. The first argument in favor of an intensification of the association between interest and achievement rests on two aspects of the person-object theory of interest. First, the development of individual interests starting with undifferentiated interest profiles in early childhood, young children do not differ much with respect to interest strength in various academic domains. In contrast, students at the end of their formal education display very distinct interest profiles. In other words, at the end of school, there are large discrepancies in individual interest strength in the various academic domains between different students (Krapp, 1998; Todt, 1990). Second, with the human self autonomously striving towards progressive integration in personality development, the maturation and growth of individual interests is intimately related to acquisition of new knowledge in domains pertinent to these interests (Krapp, 2002b). Thus, taken together, these two principles of the person-object theory of interest imply an intensification of the association between interest and achievement over the years of formal schooling as students tend to increasingly drift apart with respect to individual interest and academic achievement in a given domain. In partial support of this line of reasoning, the link between interest and achievement has been found to be particularly strong for secondary science education (Hoffmann, Häußler, et al., 1998; Krapp & Prenzel, 2011; Schiefele, et al., 1992), presumably because here, where the group-average decline of individual interest during the school years is outstandingly pronounced (Baumert & Köller, 1998; Daniels, 2008; Osborne, Simon, & Collins, 2003), the discrepancies between a minority of students with high interest as well as high achievement in science and a majority of students with low interest as well as low achievement in science are becoming exceptionally large.

For appraisal of the implications of tracking for the relation between interest and achievement in a given domain, the following gedankenexperiment is instructive: Clearly, the pedagogical intention of ability grouping is to create learning environments that are homogeneous with regard to the cognitive capacities of students, i.e. students of comparable intelligence and prior knowledge are grouped together. This is supposed to ease the task of teaching (Baumert et al., 2006b; see also LeTendre et al., 2003). Accordingly, in thought taken to the extreme, optimal ability grouping would result in learning environments consisting of students that are identical with respect to their cognitive capacities. It is not difficult to imagine that in such a situation, where there is no variance in students' cognitive capabilities in given learning environments, non-cognitive factors like individual interest should take, beyond group membership, precedence in the prediction of individual differences in academic achievement. Apparently, the result of this small gedankenexperiment renders a contribution of ability grouping to an intensification of the relation between individual interest and academic achievement in the transition from elementary to secondary education plausible. However, the potential size of this contribution in real-world settings depends on the degree to that the actual practice of tracking approaches the optimal ability grouping outlined above. In this respect, the allocation of students to the different tracks of German secondary education is far from operating in an indisputable manner (e.g., Ditton et al., 2005; Tiedemann & Billmann-Mahecha, 2010). As a consequence of this, the distributions of students' cognitive capabilities and achievement for the various school types of German secondary education show, despite differences in the respective means, considerable overlap (Baumert et al., 1997, 2006b; Helmke & Jäger, 2002; Lehmann et al., 1997).

Last but not least, the transition from elementary to secondary school and a possible alteration of the association between individual interest and academic achievement must be seen in the light of the more general transition from childhood to adolescence (Fend, 2000; Hasselhorn & Silbereisen, 2008; Silbereisen & Todt, 1994). During the period of childhood, parents serve as the central significant others. They closely monitor their children's activities. By and large, parental authority remains unquestioned. In contrast, confronted with various developmental task that must be coped with in the preparation for life as an adult, adolescents strive for growing independence from parents and teachers (Collins, 1990; Havighurst, 1972; Lam, McHale, & Crouter, 2012; Pettit, Keiley, Laird, Bates, & Dodge, 2007). The importance of interactions with peers in the course of socialization increases (Eccles, Wigfield, Flanagan, Miller, Reuman, & Yee, 1989; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991). Naturally, the process of gaining independence is accompanied by an enrichment of available activity choices. For instance, whereas children in elementary school are inclined to fulfill homework assignments simply because teachers and parents tell them to, adolescents in secondary school recognize the possibility to skip homework assignments in favor of leisure activities with peers (Hofer, Schmid, Fries, Dietz, Clausen, & Reinders, 2007; Hofer, 2010; Schnyder, Niggli, Cathomas, Trautwein, & Lüdtke, 2006; Trautwein, Lüdtke, Kastens, & Köller, 2006). Accordingly, in general terms, it is plausible that increasing freedom with respect to activity choices gives way to an increasing importance of individual interest as a predictor of academic achievement in adolescence.

Interest and socio-economic status. According to contemporary theorizing about the formation of individual interests repeated episodes of situational interest with respect to a certain object of interest precede the internalization of a corresponding individual interest (Hidi & Renninger, 2006; Krapp, 2002b). Thereby, the fulfillment of basic needs in the sense of self-determination theory is a crucial prerequisite for experiences of situational interest (Deci & Ryan, 2000; Krapp, 2002a, 2005). Insofar as children's experiences with episodes of situational interest relevant to the various academic domains vary systematically with family socio-economic status, children's social background has to be assumed to shape the development of individual interests in the realm of academic domains. Thus, material and financial resources, the quality of early childcare, as well as parental interests, expectations and values, which all vary with family socio-economic status, appear as potentially limiting factors for the provision of opportunities to experience situational interest in academic domains (Bradley & Corwyn, 2002; Evans, 2004; McLoyd, 1998). However, as the internalization of interests depends on the fulfillment of basic needs, the effect of the provision of opportunities is probably moderated by the quality of the parent-child relation, parenting practices, and the quality of parental involvement in children's education (Fan & Chen, 2001; Gottfried, Fleming, & Gottfried, 1994), which all also tend to vary systematically with family socio-economic status (Bolger, Patterson, Thompson, & Kupersmidt, 1995; Duncan & Brooks-Gunn, 1997; McLoyd, 1990). In other words, the ability to provide opportunities for episodes of situational interest represents a necessary condition for the development of enduring academic interests, which needs to be complemented by adequate parent-child interactions; both provision of opportunities and parenting practices are associated with family socio-economic status (see Figure 1). In a broader context, these mechanisms may be viewed as contributing to the transfer of social identities from parents to children.

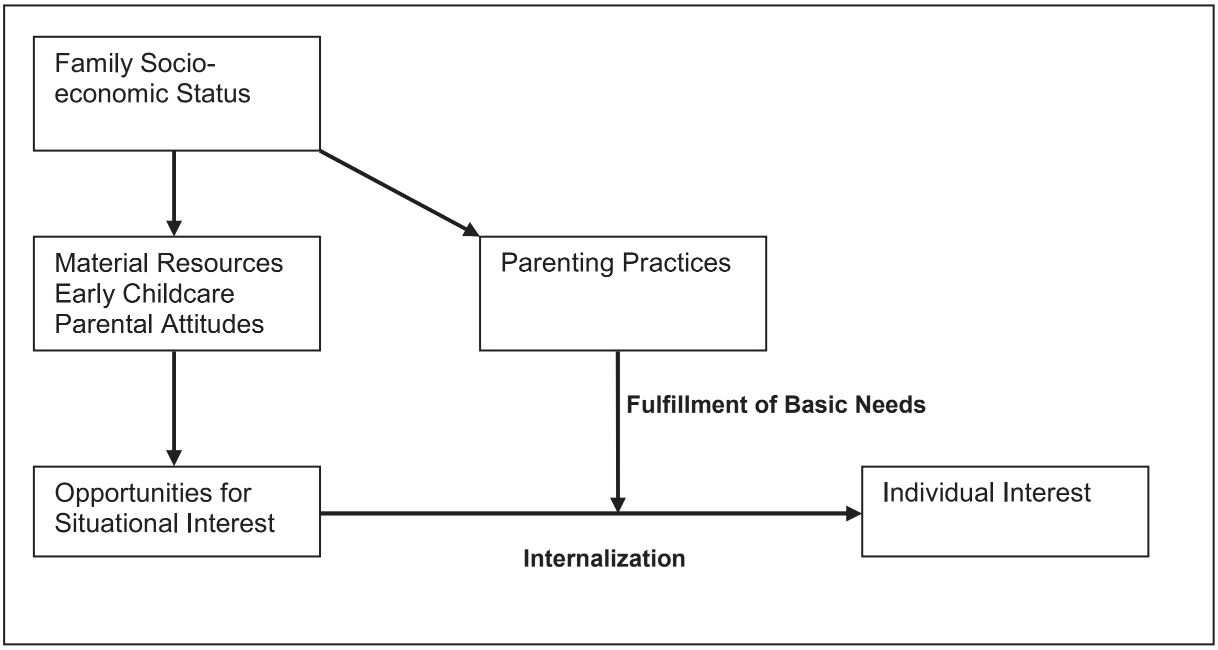

Figure 1. Tentative model of the impact of family socio-economic status on the formation of individual interest in academic activities.

Exemplifying the preceding considerations for the domain of science, longitudinal studies have revealed that parental behaviors supportive of science activities, including the purchase of science items and games, foster children's individual interest in science (Jacobs, Davis-Kean, Bleeker, Eccles, & Malanchuk, 2005). Obviously, families of high socio-economic status dispose of the financial resources necessary to invest in science items and games; they can afford to pay for visits to libraries, zoos and museums (Bradley & Corwyn, 2002; Bradley, Corwyn, Burchinal, et al., 2001; Bradley, Corwyn, McAdoo, et al., 2001). Similarly, it has been shown that parents of low socio-economic status tend to refrain from enrolling their children in non-compulsory preschool, depriving them of participation in early science education (Kratzmann & Schneider, 2009). In the same vein, coinciding with high socio-economic status (Ehmke, 2010), positive parental attitudes towards science stimulate children's interest in science (Jacobs & Bleeker, 2004; Turner, Steward, & Lapan, 2004). Obviously, the occurrence of opportunities for children to experience situational interest with respect to science appears to be constrained by family socio-economic status. In addition, parenting practices associated with low socio-economic status are prone to defy the fulfillment of basic needs and thus to hinder the formation of enduring individual interest in science. Clearly, punitive, harsh and unresponsive parenting, which is more common among families of low socio-economic status than among families of high socio-economic status, stands in opposition to children's need for social relatedness (Bradley, Corwyn, McAdoo, et al., 2001; Dodge et al., 1994). Generally, poverty is associated with impoverished social support (Atkinson, Liem, & Liem, 1986; Cochran, Larner, Riley, Gunnarson, & Henderson, 1990; Conger & Elder, 1994). Furthermore, parent-child interactions in families of low socio-economic status are relatively directive (Hart & Risley, 1995). Parents of low socio-economic status encourage conformity in their children (Evans, 2004; Kohn, 1977; Luster, Rhoades, & Haas, 1989). These characteristics of socially disadvantaged families probably deny children's need for autonomy (Adams, 1998). Last but not least, the occurrence of learning interactions that are suitable to fulfill children's need for competence with regard to academic activities is inversely related to family socio-economic status (Borduin & Henggeler, 1981; Bradley & Corwyn, 1999). So, in summary, parenting practices prevalent in families of low socio-economic status are likely to hinder the fulfillment of basic needs essential for the internalization of enduring individual interests.

At first glance, under the assumption of cumulative effects of social background on the provision of opportunities for episodes of situational interest and on the fulfillment of basic needs, it appears plausible that the association between family socio-economic status and students' individual interest for academic activities intensifies across the years of formal schooling. The seasonality of social disparities in achievement gains lends additional verisimilitude to the hypothesis of an increasing association between socio-economic status and individual interest (Alexander, Entwisle, & Olson, 2001). The relation between socio-economic status and school failure, a particular severe form of losing interest in academic activities, also points in this direction (Battin-Pearson et al., 2000). Moreover; the growth of social disparities in students' interest might be further amplified in decidedly tracked school systems as tracking creates groups of students with comparable socio-economic backgrounds and thereby presumably aligns the directionality of parental and peer influences on interest development. In fact, the sustainment of high-achieving students' interest is conventionally cited as one of the pedagogical intentions of ability grouping (Plucker et al., 2004). In spite of this, recently researchers have found evidence in favor of a big-fish-little-pond effect with regard to students' individual interest in school subjects (Frenzel et al., 2010; Trautwein et al., 2006). However, it is unclear whether the occurrence of this big-fish-little-pond effect is apt to counteract an assumed intensification of social disparities in students' interest.

Summary. Capturing students' eagerness to devote their time and energy to the study of certain domain-specific topics, individual interest in academic disciplines serves as an important motivational construct for learning in school (Krapp, 1998). Across the years of formal schooling, individual interest in academic disciplines undergoes a substantial decline (Frenzel et al., 2010; Gottfried et al., 2001). This decline is particularly profound for the domain of science (Hoffmann et al., 1998). Moreover, this decrease appears to be accompanied by a recession of the relevance of the feeling-related component and a parallel growth of the importance of the value-related component for the construct of individual interest (Frenzel et al., 2012; Krapp, 2002b). At least partly, the group-average decline of individual interest in academic disciplines has to be understood as a natural corollary of the intraindividual reduction of the various interests of early childhood to a few highly valued interests in late adolescence (Krapp, 2002b; Todt, 1990).

The person-object theory of interest assumes that the formation and existence of domain-specific individual interest is inherently tied to the growth of corresponding domain-specific knowledge (Krapp, 2002b). In fact, various forms of interest have been shown to enhance learning, in particular deep-level learning (Schiefele, 2009). The fundamental principles of the person-object theory of interest, speculations about the implications of ability grouping and the developmental increase of adolescents' independence point towards an intensification of the relation between individual interest and academic achievement in the transition from elementary to secondary education (Hasselhorn & Silbereisen, 2008; Krapp, 2002b). In spite of this, at the lower level of German secondary education empirically individual interest has been found to be a comparatively weak or even a nonsignificant predictor of academic achievement (Baumert & Köller, 1998; Köller et al., 2001; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Schiefele, 2009).

Family socio-economic status influences the formation of students' individual interest in specific academic disciplines probably both via investment in the provision of opportunities for experiences of corresponding situational interest and via the fulfillment of basic needs (Duncan & Brooks-Gunn, 1997; Jacobs et al., 2005). Under the assumption that these influences remain relatively stable across the years of formal schooling, it appears likely that cumulative effects engender a developmental intensification of the association between social background and individual interest in academic activities (Alexander et al., 2001). Recently, researchers have found evidence in favor of a big-fish-little-pond effect with regard to individual interest in school subjects (Frenzel et al., 2010; Trautwein et al., 2006). Yet, it is difficult to gauge if the implications of this big-fish-little-pond effect tend to dilute the developmentally plausible amplification of the relations of socio-economic status with academic achievement and with individual interest.

Research Questions

In general, the research questions for the present investigations had a two-layered structure. The first layer of this structure was concerned with the basic pattern of a potential mediation of the effects of socio-economic status on student achievement by cognitive and motivational propensities. Accordingly, this first layer covered general questions with regard to the associations between (a) socio-economic status and student achievement, (b) socio-economic status and cognitive and motivational propensities, and (c) cognitive and motivational propensities and student achievement. The second layer of the structure of research questions was concerned with the potential moderation of the aforementioned set of relations by students' grade level or enrollment in a specific school type. In other words, it covered the differential question whether a configuration of results obtained for an entire sample held for relevant subsamples as well. Obviously, this second layer of the structure of research questions was inspired by hypotheses with regard to the impact of developmental and institutional factors on the associations assembled in the first layer of the structure of research questions.

Moreover, in examining these associations and their moderations, due to the peculiarities of the data set used, the present investigations differed in important ways from most previous research. First, recent research on the effects of family socio-economic status on student achievement in Germany has operationalized student achievement broadly in terms of competences in specific domains (e.g., Bonsen et al., 2008; Ehmke & Jude, 2010; Stubbe et al., 2012). In contrast, in the context of the present investigations student achievement was assessed narrowly, in terms of post-instructional science achievement with regard to the topic of evaporation and condensation. Second, in field studies of the relation between social background and student achievement spanning several grade levels or school types, naturally, the content taught could not be held constant across different grade levels or school types (e.g., Ehmke et al., 2006; Opdenakker, van Damme, de Fraine, van Landeghem, & Onghena, 2002; Scott, Rock, Pollack, Ingels, & Quinn, 1995). In contrast, for the current investigations teachers from different grade levels and school types were instructed alike to provide their students with a series of lessons on the topic of evaporation and condensation (see the description of the data set below).

The first aim of the study was to establish a social gradient in science achievement for the data set at hand. Obviously, the presence of such a gradient constituted a prerequisite for the subsequent examination of the mediation of effects of socio-economic status by cognitive and motivational propensities. In sum, meta-analytic research as well as international and nationwide large-scale assessments of student achievement hint at a moderate to strong association between socio-economic status and science achievement (e.g., Bonsen et al., 2008; Hattie, 2009; Sirin, 2005; Stubbe et al., 2012; White, 1982). However, as mentioned above, these studies conceptualized science achievement predominately in broad terms as science competence covering a range of varying topics. Thus, assuming that influences of family socio-economic status on student achievement accumulate over time and over different topics, a comparatively small relation between the two variables was expected for the present data set. This expectation was located on the first layer of the structure of research questions.

Furthermore, a pair of tentative hypotheses was formulated with regard to the potential moderation of the association between socio-economic status and science achievement by grade level or school type. The accumulation of primary and secondary effects of social origin has been shown to intensify the relation between socio-economic status and student achievement in German secondary education relative to the corresponding association in elementary education (Maaz et al., 2009; Maaz & Nagy, 2009; Maaz et al., 2008; Schnabel et al., 2002). The extrapolation of this finding to the current investigations suggested an amplification of the association between socio-economic status and science achievement with an increase in students' grade level. However, as the learning sequence under consideration was limited to a specific topic and as the subsamples of fourth- and sixth-graders were positioned closely to the transition from elementary to secondary education, this extrapolation of the relevance of cumulative processes of competence acquisition bore considerable uncertainty. The aforementioned steepening of the social gradient in student achievement in German secondary education is attributable, at least to a large portion, to the differential enrollment of students of low and high family socio-economic status in the school types of secondary education and the differential development of student achievement between these school types, while there occurs little or no differential development of achievement along the lines of social background within school types (Baumert et al., 2006b; Schnabel et al., 2002; Baumert et al., 2012). Accordingly, for the current examinations, restrictions of variances in socio-economic status and achievement within school types were suspected to possibly engender a reduction of the association between social background and student achievement within specific school types of secondary education as compared to the states of affairs in the comprehensive school type of Grundschule of elementary education. However, this line of reasoning depended crucially on the occurrence of relevant restrictions of variances within the school types of secondary education. These hypotheses concerning the importance of students' grade level and school type for the association between socio-economic status and science achievement were located at the second layer of the structure of research questions.

The second aim of the present examinations was to elucidate the role of fluid ability as a potential mediator of the effects of socio-economic status on science achievement. On the first layer of the structure of research questions, this covered, as a necessary initial step, the inspection of the social gradient in fluid ability. Under the assumption that family socio-economic status affects the development of children's intelligence both directly and indirectly via the endowment of home environments and the heredity of parental intelligence (Bradley & Corwyn, 2002; Farah & Hackman, 2012), for the current data set the association between students' fluid ability and post-instructional science achievement was hypothesized to be of intermediate strength. Moreover, being a natural precursor of knowledge acquisition, fluid ability was expected to be a substantial predictor of post-instructional science achievement. Thus, in conjunction, fluid ability was anticipated to mediate a relevant portion of the effect of socio-economic status on science achievement. On the second layer of the structure of research questions, the association between social background and fluid ability for sixth-grade students was assumed to exceed the corresponding association for fourth-grade students. In other words, the relation between socio-economic status and fluid ability was expected to amplify in the transition from elementary to secondary education. This hypothesis was based on the presumption of an accumulation of adverse effects of socio-economic status on students' intelligence development during childhood and early adolescence. For this relation there were no further differential hypotheses with regard to students' enrollment in the school types of Hauptschule or Gymnasium in secondary education. The relation between fluid ability and post-instructional science achievement was expected to attenuate with the transfer from elementary to secondary education. This hypothesis followed the idea that with growing complexity of students' knowledge and of the content to be learned students' domain-specific prior knowledge replaces general mental ability as the central predictor of subsequent student achievement (Baumert et al., 2009). Thus, the association between fluid ability and science achievement was assumed to attenuate both for the entire subsample of sixth-graders and within the school types of Hauptschule and Gymnasium in secondary education. Note that if the contention that the reduction of the strength of the relation between fluid ability and student achievement was attributable mainly to reductions of variances in mental ability in specific learning environments was correct (see Jensen, 1998, for this argument), an attenuation of the association between fluid ability and post-instructional science achievement would be observable only within in the school types of Hauptschule and Gymnasium, and not for the entire subsample of sixth-graders.

The third aim of the current study was the inspection of the contribution of prior knowledge to the mediation of effects of social background on science achievement. The financial capacity to provide offspring with opportunities to attain academically relevant prior knowledge and parental behaviors fostering the acquisition of such knowledge distributed differentially along the lines of the socio-economic status, a relatively close relation between family socio-economic status and students' topic-specific prior knowledge was expected for the data set at hand. Moreover, in line with a massive bulk of research findings, students' prior knowledge was assumed to be a powerful predictor of post-instructional achievement (Hattie, 2009; Renkl, 1996). So, prior knowledge was supposed to be an important mediator of the effects of socio-economic status on post-instructional science achievement. Obviously, these hypotheses were located at the first layer of the structure of research questions.

On the second layer of the structure of research questions, the association between family socio-economic status and students' topic-specific prior knowledge was surmised to intensify in the transition from elementary to secondary education. Here, the same rationale as formulated above for the potential moderation of the social gradient in post-instructional science achievement by grade level came into effect: An accumulation of primary and secondary effects of social origin should entail an amplification of the relation between socio-economic status and prior knowledge in secondary education. Observations of increasing achievement gaps between students of low and high socio-economic status during summer break lend additional credibility to the assumption of an accumulation of adverse effects of social background on student achievement over the course of formal schooling (Alexander et al., 2001, 2007). However, it was not possible to devise further hypotheses for a differentiated moderation of the social gradient in topic-specific prior knowledge by students' enrollment in specific school types of secondary education. Complementary to the differential assumptions formed for the relevance of fluid ability as a predictor of science achievement, the relation between topic-specific prior knowledge and post-instructional science achievement was expected to intensify with students' grade level (Baumert et al., 2009). Moreover, the contention that the cumulatively growing complexity of students' knowledge and of the content to be learned renders domain-specific prior knowledge increasingly the central predictor of achievement spawned additional differential hypotheses with regard to the school types of secondary education: Students attending the school type of Hauptschule tend to possess considerably less prior knowledge than students attending the school type of Gymnasium, an inevitable corollary of ability grouping. Accordingly, the strength of prior knowledge as a predictor of achievement should be more pronounced within the school type of Gymnasium than within the school type of Hauptschule.

The fourth aim of the present investigations was the exploration of students' self-concept of ability as a potential mediator of the relation between family socio-economic status and students' science achievement. Naturally, this involved the testing of assumptions about the association between socio-economic status and self-concept of ability in science: Plausibly, both depressing top-down effects from global self-concept or self-esteem (Trautwein et al., 2006; Twenge & Campbell, 2002) and impoverished parental feedback (Baker & Stevenson, 1986; Dumont et al., 2012) cause children of low socio-economic status to form comparatively negative domain-specific self-concepts of ability, thereby constituting a social gradient in students' self-concept of ability science. Apart from that, domain-specific self-concepts of ability have the capacity to enhance domain-specific academic achievement by fostering engagement in learning behaviors (Eccles & Wigfield, 1995) and by hampering dysfunctional cognitions in learning and test situations (Urhahne et al., 2011). However, in face of the findings in support of a dominance of processes of skill-development in middle and late childhood (Helmke & van Aken, 1995; Kammermeyer & Martschinke, 2006), for the sample of the present investigations self-concept of ability in science had to be assumed to be a relatively weak predictor of subsequent science achievement. So, on the first layer of the structure of research questions, self-concept of ability appeared as a plausible, albeit weak, mediator of the relation between family socio-economic status and students' science achievement.

Which hypotheses were derived with respect to the second layer of the structure of research questions, i.e. the moderation of the strength of the relations between socio-economic status and self-concept of ability in science, on the one hand, and between self-concept of ability in science and science achievement, on the other hand, by grade level or school type? In case of the German educational system with its explicit ability grouping of students into the different school types of its secondary education, plausible general developmental trends of the two aforementioned relations are overruled by a pronounced big-fish-little-pond effect endowing children of differing socio-economic status and academic achievement with similar academic self-concepts. Thus, for the current examinations, a reduction of the associations between self-concept of ability in science and social background as well as science achievement was expected for the transition from elementary to secondary education. Under the assumption that the school type of Gymnasium represents a less ego-protective learning environment than the school types of Hauptschule and Grundschule a further differential hypothesis could be formulated for the moderator of school type: As top-down effects of negative global self-esteem should operate first and foremost in ego-protective learning environments (Trautwein et al., 2006), the association between socio-economic status and self-concept of ability in science was expected to be particularly weak within the school type of Gymnasium.

The fifth – and final – aim of the current analyses was the examination of the relevance of individual interest in science for the mediation of effects of socio-economic status on science achievement. This endeavor commenced, on the first layer of the structure of research questions, with assumptions about the presence of a social gradient in individual interest in science. Here, both disparities in families' capacities to provide opportunities for experiences of situational for interest in science (Bradley & Corwyn, 2002; Jacobs et al., 2005) and disparities in the quality and quantity of parent-child interactions (Duncan & Brooks-Gunn, 1997) were hypothesized to contribute to the manifestation of a social gradient in individual interest in science. Furthermore, individual interest in science was expected to display a small, yet reliable association with science achievement as interest is known to foster persistence in learning activities and deep-level processing of learning content (Hidi et al., 2004; Schiefele, 2009).

On the second layer of the structure of research questions, speculations about an accumulation of adverse effects of social background on the formation of individual interest in academic activities over the course of formal schooling formed the basis for the assumption that the relation between family socio-economic status and individual interest in science intensified with students' increasing age in the transition from elementary to secondary education. Similarly, inherent features of the person-object theory of interest, consequences of ability grouping and the growth of students' personal independence in adolescence were conjointly weighed as arguments in favor of an amplification of the association between individual interest in science and science achievement with increments in students' grade level. With respect to the school types of secondary education no further differential hypotheses were formulated. Nevertheless, the assumed amplifications of associations with individual interest in the transition from elementary to secondary education were possibly overshadowed by a big-fish-little-pond-effect in individual interest in science (Frenzel et al., 2010; Trautwein, Lüdtke, Marsh, et al., 2006).

Method

Design and Data Basis

The current investigations relied on data gathered in the initial cross-sectional phase of a longitudinal research project (*Professionswissen von Lehrkräften, verständnisorientierter naturwissenschaftlicher Unterricht und Zielerreichung im Übergang von der Primar- zur Sekundarstufe* – PLUS) on the development and interplay of classroom climate, science instruction, and students' interest in science in the transition from elementary to secondary education in Germany. The field study of the cross-sectional phase consisted of inspections of fourth- and sixth-grade classrooms. Participating teachers were asked to provide their students with a series of three 90-minute lessons on the topic of evaporation and condensation. The first of these lessons was videotaped. Students had to complete a variety of questionnaires and tests on three separate measurement occasions. On the first measurement occasion, before the onset of the series of lessons, students' fluid ability and topic-specific prior knowledge were assessed with written tests. Additionally, students completed questionnaires on classroom climate, on their perception of science instruction, and on different motivational constructs. On the second measurement occasion, immediately after the videotaped lesson, students reported on their appreciation of classroom climate and science instruction concerning that lesson. The third measurement occasion, after the series of three lessons, comprised concluding assessments of students' perceptions of classroom climate and science instruction as well as of several motivational constructs and science achievement with regard to the entire learning sequence. A questionnaire containing questions on family socio-economic status was send to the parents of all participating students. Moreover, teachers were probed for their topic-specific content and pedagogical content knowledge. In addition they reported on their attitudes towards teaching science (see Ewerhardy, 2011; Ewerhardy, Kleickmann, & Möller, 2012; Fricke, van Ackeren, Kauertz, & Fischer, 2012; Lange, 2010; Lange, Kleickmann, Tröbst, & Möller, 2012; Ohle, 2010; Ohle, Fischer, & Kauertz, 2011 for further analyses based on data from the cross-sectional phase of the PLUS-project).

For the examination of the mediation of effects of socio-economic status on student achievement by individual cognitive and motivational propensities six measures were picked from the cross-sectional data set. Mothers' and fathers' free reports of current occupational status were transformed into scores of the ISEI (Ganzeboom et al., 1992). The highest of these values for each family was selected as the indicator of family socio-economic status. The four possible mediators of the relation between social background and student achievement were operationalized by four measures taken from the first measurement occasion. Here, students' fluid ability was assessed with two subtests of the CFT 20-R (Cattell, 1963; Weiß, 2006). Prior knowledge with respect to the topic of evaporation and condensation was measured by a test containing 26 multiple-choice and multiple-select items. The motivational constructs of self-concept of ability in science and individual interest in science were assessed with Likert-style scales. Naturally, the central dependent variable of the potentially mediated relationship, students' topic-specific achievement in science after instruction, was taken from the third measurement occasion. Science achievement after instruction was measured with the same test as students' prior knowledge.

Data was gathered primarily in 2008. Teachers received a book about science instruction in exchange for their efforts. Classes as a whole were paid 50 Euros for their cooperation. Parents gave written consent to the participation of their children. A total of 1326 fourth-graders (621 girls and 702 boys, 3 students did not indicate gender) from 60 classrooms and 1354 sixth-graders (646 girls and 708 boys) from 54 classrooms, took part in the cross-sectional study. All participating students from fourth grade attended the comprehensive school type of Grundschule of German elementary education. Among the 54 classrooms from secondary schools 28 classrooms with 601 students stemmed from the school type of Hauptschule (the lowest track of German secondary education), whereas the other 26 classrooms with 753 students were from the school type of Gymnasium (the highest track of German secondary education). Average class size in fourth and sixth grade was 22.10 and 25.07 students, respectively. In the school types of Grundschule, Hauptschule and Gymnasium average class size equaled 22.10, 21.46 and 28.96 students, respectively. Students' mean age in elementary schools was 10.27 years. In secondary schools students' mean age amounted to 12.16 years. The questionnaire on family socio-economic status was returned by 1904 parents. This equaled a response rate of 71.04%. Furthermore, 503 of the responding families, i.e.

18.77% of the total sample, indicated that at least one parent was not born in Germany. In North Rhine-Westphalia, the federal state of Germany in which the study was conducted, in 2009 approximately 38.9% of fifth-graders – fifth grade is the first year of secondary education – attended a Gymnasium, while only 14.6% of fifth-graders were enrolled at a Hauptschule (cf. Autorengruppe Bildungsbericht, 2010, p. 246).

Although in most instances schools contributed exactly one classroom to the sample, occasionally schools participated with more than one classroom. In case of the subsample of fourth-graders, i.e. in case of the school type of Grundschule, 6 schools provided two classrooms for the study, whereas 2 further schools assigned three and four classrooms, respectively, for participation. With respect to the subsample of sixth-graders, 1 school of the school type of Hauptschule as well as 2 schools of the school type of Gymnasium contributed two classrooms to the study. In addition, 1 school of the school type of Gymnasium participated with three classrooms in the study. So, the 114 classrooms of the sample were recruited from 98 separate schools.

With regard to the population of fourth- and sixth-graders in North Rhine-Westphalia, the sampling for the current analyses deviated from random sampling in two respects, entailing specific consequences for the interpretation of results. First, the sample was a convenience sample recruited mostly within a radius of 60 kilometers around the cities of Muenster and Essen. The recruitment of classrooms was accompanied by substantial self-selection. Schools were initially approached by phone. Approximately one fifth of the contacted schools consented to participate with at least one classroom in the study. So, with one half of the participating classrooms located in the Muenster region and the other half of participating classrooms located in the Ruhr region, rural and urban classrooms were represented with roughly equal weight in the sample. However, assuming that classrooms with burned-out teachers or severe disturbances of learning climate refrain from participation in research projects, it appears plausible that self-selection engendered a positive bias with respect to overall functioning of classrooms and such variables as student achievement and motivation for the entire sample.

Second, sampling of secondary schools was restricted to the school types of Hauptschule and Gymnasium, thereby excluding a large portion of students of intermediate achievement and ability from the subsample of sixth-graders. Obviously, the fourth-graders sampled in a comparatively unrestricted manner, this introduced a certain mismatch between the subsamples of fourth- and sixth-graders with regard to their overall representativeness for the respective subpopulations. In particular, the estimation of the actual state of affairs in the subpopulation of sixth-graders based on the restricted subsample of sixth-graders is subject to specific biases. As the correlation between variables tends to be inflated by the exclusion of persons with intermediate values on those variables (for a memorable illustration see Stelzl, 1982/2005, pp. 144–158), it has to be assumed that, when the subsample of sixth-graders is considered as a whole, the strength of associations between family socio-economic status, possible mediators and student achievement is overestimated with regard to the actual relations in the subpopulation of sixth-graders. However, the severity of this biasing influence depends on the actual distributions of critical variables in the groups of students sampled from the school types of Hauptschule and Gymnasium, i.e. it depends on the actual intensity of the dilution in the central areas of the distributions of critical variables due to exclusion of school types with intermediate academic requirements. Note that the distributions of student achievement or general mental ability display considerable overlap for different school types in German secondary education. Moreover, as the school types of Hauptschule and Gymnasium represent the least and most advantageous developmental environments, respectively, for achievement in German secondary education (Baumert et al., 2006b), the deliberate exclusion of school types with intermediate academic requirements presumably amplified any differential effects of assignment of students to segregated learning environments within the subsample of sixth-graders as compared to the actual state of affairs in the subpopulation of sixth-grade students. In many respects, the use of the restricted subsample of sixth-graders for generalization to the subpopulation of sixth-graders suffers from the same methodological dubiousness as the use of a comparison of extreme groups for generalization to an intact population (Stelzl, 1982/2005).

Nonetheless, apart from the seemingly unfavorable unrepresentativeness of the subsample of sixth-graders for the subpopulation of sixth-graders, the pronounced oversampling of students from the school type of Hauptschule entails an important genuine virtue. As mentioned above, in 2009 only 14.6% of fifth-graders in North Rhine-Westphalia attended a Hauptschule. Therefore, the decided oversampling of students from the school type of Hauptschule enabled a much more accurate estimation of effects for students from this school type than would have been attainable with a subsample more representative of the subpopulation of sixth-graders. (Knight, Roosa, & Umaña-Taylor, 2009; Wainer, 2010a). In other words, the characteristics of the subsample of sixth-grade students allowed separate generalizations to the populations of sixth-graders from the school type of Hauptschule and of sixth-graders from the school type of Gymnasium with improved precision.

Analyses

Before substantive research questions were tackled, the data on family socio-economic status, on the potential mediator variables and on student achievement was submitted to specific preparatory and auxiliary analyses. For a subsample of 300 families, mothers' and fathers' free reports of occupational status were double-coded in order to investigate various indicators of intercoder reliability for the assignment of ISEI-scores (Ganzeboom et al., 1992). In particular, factual agreement, relative agreement, and structural agreement between coders were examined (Agresti, 1990; Maaz, Trautwein, Gresch, Lüdtke, & Watermann, 2009; Wirtz & Caspar, 2002). The culture fair intelligence test of fluid ability, the science achievement test, which was used to assess both prior knowledge and the dependent variable of student achievement, and the two motivational scales, self-concept of ability in science and individual interest in science, were submitted to test and scale analyses. This included calibration according to psychometric models of the Rasch family (Andrich, 1978; Masters, 1982; Nering & Ostini, 2010; Rasch, 1960), computation of central indices of classical test theory (Lord & Novick, 1968) and model-based explorations of differential item functioning (Wu, Adams, Wilson, & Haldane, 2007). This procedure served to obtain linear measures for subsequent investigations, to thoroughly describe the properties of the tests and scales and to identify peculiarities of the measures possibly relevant for the interpretation of results of subsequent inferential analyses. The problem of missing data was handled by multiple imputation with the R-package mice (van Buuren & Groothuis-Oudshoorn, 2011). The imputation model took into consideration a set of background variables otherwise unrelated to the substantial analyses. A total of 50 imputed data sets was generated for subsequent inferential analyses.

As students were nested into classrooms, it was mandatory to follow a multilevel approach in the inferential analyses in order to receive unbiased estimations of the effects (Dedrick et al., 2009; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999; see Geiser, 2010, chap. 5 for a non-technical introduction to multilevel modeling). The relations of primary interest for the current investigations, i.e. the relations between family socio-economic status, cognitive and motivational propensities as potential mediators and the central dependent variable of student achievement, were located at the individual level of the analyses. These relations were explored against a background or in a context of developmental, i.e. grade level, and institutional, i.e. school type, variables on the class level. Some of the estimated models contained only main effects for these variables on the class level. So, in these cases, a background pattern of mean differences in the respective dependent variable between grade levels or school types was estimated on the class level and additively combined with the configuration of effects on the individual level. Many of the research questions, however, were concerned with the possible moderation of effects located at the individual level by developmental or institutional variables located at the class level. The estimation of models that included terms for cross-level interactions between effects on the individual level and effects on the class level addressed this issue. In these cases, the terms for cross-level interactions captured the moderating influence of the developmental or institutional context on basic relations at the individual level. In more technical terminology, a variety of intercepts-as-outcomes and intercepts-and-slopes-as-outcomes multilevel models was estimated to answer the research questions (Geiser, 2010; Luke, 2004).

Thereby, dealing with the research questions one at a time, the inferential analyses proceeded in a stepwise fashion. First, the social gradient for student achievement after the series of lessons was assessed. In other words, the inferential analysis started with the establishment of a relation between family socio-economic status and student achievement that potentially could be mediated by individual cognitive and motivational propensities. Second, the actual mediation analyses were performed separately for each potential mediator. This included the examination of the association between socio-economic status and the respective mediator as well as the exploration of the relation between the respective mediator and the dependent variable of student achievement. Crucially, the impact of this latter relation on the association between socio-economic status and student achievement was inspected, the final evidence for the existence of a mediated relationship. Third, eventually, under consideration of the results of the previous analyses, the relative importance of each of the four mediators was evaluated. On the one hand, multilevel regression models containing more than one mediator simultaneously were fitted to the data. On the other hand, a multilevel path model incorporating all the effects and moderations identified as relevant in the preceding investigations was set up. In all multilevel analyses, both the central dependent variable of student achievement after instruction, and the potential mediators, fluid ability, topic-specific prior knowledge, self-concept of ability in science and individual interest in science, were included as standardized variables. Grade level and school type were incorporated as dummy-coded variables with fourth grade and the school type of Grundschule, respectively, as the reference categories. The inferential analyses were conducted for each imputed data set by means of maximum likelihood estimation. The results of these analyses were aggregated according to the rules devised by Rubin (1987). The software package Mplus 6.12 was used for estimation of multilevel models and aggregation of results (Muthén & Muthén, 2011).

Although structural equation modeling with full information maximum likelihood estimation offers a convenient method to integrate the formulation of psychometric models, the handling of missing data and the testing of substantial hypothesis in a single step of analysis (Arbuckle, 1996; Kaplan, 2000; Rose, Pohl, Böhme, & Steyer, 2010), for the current analyses these three aspects of data analysis were intentionally kept separated. This was thought to add clarity to the analytic procedure as the mutual interdependencies of measurement and structural parts of structural equation models as well as the dependency of full information maximum likelihood estimation on variable selection might run the risk to obscure the attribution of specific effects to one of the three aspects of data analysis (Lüdtke & Robitzsch, 2010). Moreover, the strict separation of different aspects of data analysis might foster a scientific practice that is based on the division of labor between different researchers and thus enhances critical scrutiny of results in all phases of data analysis. So, the global set up of the current analyses might be considered a trade of convenience for clarity.

Of course, besides family socio-economic status, grade level, school type and cognitive and motivational propensities, there is a plethora of further variables, for instance, characteristics of teachers and of their instruction, that contribute to the prediction of student achievement (Helmke, 2010; Helmke & Weinert, 1997; Lipowsky, 2006). In educational research, it is conventional practice to drop long lists of variables into regression models to control for the various influences on student achievement and to explain as much variance as possible. Nevertheless, in order to obtain credible effect estimates from such analyses, the respective data has to meet certain assumptions, such as the absence of multicollinearity and the linearity of effects on the dependent variable (Achen, 2005; Shieh & Fouladi, 2003). Unfortunately, even small deviations from the assumption of linear relations can cause rather severe distortions of effect estimates in regression models with more than one predictor (Achen, 2005). So, with regard to the estimation of specific regression weights, not with regard to the maximization of explained variance, the simple rule that more predictors are better only holds when additional variables conform to the underlying assumptions. However, the incorporation of more variables in a given regression model both increases the probability that certain included variables do not conform to the assumptions and complicates the interrelations between predictors (Achen, 2005; Stelzl, 1982/2005). Thus, sound theory and thorough, i.e. in many cases stepwise, data analysis are essential to avoid harmful model misspecifications. Accordingly, as the current paper did not elaborate the association of student achievement with further variables theoretically, the regression models of the inferential analyses were restricted to the predictors of socio-economic status, cognitive and motivational propensities, grade level as well as school type.

Coding of Occupational Status

In the parental questionnaire mothers and fathers were asked to provide free descriptions of their current occupation. The actual wording of those questions followed the diction of items used in previous research (Bos et al., 2005). Parents were explicitly instructed not to name the profession in which they were originally trained but to denote their current occupation as precisely as possible. On the basis of the appendix in Ganzeboom and colleagues (1992), these free descriptions of current occupation were transformed into scores of the ISEI. Relying on the International Standard Classification of Occupations (ISCO), the ISEI offers four levels of increasingly fine-grained coding of occupational status. On the broadest level of categorization the ISEI discerns effectively seven main groups ranging from professional, technical and related workers to production workers, transport equipment operators and laborers. Within the seven main groups, more detailed classifications are possible. Particularly, descriptions of current occupation can be classified into minor groups, unit groups and occupational titles, with the latter as the finest level of categorization. As a consequence of this, on the broadest level of classification, ship engineers and chemists, for instance, share a score of 67 on the ISEI (for membership in the major group of professional, technical and related workers), whereas, on the level of unit groups, they receive scores of 53 and 73, respectively (cf. Baumert & Maaz, 2006). So, in the process of coding, four variables corresponding to the four hierarchical levels of categorization were created for both mothers' and fathers' free descriptions of occupational status. Moreover, as the original construction of the ISEI was based on a sample of men working full time, unemployed and retired individuals, homemakers as well as full-time students among the participating parents were assigned a score of 0 on the four variables. This approach differed from the assessment of socio-economic status in large-scale studies of student achievement in which non-working parents were asked to report their last occupation. In those studies parents not active in the labor force could receive a score larger than 0 on the ISEI (e.g, Maaz, Trautwein, et al., 2009; Prenzel et al., 2006). However, in agreement with current research practice, mothers were given scores on the ISEI although it was devised originally for a sample of men (cf. Ganzeboom et al. 1992, pp.14-15, for a discussion of this issue). According to the scheme described above, I coded the free descriptions of current occupation for the entire data set.

Double Coding

The free descriptions of current occupation of a random sample of 375 families from the entire data set were double coded by a student research assistant. After double coding, the procedure outlined in Maaz, Trautwein and colleagues (2009) was applied to the results of the coding process in order to gauge the extent of intercoder reliability as well as the stability of indicators of criterion validity across coders. First, factual agreement, i.e. the occurrence of identical classifications across coders, was thoroughly explored. Moreover, as measures of factual agreement are rather strict indicators of intercoder reliability, relative agreement between coders, i.e. the amount of systematic association between coders, was inspected. Finally, structural agreement was investigated. For this purpose associations of scores on the ISEI with external criteria were computed for different coders and compared to each other.

Factual agreement. As measures of factual agreement between coders both the joint probability of agreement (cf. Wirtz & Caspar, 2002) and Cohen's kappa (cf. Cohen, 1960) were calculated for mothers' and fathers' scores on the ISEI. These calculations were conducted for all four levels of categorization of the ISEI. The joint probability of agreement represents the most basic indicator of agreement as it simply denotes the proportion of cases with identical classifications across coders. In contrast to this, Cohen's kappa is a standardized measure of agreement that corrects for chance agreements between coders. The computation of Cohen's kappa returns values of 0 when factual agreement does not exceed agreement expected by chance, and it returns values of 1 in case of perfect agreement, with values in between indicating increasing degrees of agreement . However, the adequacy of accounting for chance agreements is debatable as the degree to which coders really resort to chance is unknown (Uebersax, 1987). So, for the present investigations, the joint probability of agreement and Cohen's kappa might be considered as upper and lower bounds of the actual agreement between coders.

Table 1
Intercoder Agreement by Parent and Level of Coding

Parent	Level of Coding							
	Major Group		Minor Group		Unit Group		Title	
	Joint	Kappa	Joint	Kappa	Joint	Kappa	Joint	Kappa
Mother	87.5	.84	75.2	.71	66.4	.61	64.3	.58
Father	84.8	.79	61.1	.59	49.1	.47	48.5	.47

Note. Joint = Joint Probability of Agreement, Kappa = Cohen's Kappa.

The investigation of measures of factual agreement across parents and levels of classification of the ISEI revealed two general trends. Factual agreement between coders declined with increasing exactness of the levels of categorization of the ISEI. Moreover, factual agreement was higher for mothers' scores than for fathers' scores (see Table 1). The joint probability of agreement for mothers' scores decreased from 87.5 percent on the broadest level of categorization to 64.3 percent on the finest level of categorization. Similarly, Cohen's kappa for mothers' scores declined from .84 to .58. With regard to fathers' scores on the ISEI, the joint probability of agreement between coders waned from 84.8 percent on the broadest level of classification to 48.5 percent on the finest level of classification. Here, Cohen's kappa decreased from .79 to .47.

Relative agreement. If, in a given coding situation, one coder consistently assigns scores one unit higher than another coder, this results in an absence of factual agreement between the two coders despite a perfect association between the scores. Thus, indicators of factual agreement are rather harsh measures of intercoder reliability. So, to receive a balanced picture of the degree of agreement between coders for the current analyses, correlations between the scores assigned by the two coders were computed for all four levels of categorization of the ISEI. In addition, the means and standard deviations of mothers' and fathers' ISEI-scores on the level of occupational titles were computed for each coder. Moreover, the investigation of the relation between social background and student achievement necessitates the creation of a unified measure of family socio-economic status. In educational research the highest of mothers' and fathers' scores on ISEI, the so-called HISEI, is often used as a comprehensive indicator of family socio-economic status (e.g., Watermann & Baumert, 2006). Therefore, means and standard deviations on the level of occupational titles were also computed for both coders' HISEI-scores.

Table 2
Intercoder Correlations by Parent and Level of Coding

Parent	Level of Coding			
	Major Group	Minor Group	Unit Group	Title
Mother	.96***	.96***	.95***	.95***
Father	.90***	.92***	.91***	.91***

Note. *p < .05. **p < .01. ***p < .001.

The Pearson product-moment correlations between scores obtained from different coders never dropping below values of .90, there was a positive and strong association between the scores of the two coders across mothers and fathers as well as across different levels of classification of the ISEI (see Table 2). Differences between coders with regard to the means and standard deviations of mothers' and fathers' ISEI-scores on the level of occupational titles were small (see Table 3). Two-tailed t-tests for paired samples showed that these differences were statistically insignificant, $t(372) = 1.76$, $p = .08$, and $t(366) = -0.52$, $p = .60$, respectively. Correspondingly, effect sizes for these differences indicated negligible effects, Cohen's *d* = 0.03 and Cohen's *d* = 0.00, respectively. However, the difference between coders with respect to the HISEI-scores on the level of occupational titles was statistically significant, $t(375) = 2.66$, $p = .01$. Nonetheless, the calculation of the corresponding effect size, Cohen's *d* = 0.07, indicated a practically insignificant difference between coders.

Table 3
Means and Standard Deviations of ISEI-Scores by Coder

Coder	Mothers' ISEI		Fathers' ISEI		HISEI	
	M	SD	M	SD	M	SD
Coder 1	31.53	26.65	45.47	19.51	50.36	18.23
Coder 2	32.29	27.30	45.25	19.34	51.57	17.51

Structural agreement. In order to assess the validity of measures of socio-economic status obtained from different coders, it is suitable to investigate the association of these measures with external criteria across different coders. For the current analyses, the student characteristics of topic-specific science achievement prior to instruction and fluid ability as well as the number of books at home were selected as relevant external criteria. Positive associations with socio-economic status were expected for all three criteria. Weighted likelihood estimates of person ability generated from specific item response models were used as measures of science achievement and fluid ability (see the chapter below for details on the calibration of these measures). Information on the possession of books was gathered by an open item in the parental questionnaire. Correlations of socio-economic status with the external criteria were calculated for mothers' and fathers' ISEI-scores on the level of occupational titles as well as for the HISEI-scores on the level of occupational titles. Finally, to test predictive validity across coders, regressions of science achievement on socio-economic status were conducted, both with socio-economic status as the sole predictor and controlling for fluid ability. The HISEI-scores on the level of occupational titles were used as measures of socio-economic status in these regressions.

The correlations of socio-economic status with student achievement, fluid ability and books at home differed unsystematically between coders across the different forms of ISEI-scores investigated (see Table 4). Differences between coders ranged in size from .01 to .05. However, none of these differences was statistically significant, with two-tailed test statistics falling between $z = -0.39$, $p = .35$ and $z = 0.79$, $p = .22$. The regression analyses featuring HISEI-scores as predictors of science achievement yielded similar results (see Table 5). The standardized regression coefficients for socio-economic status varied by values of .03 and .04

in size between coders, whereas the corresponding explained variance differed by values of .02.

Table 4
Correlations of ISEI-Scores with External Criteria by Coder

External Criterion	Mothers' ISEI		Fathers' ISEI		HISEI	
	Coder 1	Coder 2	Coder 1	Coder 2	Coder 1	Coder 2
Science Achievement	.17**	.18**	.15**	.12*	.24**	.21**
Fluid Ability	.08	.11*	.10	.08	.12*	.15**
Books at Home	.13**	.14**	.32**	.29**	.35**	.30**

Note. *p < .05. **p < .01. ***p < .001.

Summary. Intercoder agreement in terms of joint probability of agreement and Cohen's kappa fell in a similar range as obtained by Maaz, Trautwein and colleagues (2009) for intercoder agreement between professional coding institutes and research assistants with regard to the assignment of categories of the ISCO. In fact, intercoder agreement in the current investigations tended to be somewhat higher, probably due to the use of ISEI-scores. Likewise, the differences in means and standard deviations of ISEI-scores between coders fell below the differences disclosed in previous research, with statistically, or at least practically, insignificant disparities between coders. The same holds true for differences between coders in correlations with or regressions on external criteria (cf. Maaz, Trautwein, et al., 2009). Altogether, the preceding examinations argue in favor of a great robustness of ISEI- and HISEI-scores against assignment by different coders.

Table 5
Regressions of Science Achievement on Socio-economic Status and Fluid Ability by Coder

Predictor	Coder 1		Coder 2	
	Model 1	Model 2	Model 1	Model 2
HISEI	0.24***	0.20***	0.21***	0.16**
Fluid Ability		0.38***		0.38***
R^2	*.06*	*.20*	*.04*	*.18*

Note. All regression coefficients are standardized.
*p < .05. **p < .01. ***p < .001.

Calibration of Instruments

Traditionally, the world of psychometrics is not of single mind on the issue of appropriate statistical models for measurement in the social sciences (e.g., Reckase, 2009, pp. 20/21). Whereas a faction of psychometricians contends that exclusively the application of one-parameter logistic models, i.e. the notorious Rasch model and its refinements, leads to valid measures, the majority of psychometricians considers one-, two-, and three-parameter models as a common class of appropriate models in item response theory. These two traditions have developed genuine views on the crucial aspects of test design as well as genuine indices and corresponding conventions for the evaluation of instrument quality. Without doubt, both traditions represent established approaches to the construction and analysis of psychological scales and tests (Andrich, 2004; Rupp & Zumbo, 2003; Thissen, 2009). Although the usefulness of the concept of specific objectivity, often put forward by advocates of the Rasch model as the theoretical reason for its allegedly exclusive status, may be disputed with regard to the necessity to generalize from test scores to populations of persons and items (Holland, 1990; Kane, 2011; Wainer, 2010b), the Rasch model disposes of desirable properties such as computational simplicity and a straightforward mapping of item content to person ability (for non-technical treatments of the principles of Rasch measurement see Bond & Fox, 2007; Wilson, 2005).

Thus, all scales and tests completed by the participating students of the present study were submitted to test analyses with models from the family of one-parameter item response models. In particular, the culture fair intelligence test, the science achievement test, and the two motivational scales were calibrated according to the simple Rasch model (Rasch, 1960), the partial credit model (Masters, 1982), and the Rasch rating scale model (Andrich, 1978), respectively. The program Acer Conquest 2.0 was used for these calibrations (Wu, Adams, Wilson, & Haldane, 2007). This program provided the multidimensional random coefficients multinomial logit model as a general framework for the estimation of various one-parameter item response models by means of marginal maximum likelihood estimation (Adams, Wilson, & Wang, 1997).

As its name suggests, the simple Rasch model is the basic blueprint of a one-parameter item response model. The simple Rasch model is suitable for the calibration of dichotomously scored items. Defining the probability of a correct response as an exponential function of the difference between person ability and item difficulty, the simple Rasch model locates estimates of person ability and item difficulty on a common logit scale. For a test composed of dichotomously scored items, the application of the model provides an individual estimate of difficulty for each item (Rasch, 1960). The rating scale model represents a refinement of the simple Rasch model that is adequate for the analysis of Likert-style items with ordered response categories. In order to understand the structure of the rating scale model, it is helpful to conceive the endorsement of a specific response category as the passing of successive steps of increasing difficulty. In this regard, the endorsement of the highest response category of an item with three response categories entails two steps, namely the transition from the first to the second response category and the transition from the second to the third response category. So, the application of the rating scale model to a scale of Likert-style items yields estimates of item difficulties as well as estimates of step parameters of a step structure common to all items. In order to determine the exact difficulty of a specific step of a specific item, it is necessary to add the respective item difficulty and step parameter. In other words, the formulation of the rating scale model extends the item difficulty of the simple Rasch model to a sum of item difficulty and step parameter (Andrich, 1978). The partial credit model, in turn, constitutes a refinement of the rating scale model. Whereas the rating scale model assumes a step structure that is common to all items, the partial credit model assumes a specific step structure for each item. Therefore, the partial credit model is suitable for the calibration of tests that comprise both dichotomously scored items and polytomously scored items with varying numbers of ordered response categories. The application of the partial credit model provides estimates of item difficulties as well as estimates of item-specific step parameters (Masters, 1982).

The weighted mean square residual (infit) and the unweighted mean square residual (outfit) served as indicators of the fit of the items to the three one-parameter item response models (Wilson, 2005; Wright, 1984; Wright & Masters, 1981; see Karabatsos, 2000, for a critical discussion of residual fit statistics). For these two indices, values close to 1.0 flag optimal compatibility of the observed data to the model under consideration. Values greater than 1.0 indicate greater variability in the observed responses than predicted by the statistical model whereas values smaller than 1.0 detect less variability in the observed responses than expected by the model applied. The item infit weighs responses according to the distance between respondents' ability and the difficulty of the item under consideration. A small distance between ability and difficulty implies relative amplification of a respondent's impact on the item infit. In contrast, investigated respondents contribute to the item outfit in the same way, irrespective of their individual ability (Bond & Fox, 2007; Wilson, 2005; Wright & Stone, 1999). Popular recommendations for the evaluation of item fit define reasonable fit ranges with respect to the nature of the instrument at hand. For conventional multiple-choice tests fit values between 0.7 and 1.3 are widely considered as acceptable. For high stakes testing fit values of 0.8 and 1.2 have been proposed as boundaries of reasonable item fit (Adams & Khoo, 1996; Bond & Fox, 2007; Smith, 2000; Wright, Linacre, Gustafsson, & Martin-Löf, 1994). For rating scale instruments fit values between 0.6 and 1.4 have been suggested as constituting adequate item fit (for an elaborate classification see Fisher, 2007; Wright et al., 1994). Generally, mean square residuals above 1.0, i.e. underfit to the model applied, are considered more detrimental to the purpose of accurate measurement than mean square residuals below 1.0, i.e. overfit to the model applied (Bond & Fox, 2007; Wilson, 2005). However, items with substantial overfit tend to inflate estimated standard errors and reliability (Bond & Fox, 2007).

Beyond item infit and item outfit, the targeting of the items on the persons was consulted for the evaluation of instrument quality as well. An item provides most information for the estimation of a person's ability when its difficulty is similar to that person's ability. That is why a calibration in which the person ability distribution matches the item difficulty distribution yields an overall more accurate estimation of person ability than a calibration in which the person ability distribution and the item difficulty distribution are located far apart (Wright & Stone, 1999). In order to gauge this aspect of instrument quality, the means and standard deviations of the ability and difficulty distributions were computed for each instrument.

Estimates of individual person ability for each instrument were obtained by weighted likelihood estimation (Warm, 1989). The weighted likelihood estimates of person ability were used as linear measures in the subsequent substantial analyses. Correspondingly, the person separation reliability of these estimates was inspected as an indicator of instrument quality. As its name suggests, this reliability index – adopting values between 0 and 1 – quantifies the capacity of an item pool to differentiate persons from each other with respect to their ability. It is similar to Cronbachs' alpha (Cronbach, 1951; Wright & Stone, 1999).

For comparative purposes, indicators of instrument quality pertinent to classical test theory were computed for all instruments as well. With respect to individual items, these included the proportion of correct responses or the mean raw score, depending on the type of item, and the item discrimination. With regard to instruments as a whole, the percentage of the total variance explained by the first component of a principal components analysis as well as the ratio of the first and second eigenvalues of a principal components analysis were calculated as indicators of unidimensionality (Hattie, 1985). As a rough guideline, it has been suggested to assume unidimensionality when the first component explains at least 20% or 40% of the total variance (Carmines & Zeller, 1979; Reckase, 1979). Similarly, a large ratio of the first and second eigenvalues in combination with a second eigenvalue that is not considerably larger than the remaining eigenvalues can be considered as evidence for unidimensionality (Hutten, 1980; Lord, 1980). Furthermore, the value of Cronbach's alpha was calculated as an index of internal consistency (Cronbach, 1951). The program IBM SPSS 19 was used for the outlined analyses in the framework of classical test theory.

Test and scale analyses concluded with an exploration of differential item functioning. Historically, the concept of differential item functioning was elaborated in the context of high-stakes testing (Zumbo, 2007a). According to common definitions, an item exhibits differential item functioning when respondents from a focal group (e.g., girls) solve that item more or less often than equally proficient respondents from a corresponding reference group (e.g., boys), i.e. when group membership exerts an influence on solution rates under control of proficiency (Meredith & Millsap, 1992; Zumbo, 2007a, 2007b). Emphasizing the computational implications of specific objectivity, dogmatic interpretations of the Rasch model view the absence of differential item functioning as a central prerequisite for the its applicability (e.g., Kubinger & Draxler, 2007). Nevertheless, the occurrence of differential item functioning does not threaten the obtainment of unbiased estimators of person ability as long as the effects of differential item functioning against and in favor of a specific group of respondents counterbalance each other (e.g., Artelt & Baumert, 2004). So, in a pragmatic approach, the inspection of differential item functioning can serve at least two purposes. First off, it can render information on the existence and size of biasing influences on the estimation of person ability, and thereby potentially qualify the results of subsequent substantial analyses or reveal the necessity of additional sensitivity analyses. Furthermore, the investigation of differential item functioning can disclose the presence of response patterns typical for specific groups of respondents, without necessarily compromising the unbiasedness of estimations of person ability in a practically relevant way (see also Zumbo & Rupp, 2004; Zumbo, 2007a, 2007b). In other words, with regard to the distinction between the general facet and the differential facet of psychological research (Asendorpf, 1990; Cronbach, 1957; Stern, 1911), a thorough exploration of differential item functioning can provide valuable insights into the adequacy of a general interpretation of ability estimations.

In face of a plethora of methods for the detection of differential item functioning (Camilli & Shepard, 1994; Holland & Wainer, 1993; Mapuranga, Dorans, & Middleton, 2008; Wainer, 2010a), a simple and conceptually straightforward – though not computationally optimal – way of inspecting differential item functioning is to conduct separate instrument calibrations for two disjunctive groups (e.g., boys vs. girls) and to compare the resulting item difficulties. For illustrative purposes this approach was taken as a first step of the analysis of differential item functioning for the culture fair test of intelligence. In particular, graphs contrasting item difficulties from separate calibrations were inspected for the presence of differential item functioning (Wright & Stone, 1979, 1999). However, a model-based approach was implemented as the central form of exploration of differential item functioning for all scales and tests (Janssen, 2011; Thissen, Steinberg, & Wainer, 1993; Wu et al., 2007). This approach incorporated the successive comparison of nested models by means of the likelihood ratio test (Thissen, Steinberg, & Gerrard, 1986; Thissen et al., 1993). It was explored whether the inclusion of additional interactions between grouping variables (e.g., gender) and item difficulties or step parameters beyond main effects of grouping variables, item difficulties and step parameters significantly improved the overall fit of the respective item response models (Wu et al., 2007, chap. 8). Moreover, these models were used to calculate the differences in difficulty of specific items or specific combinations of items and steps on the logit scale of the Rasch model as an effect size for differential item functioning between certain disjunctive groups of respondents. Absolute differences of less than .426 logits and of more than .638 logits were evaluated as negligible and large amounts of differential item functioning, respectively. In accordance with that, absolute differences between .426 and .638 logits were considered as intermediate amounts of differential item functioning (Paek, 2002; Paek & Wilson, 2011; Wilson, 2005). Note, however, that the aforementioned procedure does not constitute a strict analysis of differential item functioning for specific items in which items under investigation are tested one by one against a context of established items free of differential item functioning but rather represents an exploration of differential item functioning for predefined tests and scales.

The grouping variables for the analyses of differential item functioning included gender (boys vs. girls), socio-economic status (low socio-economic status vs. high socio-economic status), immigrant background (both parents born in Germany vs. at least one parent born abroad), grade level (fourth-graders vs. sixth-graders), and raw score on the respective instrument (low scorers vs. high scorers). Moreover, school type was investigated as a grouping variable. This entailed comparisons of all three possible pairings of groups formed by school type (Grundschule vs. Hauptschule, Grundschule vs. Gymnasium, and Hauptschule vs. Gymnasium). The grouping variable for socio-economic status was created by a median-split of the HISEI-score. In the same vein, the grouping variable for scoring on the instrument was generated by a median-split of the respective raw score.

Culture Fair Assessment of Intelligence

Students completed two subtests of the CFT 20-R, a German adaptation of the Culture Fair Intelligence Test (Cattell, 1957, 1963; Weiß, 2006). These subtests were Perceptual Series and Topology, both taken from the first part of the CFT 20-R. They featured multiple-choice items with five response alternatives. Perceptual Series afforded students to select logical continuations of series of three figures. It consisted of 15 items. Topology confronted students with complex figures of geometrical shapes. Students had to select figures in such a way that the placement of dots within geometrical shapes conformed to certain given rules. It comprised 11 items (Jacobs, Petermann, & Weiß, 2007). The intention of the use of the CFT 20-R was to assess students' general mental capacity in terms of fluid ability, i.e. free from influences of ethnic, cultural, social and educational background.

Test analysis. The 26 items of the two subtests of the CFT 20-R were submitted to a joint calibration according to the simple Rasch model (Rasch, 1960). The calibration relied on test responses from 2500 students. The estimated item difficulties ranged from -2.676 logits to 3.340 logits. Generally, the items of Perceptual Series were less difficult (M = -0.526, SD = 1.461) than the items of Topology (M = 0.718, SD = 1.220). Item infit values ranged from 0.94 to 1.06, whereas item outfit values varied between 0.77 and 1.34 (see Table 6).

Apart from item fit, the matching of students and items was investigated as a criterion of instrument quality. The distribution of item difficulty had a – for the purpose of model identification predefined – mean of 0.00 logits and a standard deviation of 1.48 logits. The mean of the distribution of person ability amounted to 0.59 logits with a standard deviation of 0.74 logits. So, although the average student tended to be somewhat more able than demanded by the average item, the distribution of item difficulty efficiently spanned the entire range of person ability, with the easiest item solved by almost every student and the hardest item solved by very few students. The separation reliability of the weighted likelihood estimates of person ability was 0.70.

Finally, for comparative purposes, indices of instrument quality relevant to classical test theory were computed. Axiomatically, the actual solution rates for the items mirrored the estimated item difficulties. In particular, the easiest items (cft_ps1, cft_ps2, cft_ps3) were solved by 95% or more of the respondents, whereas only 7% of the students responded correctly to the hardest item (cft_tp11). Item discriminations, i.e. corrected item-total correlations, ranged from .05 to .37 (see Table 6). In order to tackle the issue of dimensionality, a principal component analysis was conducted. The first component obtained in this analysis explained 13.08% of the total variance. The principal component analysis identified seven components with eigenvalues greater than 1.00. In order, the eigenvalues of the first five components were 3.40, 1.66, 1.27, 1.22, and 1.09. Correspondingly, the ratio of the first and second eigenvalues was 2.05. Cronbachs' alpha of the test of fluid ability was .70.

Table 6
Results of Probabilistic and Traditional Item Analyses for the Culture Fair Intelligence Test

Item	Difficulty		Infit	Outfit	*P*	r_{it}
	Estimate	SE				
cft_ps1	-2.676	0.061	0.95	0.77	.95	.28
cft_ps2	-2.733	0.061	0.94	0.67	.96	.31
cft_ps3	-2.648	0.060	0.94	0.74	.95	.30
cft_ps4	-1.009	0.044	1.01	1.07	.81	.20
cft_ps5	-1.413	0.047	0.96	0.92	.86	.29
cft_ps6	-0.978	0.044	0.95	0.91	.80	.33
cft_ps7	-0.962	0.044	0.97	0.97	.80	.29
cft_ps8	0.301	0.037	1.09	1.11	.56	.12
cft_ps9	-0.274	0.039	0.99	0.98	.68	.28
cft_ps10	0.132	0.038	1.03	1.04	.60	.23
cft_ps11	-0.180	0.039	1.01	1.02	.67	.24
cft_ps12	-0.095	0.038	0.98	0.97	.65	.30
cft_ps13	1.318	0.039	1.01	1.02	.34	.23
cft_ps14	1.512	0.039	1.00	1.01	.30	.25
cft_ps15	1.811	0.041	1.01	1.09	.25	.18
cft_tp1	-0.824	0.042	0.98	0.95	.78	.29
cft_tp2	-0.927	0.043	1.04	1.10	.80	.17
cft_tp3	1.340	0.039	0.99	1.01	.34	.27
cft_tp4	0.132	0.038	0.97	0.96	.60	.32
cft_tp5	0.290	0.037	0.94	0.93	.57	.37
cft_tp6	0.200	0.038	0.98	0.98	.59	.30
cft_tp7	1.127	0.038	1.02	1.04	.38	.21
cft_tp8	0.178	0.038	1.04	1.05	.59	.21
cft_tp9	1.601	0.040	1.02	1.03	.29	.21
cft_tp10	1.436	0.039	1.06	1.10	.32	.13
cft_tp11	3.340[a]	0.216	1.03	1.34	.07	.05

Note. Item labels provide information on subtest (ps = Perceptual Series, tp = Topology) and item position.
[a]Parameter estimate is constrained.

Differential item functioning. Due to missingness on the grouping variables and on the test of fluid ability, examinations of differential item functioning were based on changing numbers of students that deviated to varying degrees from the entire sample of 2680 students. The investigation of differential item functioning for gender as the grouping variable comprised data from 1326 boys and 1173 girls. Additionally, differential item functioning was explored for the raw score in the culture fair intelligence test as the grouping variable. Both the low-scoring group and the high-scoring group, formed by a median-split of the raw score, consisted of 1250 students. Furthermore, analyses of differential item functioning covered a comparison of 893 students of low socio-economic status with 922 students of high socio-economic status as well as a comparison of 1323 students without immigrant background with

466 students with immigrant background. The examination of differential item functioning for the grouping variable of grade level relied on data from 1255 fourth-graders and 1245 sixth-graders. Correspondingly, the three pairwise comparisons constituting the investigation of differential item functioning for the grouping variable of school type comprised 1255 students from the school type of Grundschule, 528 students from the school type of Hauptschule and 717 students from the school type of Gymnasium.

The exploration of differential item functioning for the test of fluid ability started with an inspection of graphs contrasting separate calibrations for the groups of the various grouping variables (Wright & Stone, 1979, 1999). As examples for this procedure, the graph for the grouping variable of socio-economic status and the graph comparing the calibration for students from the school type of Grundschule with the calibration for students from the school type of Hauptschule are presented here (see Figure 2). The graph for the grouping variable of socio-economic status revealed no apparent disparities between the calibrations for students of low and for students of high socio-economic status. In this graph the pairs of item difficulties

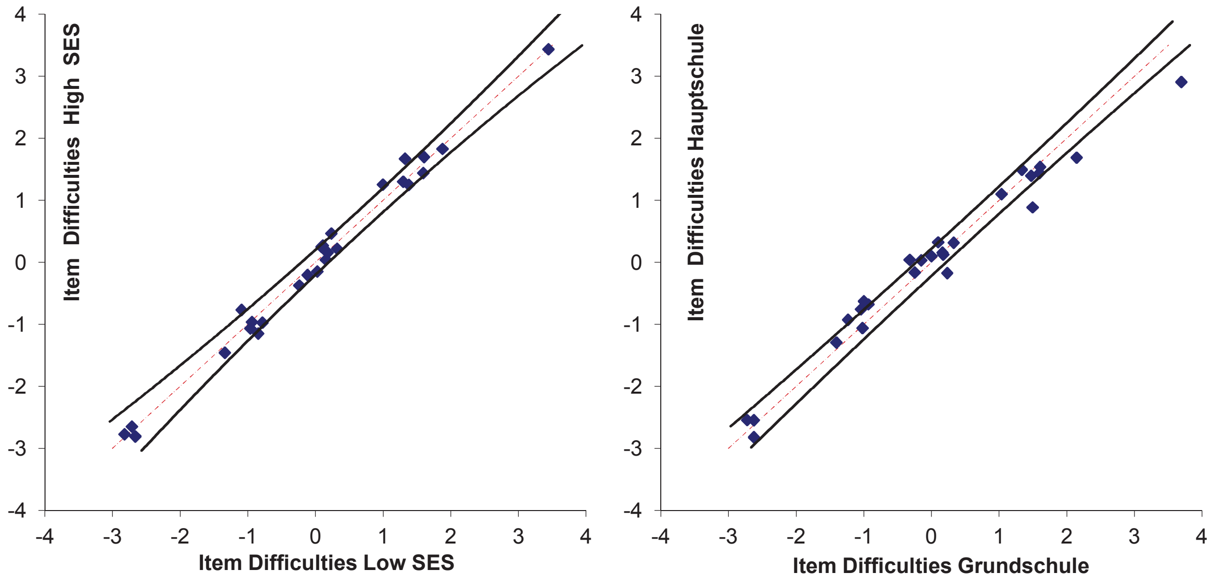

Figure 2. Plots of common item linkings for the grouping variables of socio-economic status and school type (Grundschule vs. Hauptschule) with control lines following Wright and Stone (1979).

clustered unsystematically around the identity line and none of the pairs fell prominently outside the control lines, i.e. the separate calibrations yielded similar item difficulties. The juxtaposition of calibrations for students from the school types of Grundschule and Hauptschule, however, disclosed relevant discrepancies between the groups. Four pairs of item difficulties fell into the area outside the control lines and correspondingly indicated estimations of lowered item difficulties for students from the school type of Hauptschule as compared to for students from the school type of Grundschule.

Table 7
Differential Item Functioning in Terms of Logit Differences for the Culture Fair Intelligence Test

Item	Difference in Difficulty							
	Boys – Girls	Loscore – Hiscore	Loses – Hises	Migno – Migyes	Fourth – Sixth	Grund – Haupt	Grund – Gymna	Haupt – Gymna
cft_ps1	0.328	***0.706***	-0.078	-0.100	0.076	0.190	-0.032	-0.220
cft_ps2	-0.006	***1.498***	-0.060	-0.018	-0.026	-0.186	0.366	**0.554**
cft_ps3	0.186	**0.498**	0.134	-0.096	0.012	-0.088	0.254	0.338
cft_ps4	0.192	-0.388	0.096	-0.030	**-0.474**	-0.314	**-0.634**	-0.320
cft_ps5	0.026	0.328	0.110	-0.404	-0.004	-0.118	0.190	0.308
cft_ps6	0.134	0.348	0.294	-0.080	-0.056	-0.376	**0.438**	***0.816***
cft_ps7	-0.046	-0.056	0.014	0.200	-0.180	-0.288	-0.032	0.256
cft_ps8	0.066	***-0.678***	-0.236	0.026	-0.248	0.046	**-0.514**	**-0.562**
cft_ps9	0.274	0.066	0.128	-0.102	0.060	-0.088	0.204	0.292
cft_ps10	-0.148	-0.302	-0.168	0.214	-0.266	-0.104	**-0.428**	-0.324
cft_ps11	-0.108	-0.234	0.080	0.148	-0.288	-0.364	-0.232	0.134
cft_ps12	-0.136	0.126	0.166	0.012	-0.114	-0.188	-0.064	0.126
cft_ps13	0.122	-0.210	-0.008	0.012	0.070	-0.160	0.140	0.300
cft_ps14	0.228	-0.096	0.150	0.182	0.214	0.066	0.232	0.166
cft_ps15	0.290	-0.296	0.046	-0.270	**0.636**	**0.454**	***0.660***	0.204
cft_tp1	-0.172	0.216	0.178	0.030	-0.240	-0.262	-0.212	0.050
cft_tp2	0.042	-0.362	-0.336	0.352	-0.210	0.036	**-0.450**	**-0.486**
cft_tp3	-0.198	0.092	0.128	-0.138	0.292	0.076	0.358	0.284
cft_tp4	-0.192	0.192	0.108	-0.204	0.078	0.002	0.118	0.116
cft_tp5	-0.248	**0.446**	0.096	-0.120	0.094	0.010	0.134	0.126
cft_tp6	0.138	0.156	0.016	-0.058	-0.198	-0.228	-0.202	0.026
cft_tp7	0.182	-0.286	-0.262	0.102	-0.150	-0.064	-0.264	-0.202
cft_tp8	0.018	-0.354	-0.168	0.070	0.130	0.402	-0.130	**-0.534**
cft_tp9	-0.174	-0.126	-0.096	0.048	0.030	0.156	-0.116	-0.272
cft_tp10	-0.212	**-0.572**	-0.348	0.228	0.150	**0.606**	-0.224	***-0.834***
cft_tp11	**-0.584**	***-0.716***	0.010	-0.008	**0.616**	***0.786***	**0.444**	-0.344

Note. Calculations of differential item functioning were based on interaction models comprising main effects for item and the respective grouping variable as well as the interaction between the grouping variable and item. Bold print highlights intermediate, bold and italic print signifies large differential item functioning. Loscore/Hiscore = low/high raw score on the scale; Loses/Hises = low/high socio-economic status; Migno/Migyes = non-immigrant/immigrant background; Fourth/Sixth = fourth/sixth grade; Grund = Grundschule; Haupt = Hauptschule; Gymna = Gymnasium.

Eventually, the issue of differential item functioning was tackled in a model-based approach (cf. Wu et al., 2007, chap. 8). On the one hand, for statistical inference with regard to the general presence of differential item functioning in the test, for each grouping variable the overall fit of two nested models was compared to each other by means of the likelihood ratio test, i.e. models containing only main effects for the items and the respective grouping variables, henceforth denoted main-effects-only models, served as a baseline for the evaluation of models comprising additionally a term for the interaction between the items and the respective grouping variables, henceforth called interaction models. Specifically, this model-based approach allowed the investigation of differential item functioning under simultaneous control of mean differences between student groups – via the main effect for the grouping variables – in a common framework (cf. Paek & Wilson, 2011; Wilson, 2005). On the other hand, to assess the practical relevance of differential item functioning, parameter estimates from interaction models were used to calculate differences in estimated difficulty between groups of respondents as effect sizes of differential item functioning for individual items (Paek, 2002; Paek & Wilson, 2011).

With regard to the grouping variable of gender, comparison of the two nested models revealed a significantly better fit of the interaction model, $\chi^2(25) = 92.47$, $p < .001$. Via inspection of differences in estimated difficulty between boys and girls, one item (cft_tp_11) was found to display intermediate differential item functioning in favor of boys (see Table 7). Similarly, for the grouping variable of raw score of fluid ability the likelihood ratio test identified the interaction model as fitting better to the data $\chi^2(25) = 265.29$, $p < .001$. Here four items displayed large differential item functioning and three items showed intermediate differential item functioning. Notably, the first three items, which were also the easiest items (cft_ps1, cft_ps2, cft_ps3), showed intermediate and large differential item functioning in favor of high-scoring students. In contrast, the last two items, one of them was the hardest item (cft_tp10, cft_tp11), exhibited intermediate and large differential item functioning in favor of low-scoring students. In other words, low-scoring students performed worse than expected due to their overall fluid ability at the beginning of the test and better than expected due to their overall fluid ability at the end of the test. Specific warm-up problems and successful guessing towards the end of the test might explain this pattern. In addition, for items with extremely low or high solution rates, few unexpected responses are sufficient to distort estimations of item difficulty. However, as one item showed intermediate and two items showed large differential item functioning in favor of low-scoring students while two items displayed intermediate and two items displayed large differential item functioning in favor of high-scoring students, the net impact of differential item functioning on the estimation of person ability was almost balanced across student groups.

Moreover, the interaction model was found to fit best to the data with respect to both the grouping variable of socio-economic status, $\chi^2(25) = 55.73$, $p < .001$, and the grouping variable of immigrant background, $\chi^2(25) = 38.11$, $p = .045$. Nevertheless, for socio-economic status as well as for immigrant background as the grouping variable differential item functioning of all items was negligible. In case of the grouping variable of grade level the interaction model fitted the data better than the main-effects-only model, $\chi^2(25) = 148.23$, $p < .001$. Specifically, one item displayed intermediate differential item functioning in favor of fourth-graders, whereas two items displayed intermediate differential item functioning in favor of sixth-graders. In the course of the examination of differential item functioning for students from the school types of Grundschule and Hauptschule, the interaction model was found to fit the data best, $\chi^2(25) = 113.86$, $p < .001$. Three items exhibited intermediate differential item functioning in favor of students from the school type of Hauptschule. Had all items of the test possessed differential item functioning like these three items, the mean ability of students from the school type of Hauptschule would have been inflated by 0.615 logits relative to the mean ability of students from the school type of Grundschule (cf. Wu et al., 2007, chap. 8). However, only three out of a total of twenty-six items showed differential item functioning of intermediate size in favor of students from the school type of Hauptschule. The inspection of data of students from the school types of Grundschule and Gymnasium entailed the finding that the interaction model fitted the data better than the main-effects-only model, $\chi^2(25) = 181.93$, $p < .001$. In particular, four items exhibited intermediate differential item functioning in favor of students from the school type of Grundschule, whereas two items showed intermediate and one item showed large differential item functioning in favor of students from the school type of Gymnasium, an almost balanced distribution of influences of differential item functioning. Last but not least, for the investigation of responses of students from the school types of Hauptschule and Gymnasium, the interaction model was found to fit the data best as well, $\chi^2(25) = 170.95$, $p < .001$. On the one hand, three items showed intermediate and one item showed large differential item functioning in favor of students from the school type of Hauptschule. On the other hand, one item possessed intermediate and one item possessed large differential item functioning in favor of students from the school type of Gymnasium. Thus, the distribution of effects of differential item functioning across student groups was comparatively unbalanced.

Summary. Overall, with regard to popular conventions (e.g., Bond & Fox, 2007), item fit values indicated excellent fit to the simple Rasch model (Rasch, 1960). Amounting to 0.70, the separation reliability of measures of person ability was found to be adequate. Analyses in the framework of classical test theory supported this general impression. Nevertheless, though decontextualized reasoning tasks undeniably share common variance in the assessment of fluid ability (Carroll, 1993; Cattell, 1987; but see Johnson & Bouchard, 2005 for an alternative conceptualization), the first component of the principal component analysis of the two subtests of the CFT 20-R (Weiß, 2006) explained a comparatively small portion of the total variance and the first eigenvalue was twice as large as the second eigenvalue. Despite this hint at the presence of modest multidimensionality, the unidimensional measure of person ability derived here was transferred to the next steps of data analysis in order to avoid the complexities of latent variable modeling with subscales as parcels and to maintain congruence of measures of the potential mediators in the subsequent analyses. Due to the application of the simple Rasch model (Rasch, 1960) all individual items contributed with equal weight to the estimation of person ability.

In agreement with the claim of culture fair assessment of intelligence, differential item functioning for the grouping variables of gender, socio-economic status and immigrant background was negligible in practical terms. In contrast, for the raw score of fluid ability as the grouping variable seven items displayed non-negligible differential item functioning. However, attributable to extreme solution rates for the easiest and hardest items, the effects of differential item functioning were balanced across the groups of low- and high-scoring students. With respect to the grouping variable of school type, the explorations revealed a tendency towards slight inflation of estimation of person ability for students from the school type of Hauptschule. Finally, the CFT 20-R representing an established and, in both research and practice, widely used instrument, the results of the preceding calibration and analyses might be viewed, besides delineating thoroughly the properties of the measures of fluid ability, as a frame of reference for the judgment of the quality of other instruments.

Self-concept of Ability in Science

The scale assessing students' self-concept of ability in science consisted of seven items. Employing the response categories *not at all*, *a little*, *almost* and *exactly*, these items featured a 4-point Likert-scale (see Blumberg, 2008 for a precursor of the scale). Students were instructed to think of acoustics, magnetism and optics when answering the items. As students were not necessarily familiar with those topics, instructions included brief descriptions of their characteristics. The items comprised the absolute evaluation of competence in science ("I am good at these topics." [sca1]; "I know a lot about these topics." [sca3]; "I can answer questions concerning these topics most of the time already on my own." [sca5]), the perceived ease of understanding science ("Often I do not understand such topics correctly." [sca2]; "These topics are too difficult for me." [sca4]; "Understanding such topics is easy for me." [sca6]), as well as the perception of relative science competence in comparison to peers

("With regard to such topics, I am among the best students in my class." [sca7]). Thus, by and large, the scale covered the same aspects of self-concept of ability as the Academic Self Description Questionnaire I (Byrne, 1996; Marsh, 1990b). However, intra-individual comparisons across content domains were neither explicitly addressed in the items nor implicitly suggested by presentation of self-concept scales for other areas but science (Harter, 1982; Marsh, 1990b). Prior to test analysis raw scores between 0 (*not at all*) and 3 (*exactly*) were assigned to students' responses. For the two negatively worded items (sca2, sca4) the assignment of raw scores was reversed.

Scale analysis. The calibration of the scale for students' self-concept of ability in science was based on the Rasch rating scale model (Andrich, 1978, 2010). Accordingly, for each item a mean location on the latent continuum of attitude strength as well as a step structure common to all items were retrieved from the data (see Table 8). In order to determine the precise location of a specific category boundary it is necessary to add the respective estimates of mean item location and step value. The resulting value of the specific category boundary represents the amount of attitude strength needed by a person to score with a probability of 50% on that step for that item. For instance, a student with a value of attitude strength of 0.339 logits on the latent continuum of self-concept of ability in science had a probability of 50% to respond in the category *almost*, i.e. to achieve a raw score of 2, on the third item (sca3).

Table 8

Probabilistic and Traditional Item Analyses for the Scale of Self-concept of Ability in Science

Parameter	Estimate	SE	Infit	Outfit	*M*	r_{it}
			Items			
sca1 (good at)	0.196	0.028	0.71	0.72	1.79	.73
sca2 (not corr.)[a]	-0.626	0.030	1.37	1.39	2.14	.50
sca3 (know)	0.504	0.028	0.83	0.83	1.66	.70
sca4 (too diff.)[a]	-1.504	0.035	1.38	1.34	2.44	.50
sca5 (answer)	0.242	0.028	0.94	0.94	1.77	.65
sca6 (easy)	0.022	0.029	0.85	0.85	1.87	.71
sca7 (among)	1.166[b]		1.14	1.12	1.36	.65
			Steps			
1 (little)	-1.682	0.029	1.27	1.17	---	---
2 (almost)	-0.165	0.023	1.40	1.36	---	---
3 (exactly)	1.847[b]		1.30	1.66	---	---

Note. Abbreviated item wordings and response category labels are given in parentheses. Diff. = difficult; corr. = correctly.

[a]Scoring reversed. [b]Parameter estimate is constrained.

The results of the calibration summarized responses from 2486 students. With mean item locations of -1.504 logits and -0.626 logits, the negatively worded items (sca2, sca4) were particularly easy to reject. At the opposite end of the latent continuum of self-concept of ability in science, the item calling for interpersonal comparisons (sca7) was the hardest to endorse. Its mean item location was 1.166 logits. Apparently, self-attributions of non-competence and competence were not symmetric, i.e. it was easier to deny non-competence than to approve competence. Infit values for mean item locations ranged from 0.71 to 1.38. Similarly, corresponding outfit values varied between 0.72 and 1.39. Notably, the negatively worded items (sca2, sca4) displayed, with some margin, the worst infit and outfit values of all items. With estimated step values of -1.682 logits, -0.165 logits and 1.847 logits, the step structure common to all items gave evidence for clearly separated response categories. The infit values for the step structure varied between 1.27 and 1.40, whereas the associated outfit values ranged from 1.17 to 1.66 (see Table 8).

In order to enable model identification, the distribution of item locations had a predefined mean of 0.00 logits. The corresponding standard deviation amounted to 0.85 logits. Note, however, that the distribution of actual category boundaries defined by a combination of mean item locations and step values spanned a wider range of the latent continuum of attitude strength than the distribution of mean item locations. The distribution of category boundaries, or, expressed in the terminology used within ACER ConQuest 2.0, of item deltas, had a predefined mean of 0.00 logits and a standard deviation of 1.69 logits. On the person side, the mean of the distribution of attitude strength was 0.83 logits with a standard deviation of 1.51 logits. The separation reliability of the weighted likelihood estimates of attitude strength was 0.83.

Last but not least, indices from the framework of classical test theory were computed for the scale. In correspondence to the results of the Rasch rating scale analysis inspection of the mean raw scores of the items revealed that the two negatively worded items (sca2, sca4) were particularly easy to reject (see Table 8). Item discriminations, i.e. corrected item-total correlations, fell between $r_{it} = .50$ and $r_{it} = .73$. The two negatively worded items (sca2, sca4) had the lowest item discriminations, $r_{it} = .50$ for both, of all items. In order to investigate the dimensionality of the scale, students' responses were submitted to a principal component analysis. The first component explained 55.10% of the total variance. The eigenvalue of the first component amounted to 3.86. No other component had an eigenvalue greater than 1. The ratio of the first and second eigenvalues was 3.97. The computation of Cronbach's alpha returned a value of .86 for the internal consistency of the scale.

Differential item functioning. For each grouping variable, the model-based inferential analysis of differential item functioning comprised comparisons of four nested models. Main-effects-only models containing main effects for items, steps and grouping variables in the fixed part of the model formulation constituted the baseline for these model comparisons. On the one hand, the baseline models were compared with models additionally incorporating interactions between items and grouping variables. Henceforth, these models are called item interaction models. The item interaction models assumed that different subgroups of students used the scheme of response categories in the same way whereas the mean location of items was set free to vary between different subgroups. On the other hand, the baseline models were compared with models additionally covering interactions between steps and grouping variables. Henceforth, these models are called step interaction models. The step interaction models presumed that the mean location of items was identical for different subgroups of students whereas the use of the scheme of response categories was allowed to differ between subgroups. Thus, step interaction models were capable of capturing the presence of response tendencies, such as leniency or a tendency towards the scale center, in specific subgroups of respondents. Eventually, models containing interactions between items and grouping variables as well as interactions between steps and grouping variables – henceforth these models are denoted full interaction models – were evaluated both against item interaction models and against step interaction models (see Figure 3). In order to retrieve effect sizes for differential item functioning, parameter estimates from full interaction models were used to compute logit differences in attitude strength between subgroups for all individual category boundaries.

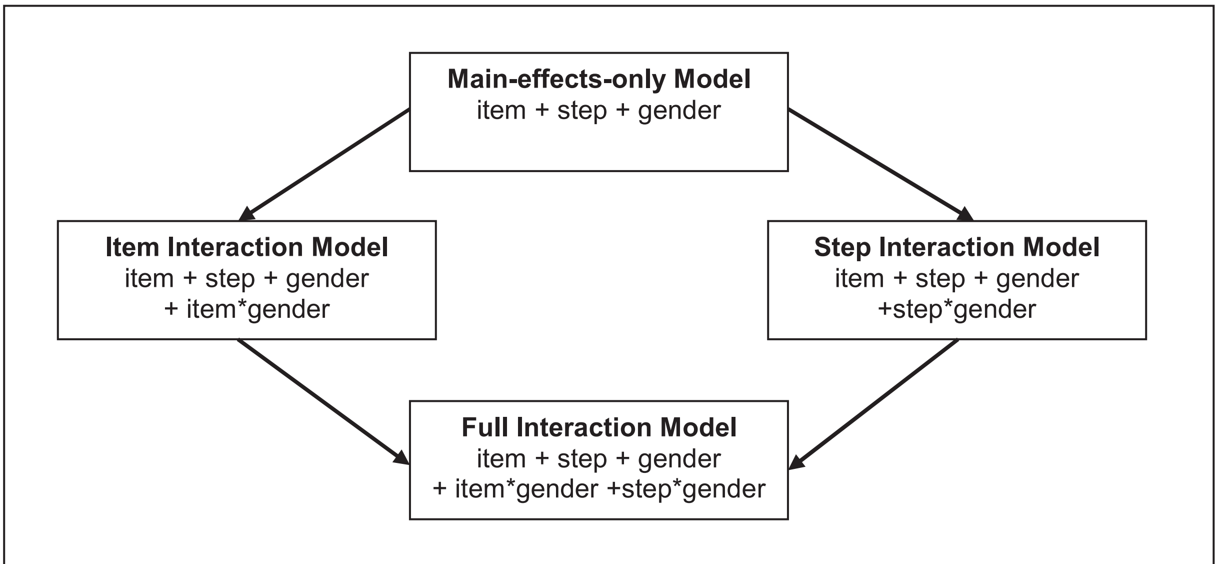

Figure 3. Overview of model comparisons for the inferential analysis of differential item functioning in the scale of self-concept of ability in science for the grouping variable of gender.

Data missing both for the grouping variables and for the scale of self-concept of ability, analyses of differential item functioning made use of groups of students deviating to varying extent from the entire sample of 2680 students. The inspection of differential item functioning with regard to the grouping variable of gender took responses from 1321 boys and 1164 girls into consideration. The item interaction model showed better fit than the main-effects-only model, $\chi^2(6) = 103.72$, $p < .001$, and the full interaction model showed better fit than the item interaction model , $\chi^2(4) = 18.76$, $p = .001$. Thus, the model comparisons returned the full interaction model as fitting best to the data (see Table 9). Three category boundaries displayed intermediate and two category boundaries displayed large differential item functioning (see Table 10). For boys it was more difficult than for girls to pass the first category boundary of the negatively worded items (sca2, sca4). In other words, here girls avoided – stronger than expected in face of their overall self-concept of ability – selecting the response category *exactly*. (Remember that the scoring for negatively worded items was reversed.) In contrast, it was easier for boys than for girls to ascribe themselves high competence in comparison to their peers (sca7). Girls refrained from attributing themselves superior competence in social comparisons. However, differential item functioning with regard to gender was approximately balanced, with three and two specific category boundaries favoring boys and girls, respectively.

Table 9
Model Comparisons for Inferential Analyses of Differential Item Functioning for the Scale of Self-concept of Ability in Science

Grouping Variable	Comparison MEOM – IIM			Comparison MEOM – SIM			Comparison IIM – FIM			Comparison SIM – FIM		
	χ^2	*df*	*p*	χ^2	*df*	*p*	χ^2	*df*	*p*	χ^2	*df*	*p*
Gender	103.72	6	<.001	1.42	4	.841	18.76	4	.001	121.06	6	<.001
SCA-score	71.31	6	<.001	27.92	4	<.001	32.09	4	<.001	75.48	6	<.001
SES	51.36	6	<.001	0.47	4	.976	10.53	4	.032	61.42	6	<.001
Immigration	54.33	6	<.001	1.03	4	.905	2.51	4	.643	55.81	6	<.001
Grade	71.72	6	<.001	25.03	4	<.001	6.17	4	.187	52.86	6	<.001
GS vs. HS	20.24	6	.003	15.81	4	.003	14.47	4	.006	18.90	6	.004
GS vs. GYM	115.45	6	<.001	20.04	4	<.001	0.12	4	.998	95.53	6	<.001
HS vs. GYM	61.16	6	<.001	3.79	4	.435	10.70	4	.030	68.07	6	<.001

Note. MEOM = main-effects-only model; IIM = item interaction model; SIM = step interaction model; FIM = full interaction model; SCA = self-concept of ability in science; SES = socio-economic status; GS = Grundschule; HS = Hauptschule; GYM = Gymnasium.

The investigation of differential item functioning for the grouping variable of raw score of self-concept of ability relied on responses from 1243 low-scoring and 1243 high-scoring students. The inferential model comparisons via likelihood ratio tests established that the full interaction model possessed the best fit to the data. Examination of logit differences between the two student groups – obtained from the full interaction model – for the location of specific category boundaries revealed that five category boundaries exhibited intermediate differential item functioning, while two category boundaries exhibited large differential item functioning. Of these category boundaries, three displayed intermediate and one displayed large differential item functioning in favor of low-scoring students, whereas the remaining three category boundaries were advantaging high-scoring students. So, save one category boundary possessing intermediate differential item functioning, the distribution of effects of differential item functioning across the groups of low- and high-scoring students appeared as practically balanced.

With regard to socio-economic status as the grouping variable, data of 881 students with low socio-economic status and 913 students of high socio-economic status was examined. As the item interaction model displayed better fit than the main-effects-only model, $\chi^2(6) = 51.36$, $p < .001$, and the full interaction model displayed better fit than the item interaction model, $\chi^2(4) = 10.53$, $p = .032$, in the course of the model comparisons by likelihood ratio tests the full interaction model was found to fit best to the data. Specifically, one category boundary exhibited differential item functioning of intermediate size in favor of students with low socio-economic status, whereas three category boundaries exhibited intermediate differential item functioning in favor of students with high socio-economic status.

The exploration of differential item functioning with respect to immigrant background incorporated responses of 1308 students without immigrant background, i.e. both parents were born in Germany, and 466 students with immigrant background, i.e. at least one parent was born abroad. In the inferential analyses, the inclusion of the interaction between steps and immigrant background did not improve overall model fit, neither in the comparison of the main-effects-only model with the step interaction model, $\chi^2(4) = 1.03$, $p = .905$, nor in the comparison of the item interaction model with the full interaction model, $\chi^2(4) = 2.51$, $p = .643$. In contrast, the item interaction model displayed better overall model fit than the main-effects-only model, $\chi^2(6) = 54.33$, $p < .001$. However, merely one specific category boundary showed non-negligible differential item functioning. It was of intermediate size.

The examination of differential item functioning for the grouping variable of grade level relied on responses from 1239 fourth- and 1247 sixth-graders. The model comparisons disclosed the full interaction model as the best fitting model as the step interaction model possessed better model fit than the main-effects-only model, $\chi^2(4) = 25.03$, $p < .001$, and the full interaction model possessed better model fit than the step interaction model, $\chi^2(6) = 52.86$, $p < .001$. Nevertheless, only one specific category boundary displayed differential item functioning of intermediate size in favor of sixth-graders.

The investigation of differential item functioning for school type as the grouping variable comprised three pairwise comparisons. It used data of 1239 students from the school type of Grundschule, 530 students from the school type of Hauptschule, and 717 students from the school type of Gymnasium. For the comparison of students from the school type of Grundschule with students from the school type of Hauptschule, inferential analyses of overall model fit identified the full interaction model as the best fitting model. Nevertheless, for this comparison none of the category boundaries showed differential item functioning of non-negligible size.

Table 10
Differential Item Functioning in Terms of Logit Differences for the Scale of Self-concept of Ability in Science

Item	Step	Difference in Difficulty							
		Boys – Girls	Loscore – Hiscore	Loses – Hises	Migno – Migyes	Fourth – Sixth	Grund – Haupt	Grund – Gymna	Haupt – Gymna
sca1	1	0.156	**-0.559**	-0.044	0.041	-0.043	0.048	-0.112	-0.161
	2	-0.124	0.183	-0.129	0.170	-0.098	-0.042	-0.149	-0.101
	3	-0.284	-0.139	-0.386	0.256	0.093	0.336	-0.130	**-0.458**
sca2[a]	1	**0.532**	***-0.679***	**0.504**	-0.399	0.313	0.068	**0.542**	**0.459**
	2	0.252	0.063	0.419	-0.270	0.258	-0.022	**0.505**	**0.519**
	3	0.092	-0.259	0.162	-0.184	**0.449**	0.356	**0.524**	0.162
sca3	1	-0.014	**-0.507**	-0.074	0.133	0.009	0.080	-0.046	-0.129
	2	-0.294	0.235	-0.159	0.262	-0.046	-0.010	-0.083	-0.069
	3	**-0.454**	-0.087	-0.416	0.348	0.145	0.368	-0.064	**-0.426**
sca4[a]	1	***0.686***	-0.153	**0.564**	**-0.529**	0.137	-0.206	**0.468**	***0.661***
	2	0.406	**0.589**	**0.479**	-0.400	0.082	-0.296	**0.431**	***0.721***
	3	0.246	0.267	0.222	-0.314	0.273	0.082	**0.450**	0.364
sca5	1	0.256	**-0.509**	-0.130	0.201	-0.153	-0.018	-0.260	-0.241
	2	-0.024	0.233	-0.215	0.330	-0.208	-0.108	-0.297	-0.181
	3	-0.184	-0.089	**-0.472**	0.416	-0.017	0.270	-0.278	**-0.538**
sca6	1	0.350	-0.257	0.112	-0.099	-0.159	-0.216	-0.106	0.105
	2	0.070	**0.485**	0.027	0.030	-0.214	-0.306	-0.143	0.165
	3	-0.090	0.163	-0.230	0.116	-0.023	0.072	-0.124	-0.192
sca7	1	-0.286	-0.047	0.062	-0.153	-0.291	-0.216	-0.358	-0.141
	2	**-0.566**	***0.695***	-0.023	-0.024	-0.346	-0.306	-0.395	-0.081
	3	***-0.726***	0.373	-0.280	0.062	-0.155	0.072	-0.376	**-0.438**

Note. Calculations of differential item functioning were based on full interaction models comprising main effects for item, step and the respective grouping variable as well as interactions of the grouping variable with item and step. Bold print highlights intermediate, bold and italic print signifies large differential item functioning. Loscore/Hiscore = low/high raw score on the scale; Loses/Hises = low/high socio-economic status; Migno/Migyes = non-immigrant/immigrant background; Fourth/Sixth = fourth-/sixth grade; Grund = Grundschule; Haupt = Hauptschule; Gymna = Gymnasium.
[a]Scoring reversed.

The step interaction model displaying better model fit than the main-effects-only model, $\chi^2(4) = 20.04$, $p < .001$, and the full interaction model displaying better model fit than the step interaction model, $\chi^2(6) = 95.53$, $p < .001$, the full interaction model appeared as the best fitting model for the combined data of students from the school types of Grundschule and Gymnasium. Inspection of estimated logit differences unveiled intermediate differential item functioning in favor of students from the school type of Gymnasium for six category boundaries. These six category boundaries were associated with the two negatively worded items (sca2, sca4). In face of the strength of their overall self-concept of ability in science, students from the school type of Grundschule were seemingly less hesitant than students from the school type of Gymnasium to ascribe themselves non-competence in science. In addition, or alternatively, it is plausible that students from the school type of Grundschule were more likely to miss the negative orientation of items. Had all category boundaries exhibited differential item functioning of the size of the category boundaries of the negatively worded items (sca2, sca4), that would have depressed – relative to students from the school type of Gymnasium – the estimated mean attitude strength of students from the school type of Grundschule by 0.487 logits, slightly less than one third of the standard deviation of the person distribution of attitude strength (cf. Wu et al., 2007, chap. 8) However, only six out of twenty-one category boundaries displayed differential item functioning of intermediate size.

In the analyses of students from the school types of Hauptschule and Gymnasium, the item interaction model exhibited better model fit than the main-effects-only model, $\chi^2(6) = 61.16$, $p < .001$, and the full interaction model exhibited better model fit than the item interaction model, $\chi^2(4) = 10.70$, $p = .030$. Thus, the full interaction model was found to fit the data best. With regard to the size of logit differences between groups, two category boundaries showed large differential item functioning and six category boundaries showed intermediate differential item functioning. Notably, those four category boundaries exhibiting intermediate differential item functioning in favor of students from the school type of Hauptschule were all located at the transition from the response category *almost* to the response category *exactly*. In other words, for four items (sca1, sca3, sca5, sca7) students from the school type of Hauptschule were more likely to select the response category *exactly* than expected according to their overall self-concept of ability in science. On the contrary, those four category boundaries exhibiting intermediate and large differential item functioning in favor of students from the school type of Gymnasium were located at the first two transitions of the negatively worded items (sca2, sca4). As four category boundaries displayed differential item functioning in favor of students from the school type of Hauptschule and four category boundaries displayed differential item functioning in favor of students from the school type of Gymnasium, the net effects of differential item functioning – setting aside the differentiation of large and intermediate differential item functioning –on the extraction of person estimates of self-concept ability in science roughly canceled out each other.

Summary. Infit values for items and steps indicated very good conformity with the Rasch rating scale model (Fisher, 2007). Notably, the negatively worded items had the worst fit values of all items. Clearly, an instantiation of the disturbances often introduced by reversely coded items into measurements. These disturbances were also reflected by the comparatively low item discriminations of the negatively worded items. The distributions of person attitude strength and category boundaries were shifted against each other. The mean of the distribution of person attitude strength exceeded the mean of the distribution of category boundaries by approximately half of the corresponding standard deviations. In other words, the scale lacked some items that were hard to endorse. Nevertheless, the weighted likelihood estimates of attitude strength possessed sufficient reliability. The results of the principal components analysis argued in favor of unidimensionality of the scale.

The results of the explorations of differential item functioning spawned some speculation about the impact of group-specific response tendencies – partly involving the negatively worded items – on the estimation of person attitude strength. For the grouping variable of gender, girls displayed a certain tendency towards the center of the scale. On the one hand, they shunned the report of extreme non-competence on the negatively worded items. On the other hand, they avoided the self-ascription of competence in interpersonal comparisons. So, operating in opposite directions on the different items, the effects of this tendency towards the center appeared to cancel out each other in the estimation of person attitude strength. Similarly, effects of differential item functioning were approximately balanced for the grouping variable of raw score of self-concept of ability in science. When socio-economic status was considered as the grouping variable, net effects disadvantaging students of low socio-economic status equaled differential item functioning of intermediate size for two category boundaries. For the grouping variables of immigrant background and grade level differential item functioning of intermediate size was found for one category boundary, respectively.

With regard to the grouping variable of school type, the occurrence of non-negligible differential item functioning was restricted to those comparisons that involved the school type of Gymnasium. In particular, students from the school type of Gymnasium were more hesitant than expected due to their overall self-concept of ability in science to ascribe themselves non-competence than students from the two other school types. This response tendency implicated a small depression of estimates of person attitude strength for students from the school type of Grundschule relative to students from the school of Gymnasium. In case of the comparison with students from the school type of Hauptschule, the relative unwillingness of students from the school type of Gymnasium to endorse the negatively worded items was counterbalanced by a comparatively progressive use of the response category *exactly* for four of the unreversed items.

Individual Interest in Science

Measurement of students' individual interest in science made use of five items. These items featured the same four response categories as the items assessing students' self-concept of ability in science (again see Blumberg, 2008 for a precursor of the scale). Moreover, in further correspondence with the assessment of self-concept of ability, students were reminded to consider the topics of acoustics, magnetism and optics when responding to the items. The items were concerned with feeling-related ("To occupy myself with these topics is a lot of fun." [int5]) and value-related aspects ("I am eager to get to know more about these topics." [int4]) of individual interest. Moreover, the experience of flow ("When I am occupied with these topics, I forget everything around me." [int3]) and behavioral manifestations of interest ("At home I often read about these topics." [int1]; "I am delighted, when something about these topics is broadcasted in television." [int2]) were covered. All items addressed the positive affective tone that accompanies interest-driven behaviors. In this respect, a thoughtful look at the item content reveals that the scale assessing students' individual interest in science might be denoted with equal justification a scale measuring academic intrinsic motivation in science (see Gottfried, 1985 for items intended to assess academic intrinsic motivation).

Scale analysis. In correspondence to the scale of self-concept of ability in science, the scale assessing students' individual interest in science was calibrated according to the Rasch rating scale model (Andrich, 1978, 2010). The calibration of the scale of individual interest in science relied on data of 2484 students. Mean item locations ranged from -0.580 logits to 0.829 logits on the latent continuum of individual interest in science. At one end of the continuum, the items concerned with the behavior of reading (int1) and the experience of flow (int3) constituted a pair of items that was difficult to affirm, whereas, at the other end of the continuum, the remaining items (int2, int4, int5) formed a cluster of items relatively easy to endorse. Infit values associated with mean item locations varied between 0.84 and 1.17, while corresponding outfit values covered an almost identical range from 0.83 to1.18. The step structure common to all items featured step values of -1.313 logits, 0.142 logits and 1.171 logits, thereby indicating clearly separated transitions from one response category to the next, although the estimated range of the step structure was somewhat smaller than for the scale of self-concept of ability in science. The infit values related to the step structure fell between 1.11 and 1.13. The corresponding outfit values ranged from 1.06 to 1.51 (see Table 11).

Table 11
Results of Probabilistic and Traditional Item Analyses for the Scale of Individual Interest in Science

Parameter	Estimate	SE	Infit	Outfit	*M*	r_{it}
			Items			
int1 (read)	0.829	0.027	1.02	1.05	1.07	.56
int2 (television)	-0.433	0.026	1.11	1.05	1.72	.66
int3 (forget)	0.705	0.026	1.17	1.18	1.13	.58
int4 (eager)	-0.521	0.026	0.90	0.89	1.76	.70
int5 (fun)	-0.580[a]		0.84	0.83	1.79	.72
			Steps			
1 (little)	-1.313	0.029	1.11	1.09	---	---
2 (almost)	0.142	0.027	1.11	1.06	---	---
3 (exactly)	1.171[a]		1.13	1.51	---	---

Note. Abbreviated item wordings and response category labels are given in parentheses. Diff. = difficult; corr. = correctly.
[a]Parameter estimate is constrained.

Its mean fixed for the purpose of identification to a value of 0.00 logits prior to model estimation, the distribution of item locations displayed a standard deviation of 0.70 logits. The corresponding distribution of category boundaries, or item deltas, possessed a predefined mean of 0.00 logits and a standard deviation of 1.24 logits. The distribution of person attitude strength had a mean of -0.02 logits and a standard deviation of 1.48 logits. The separation reliability pertaining to the weighted likelihood estimates of person attitude strength amounted to 0.78.

As the final step of the scale analysis, statistics from the framework of classical test theory were inspected. Naturally, the arrangement of mean raw scores obtained for the items resembled the collocation of mean item locations retrieved from the application of the Rasch rating scale model. Note in this context that higher raw scores imply that items are comparatively easy to endorse, which is reflected by smaller values for item locations on the logit scale of the Rasch rating scale model. Item discriminations attained values between r_{it} = .56 and r_{it} = .72 (see Table 11). The first component of a principal component analysis explained 61.02% of the total variance. A value of 3.05 was retrieved for the first eigenvalue, with the values of all other eigenvalues falling below 1.00. Correspondingly, the ratio of the first and second eigenvalues equaled 5.08. The internal consistency of the scale amounted to .84 in terms of Cronbach's alpha.

Differential item functioning. The calibrations for both the scale assessing self-concept of ability in science and the scale of individual interest in science were based on the Rasch rating scale model (Andrich, 1978, 2010). Accordingly, the model-based inferential analysis of differential item functioning for the scale of individual interest in science followed the pattern outlined for the scale of self-concept of ability. In particular, for each grouping variable four nested models – the main-effects-only model, the item interaction model, the step interaction model and the full interaction model – were tested against each other with respect to their fit to the data. Moreover, based on the parameter estimates of the full interaction models, logit differences in attitude strength between relevant subgroups were calculated for all specific category boundaries in order to gauge the size of differential item functioning. As for the other tests and scales, the explorations of differential item functioning made use of fluctuating samples of students because data was missing in differing degrees for the grouping variables as well as for the scale of individual interest in science.

Table 12
Model Comparisons for Inferential Analyses of Differential Item Functioning for the Scale of Individual Interest in Science

Grouping Variable	Comparison MEOM – IIM			Comparison MEOM – SIM			Comparison IIM – FIM			Comparison SIM – FIM		
	χ^2	*df*	*p*	χ^2	*df*	*p*	χ^2	*df*	*p*	χ^2	*df*	*p*
Gender	11.46	4	.022	7.16	4	.128	5.26	4	.262	9.56	4	.049
INT-score	125.11	4	<.001	9.61	4	.048	12.69	4	.013	128.19	4	<.001
SES	3.10	4	.541	8.58	4	.072	6.71	4	.152	1.23	4	.873
Immigration	11.28	4	.024	4.50	4	.343	6.54	4	.162	13.32	4	.010
Grade	20.29	4	<.001	16.18	4	.003	15.77	4	.003	19.88	4	.001
GS vs. HS	28.29	4	<.001	11.20	4	.024	12.25	4	.016	29.34	4	<.001
GS vs. GYM	23.02	4	<.001	12.99	4	.011	11.67	4	.020	21.70	4	<.001
HS vs. GYM	36.05	4	<.001	4.03	4	.402	4.61	4	.330	36.63	4	<.001

Note. MEOM = main-effects-only model; IIM = item interaction model; SIM = step interaction model; FIM = full interaction model; INT = individual interest in science; SES = socio-economic status; GS = Grundschule; HS = Hauptschule; GYM = Gymnasium.

Relying on data of 1321 boys and 1162 girls, the model-based examination of differential item functioning for the grouping variable of gender resulted in the identification of the item interaction model as fitting best to the data. The incorporation of the interaction between step and gender enhanced overall model fit neither with regard to main-effects-only model, $\chi^2(4) = 7.16$, $p = .128$, nor with regard to the item interaction model, $\chi^2(4) = 5.26$, $p = .262$ (see Table 12). None of the category boundaries of the scale exhibited differential item functioning of non-negligible size (see Table 13).

The exploration of differential item functioning for the grouping variable of raw score of individual interest in science, created by a median-split, took responses of 1242 low-scoring and 1242 high-scoring students into consideration. The model comparisons for the inferential investigation of differential item functioning established that the full interaction model possessed the best relative fit to the data. In particular, three category boundaries were found to show differential item functioning of intermediate size whereas one category boundary was found to display large differential item functioning. However, as one half of these category boundaries exhibited differential item functioning in favor of and the other half against low-scoring students, differential item functioning exerted only a very small biasing influence on the estimation of person ability with respect to the grouping variable of raw score of individual interest in science.

Table 13
Differential Item Functioning in Terms of Logit Differences for the Scale of Individual Interest in Science

Item	Step	Difference in Difficulty							
		Boys – Girls	Loscore – Hiscore	Loses – Hises	Migno – Migyes	Fourth – Sixth	Grund – Haupt	Grund – Gymna	Haupt – Gymna
int1	1	0.126	**-0.513**	0.197	-0.110	0.230	0.013	0.406	0.403
	2	0.011	-0.381	-0.040	0.088	-0.076	-0.233	0.040	0.280
	3	-0.106	***-0.787***	-0.085	0.232	0.260	0.202	0.287	0.085
int2	1	-0.036	**0.445**	0.109	-0.028	-0.120	-0.299	0.042	0.359
	2	-0.151	**0.577**	-0.128	0.170	**-0.426**	**-0.545**	-0.324	0.236
	3	-0.268	0.171	-0.173	0.314	-0.090	-0.110	-0.077	0.041
int3	1	0.152	-0.001	0.181	-0.116	0.098	0.231	-0.004	-0.249
	2	0.037	0.131	-0.056	0.082	-0.208	-0.015	-0.370	-0.372
	3	-0.080	-0.275	-0.101	0.226	0.128	0.420	-0.123	**-0.567**
int4	1	0.152	0.121	0.179	-0.270	0.106	0.047	0.172	0.127
	2	0.037	0.253	-0.058	-0.072	-0.200	-0.199	-0.194	0.004
	3	-0.080	-0.153	-0.103	0.072	0.136	0.236	0.053	-0.191
int5	1	0.186	0.183	0.199	-0.376	0.146	0.103	0.196	0.093
	2	0.071	0.315	-0.038	-0.178	-0.160	-0.143	-0.170	-0.030
	3	-0.046	-0.091	-0.083	-0.034	0.176	0.292	0.077	-0.225

Note. Calculations of differential item functioning were based on full interaction models comprising main effects for item, step and the respective grouping variable as well as interactions of the grouping variable with item and step. Bold print highlights intermediate, bold and italic print signifies large differential item functioning. Loscore/Hiscore = low/high raw score on the scale; Loses/Hises = low/high socio-economic status; Migno/Migyes = non-immigrant/immigrant background; Fourth/Sixth = fourth-/sixth grade; Grund = Grundschule; Haupt = Hauptschule; Gymna = Gymnasium.

The investigation of differential item functioning for the grouping variable of socio-economic status was based on data from 880 students of low socio-economic status and 913 students of high socio-economic status. Neither the inclusion of the interaction between item and socio-economic status, $\chi^2(4) = 3.10$, $p = .541$, nor the incorporation of the interaction between step and socio-economic status, $\chi^2(4) = 8.58$, $p = .072$, significantly improving overall fit, the main-effects-only model was identified by a sequence of model comparisons to fit the data best. In correspondence with this, in terms of logit differences in attitude strength between groups, all category boundaries of the scale displayed negligible differential item functioning.

Responses from 1309 students without immigrant background and from 464 students with immigrant background were used to explore differential item functioning for the grouping variable of immigrant background. The sequence of four model comparisons established that the item interaction model fitted best to the data. Despite this, the size of differential item functioning was negligible for all category boundaries.

The exploration of differential item functioning for the grouping variable of grade level included data of 1239 fourth-graders and 1245 sixth-graders. All four comparisons of nested models indicating preference for the more complex model, the full interaction model was found to fit best to the data. One specific category boundary displayed differential item functioning of intermediate size in favor of fourth-graders.

Three pairwise comparisons of student groups constituted the exploration of differential item functioning for the grouping variable of school type. In particular, here the investigation of differential item functioning was based on responses of 1239 students from the school type of Grundschule, 531 students from the school type of Hauptschule and 714 students from the school type of Gymnasium. With respect to the investigation of responses of students from the school types of Grundschule and Hauptschule, the full interaction model was identified as fitting best to the data. Differential item functioning of intermediate size was shown by one specific category boundary. In case of responses of students from the school types of Grundschule and Gymnasium, the full interaction model was found to fit the data best. However, none of the category boundaries of the scale displayed non-negligible differential item functioning. For the combined data of students from the school types of Hauptschule and Gymnasium, the incorporation of the interaction between step and school type (Hauptschule vs. Gymnasium) improved overall model fit neither in relation to the main-effects-only model, $\chi^2(4) = 4.03$, $p = .402$, nor in relation to the item interaction model, $\chi^2(4) = 4.61$, $p = .330$. As a consequence of this, the item interaction model was identified as fitting best to the data. One specific category boundary showed intermediate differential item functioning in favor of students from the school type of Hauptschule.

Summary. Infit values for items and steps flagged excellent fit to the Rasch rating scale model (Fisher, 2007). Despite a minor discrepancy in the respective standard deviations, the distributions of person attitude strength and category boundaries neatly matched each other. Weighted likelihood estimates of person attitude strength were sufficiently reliable. With respect to the content of the scale, the examinations revealed that experiences of flow and the reading of domain-specific books were endorsed relatively reluctantly as indicators of individual interest in science whereas general enjoyment and openness as well as watching domain-specific television programs were reported more willingly. The first component explaining more than 60% of the total variance, the results of the principal component analysis hinted at unidimensionality of the scale. The results for Cronbachs' alpha and for the item discriminations pointed into the same direction. For the grouping variables of gender, socio-economic status and immigrant background as well as for the comparison of students from the school types of Grundschule and Gymnasium all category boundaries displayed differential item functioning of negligible size. For the other grouping variables and comparisons non-negligible differential item functioning was either relatively balanced or restricted to one category boundary. In essence, it was not possible to identify a meaningful or practically relevant pattern of differential item functioning for the investigated grouping variables.

Science Achievement

The test of students' science achievement comprised 26 item stems, with 3 to 9 response alternatives nested under each item stem. For 22 item stems students had to endorse or reject response alternatives individually. For the other 4 item stems students had to select their responses directly from the corresponding set of alternatives (i.e. selection of a specific alternative implied rejection of remaining alternatives). With respect to the content covered, 12 item stems were devoted to phenomena of evaporation of water, 10 item stems were dedicated to phenomena of condensation, and 4 item stems presented the issue of phase changes in a decontexualized manner.

The test combined the measurement of two forms of knowledge. On the one hand, the test assessed the knowledge of simple facts. This encompassed, for instance, the knowledge of appropriate scientific terminology and the knowledge of conditions promoting the process of evaporation. On the other hand, the test assessed the knowledge of correct explanations for phenomena of evaporation and condensation in different situations. In particular, the measurement of this latter form of knowledge featured the presentation of common student misconceptions as distractor response alternatives. A group of 11 item stems was devoted to the assessment of the knowledge of simple facts, whereas the remaining 13 item stems covered the recognition of correct explanations.

The achievement test was administered to the participating children both before and after the series of three lessons on the topic of evaporation and condensation. All instructions, item stems and response alternatives were read out aloud to the students in their regular science classrooms. This procedure was expected to minimize the influence of individual reading ability on test performance. Furthermore, the instructions ensured that classrooms proceeded in unison through the test.

The test embodied, most prominently with regard to the 13 item stems concerned with explanations for evaporation and condensation, several potential sources of local dependencies. As already said, groups of autonomous response alternatives were nested under common item stems. Moreover, the conceptions represented in those response alternatives were repeated, albeit often in slightly different wording, across item stems. In this vein, the test contained even actual twins of item stems that shared both the guiding question within the item stem and the corresponding response alternatives in identical wording. Thus, a relatively complex scoring procedure was applied to students' responses before final test calibration. The scoring procedure served three purposes. First, it was intended to counter the potential overestimation of reliability due to local dependencies by introduction of superitems, i.e. by summarizing over possibly dependent responses (Marais & Andrich, 2008; Swygert, MacLeod, & Thissen, 2001). Second, it was devised to reduce problems of parameter estimation due to extreme response probabilities by summarizing over response alternatives with extremely high solution rates. Third, it was set up to obtain estimates of difficulty related exclusively to specific student conceptions concerning evaporation and condensation. Due to ambiguous test instructions, student responses to 2 item stems could not be scored consistently. These item stems were excluded from subsequent analyses.

Eventually, application of the scoring scheme resulted in 21 dichotomously and 12 polytomously scored items. Maxima ranging from two to three raw points could be achieved on the polytomously scored items. Correspondingly, they featured two or three category boundaries in test calibration. In particular, superitems scored across item stems covered the student conceptions that objects sweat condensed water out (sat_as_1), that water turns into or results from smoke during phase changes (sat_as_2), that evaporated water turns into oxygen (sat_as_3), that oxygen turns into condensed water (sat_as_4), that evaporated water in closed rooms concentrates near the ceiling (sat_as_7), that heat transforms into water and vice versa (sat_as_8), that evaporated water ceases to exist (sat_as_9), that light creates condensed water (sat_as_10), that coldness turns into condensed water (sat_as_11), that evaporated water turns into steam (sat_as_12), that the sun directly attracts water (sat_as_14), and that water completely emerges from or goes to hidden places, e.g. is soaked in completely by the soil (sat_as_15).

Wenn es sehr heiß ist, ist die Erde an manchen Stellen ganz hart und rissig. Die Erde ist aus-getrocknet. Das Wasser, das vorher in der Erde war, ist nicht mehr dort. Was ist mit dem Wasser passiert?
[When it is very hot, in some places the soil is all hard and cracked. The soil is dried out. The water that previously has been in the soil is not there anymore. What has happened to the water?]

Das fehlende Wasser ist jetzt… [The missing water is now…]
Kreuze nach jedem Wort „richtig“ oder „falsch“ an! [Check “correct” or “false” after each word!]

	Richtig [Correct]	Falsch [False]
…Nichts. […nothing.]	☐	☐
…nicht sichtbarer Sauerstoff in der Luft. […invisible oxygen in the air.]	☐	☐
…Rauch. […smoke.]	☐	☐
…Wasserdampf. […water vapor.]	☐	☐
…nicht sichtbar in der Luft verteilt […dispersed invisibly in the air.]	☐	☐
…Wärme. […heat.]	☐	☐

Figure 4. Sample item stem and corresponding response alternatives from the science achievement test in original German and English translation.

Test analysis. After application of the scoring scheme, the entire set of 21 dichotomously and 12 polytomously scored items was calibrated according to the partial credit model (Masters, 1982). As the science achievement test was administered to all students in identical form at the pre- and post-instructional measurement occasion, and thus individual tests were linked perfectly across measurement occasions and student groups, pre- and posttest responses were submitted jointly to a concurrent calibration (Hanson & Béguin, 2002; Kim & Kolen, 2006; Lee & Ban, 2010; see Robitzsch, 2009, for a discussion of issues surrounding the choice of concurrent or separate calibration). In particular, pretest responses of 2519 students and posttest responses of 2512 students accumulated to a calibration sample of 5031 virtual students. The estimated item difficulties ranged from -1.272 logits to 1.518 logits. On average, items covering phenomena of condensation were more difficult (M = 0.122, SD = 0.789) than items addressing phenomena of evaporation (M = -0.037, SD = 0.761) or treating the issue of phase changes in a decontexualized manner (M = -0.240, SD = 0.668). The average difficulty of items calling for the knowledge of simple facts (M = -0.025, SD = 0.703) was comparable to the average difficulty of items asking for the recognition of correct explanations for phenomena of evaporation and condensation (*M* = 0.024, *SD* = 0.779). In support of the validity of the test, those items containing scientifically adequate explanations acknowledging that evaporation and condensation involve changes in both the phase and visibility of water were among the most difficult items of the test (sat_as_5, sat_as_6). Item infit values varied between 0.89 and 1.18, while item outfit values spanned the range between 0.79 and 1.21. Infit values associated with item-specific step parameters fell between 0.92 and 1.04, whereas corresponding outfit values ranged from 0.94 to 1.76 (see Table 14).

For the virtual data set, the distribution of person ability possessed a mean of 0.85 logits and a standard deviation of 0.63 logits. With respect to the distribution of item difficulty, a predefined mean of 0.00 logits was accompanied by a standard deviation of 0.73 logits. The corresponding distribution of category boundaries displayed a mean of 0.20 logits with a standard deviation of 0.87 logits. (Please note that the mean of the distribution of category boundaries deviated from 0.00 logits because the science achievement test featured an unbalanced number of dichotomously and polytomously scored items.) The separation reliability of the weighted likelihood estimates of person ability amounted to 0.82.

The virtual data set comprising student responses from both the pre- and the post-instructional measurement occasion was submitted to analyses in the framework of classical test theory as well. For the 21 dichotomously scored items the mean relative proportion of correct solutions was 0.73 with a standard deviation of 0.13. For the 7 items with theoretical maxima of two raw points the mean of achieved raw points equaled 1.31 with a standard deviation of 0.22. Finally, with regard to the 5 items with theoretical maxima of three raw points the mean of achieved raw points equaled 1.46 with a standard deviation of 0.42. Item discriminations varied between r_{it} = -.10 and r_{it} = .46 (see Table 14). Specifically, for two out of thirty-three items corresponding item discriminations fell below r_{it} = .10. For another five items corresponding item discriminations attained values above r_{it} = .10 and below r_{it} = .20. The first component identified in a principal component analysis explained 13.85% of the total variance. There were ten components with eigenvalues greater than 1.00. The first through fifth components had eigenvalues of 4.57, 1.86, 1.62, 1.44, and 1.30, respectively. Thus, the ratio of the first and second eigenvalues amounted to 2.46. Moreover, Cronbachs' alpha of the virtual data set for the science achievement test equaled .77.

Differential item functioning. As the calibration of the science achievement test relied on a virtual data set comprising both pre- and posttest responses, the examination of differential item functioning featured measurement occasion (pretest vs. posttest) as an additional

Table 14
Probabilistic and Traditional Item Analyses for the Science Achievement Test

Parameter	Estimate	SE	Infit	Outfit	*M*	r_{it}
			Items			
sat_di_1	-0.689	0.036	0.99	1.01	0.81	.18
sat_di_2	0.239	0.030	1.01	1.02	0.64	.20
sat_di_3	-0.944	0.039	0.99	0.94	0.84	.23
sat_di_4	0.119	0.030	0.95	0.92	0.67	.34
sat_pc_1	0.685	0.019	0.97	0.96	1.11	.39
sat_pc_2	-0.640	0.034	0.98	0.96	1.60	.28
sat_pc_3	0.505	0.022	0.94	0.94	1.17	.40
sat_di_5	-0.940	0.039	0.98	0.92	0.84	.24
sat_di_6	-0.317	0.033	0.97	0.93	0.75	.28
sat_di_7	0.731	0.029	1.18	1.21	0.53	-.10
sat_di_8	-0.625	0.036	0.98	0.98	0.80	.23
sat_di_9	1.382	0.030	0.97	0.96	0.39	.30
sat_di_10	0.298	0.030	0.98	0.97	0.63	.27
sat_di_11	-0.382	0.034	1.02	1.03	0.76	.18
sat_di_12	-0.537[a]	0.035	1.00	1.01	0.79	.20
sat_di_13	0.715	0.029	1.03	1.04	0.53	.19
sat_as_1	-0.849	0.038	0.91	0.83	0.84	.34
sat_as_2	-0.712	0.036	0.91	0.81	0.82	.38
sat_as_3	0.672	0.016	1.03	1.03	1.67	.42
sat_as_4	0.484	0.018	1.10	1.10	1.84	.34
sat_as_5	1.237	0.017	1.04	1.13	1.15	.39
sat_as_6	1.518	0.018	1.00	1.18	0.88	.38
sat_as_7	-0.420	0.034	1.06	1.14	0.77	.05
sat_as_8	-0.011	0.023	0.89	0.87	1.44	.46
sat_as_9	-0.881	0.038	0.91	0.79	0.84	.37
sat_as_10	-1.272	0.044	0.94	0.83	0.89	.28
sat_as_11	0.522	0.021	1.04	1.05	1.75	.31
sat_as_12	-0.478	0.034	1.03	1.05	1.47	.20
sat_as_13	0.111	0.031	0.97	0.94	0.66	.32
sat_as_14	0.143	0.022	1.14	1.20	1.39	.15
sat_as_15	0.855	0.021	1.01	1.01	1.01	.33
sat_di_14	-0.180	0.032	0.95	0.92	0.73	.32
sat_di_15	-0.339[a]		1.04	1.06	0.75	.13
			Steps			
sat_pc_1(1)	0.080	0.031	1.00	0.99	---	---
sat_pc_1(2)	-0.080[a]		0.96	0.94	---	---
sat_pc_2(1)	-0.777	0.042	1.01	1.00	---	---
sat_pc_2(2)	0.777[a]		1.00	0.99	---	---
sat_pc_3(1)	-0.698	0.029	0.99	0.98	---	---
sat_pc_3(2)	0.698[a]		0.98	0.99	---	---
sat_as_3(1)	-0.170	0.036	0.98	0.94	---	---
sat_as_3(2)	0.249	0.045	1.01	1.03	---	---
sat_as_3(3)	-0.079[a]		0.92	0.96	---	---
sat_as_4(1)	-0.393	0.040	1.00	0.98	---	---
sat_as_4(2)	-0.084	0.042	1.01	1.02	---	---

Table 14 (continued)
Probabilistic and Traditional Item Analyses for the Science Achievement Test

Parameter	Estimate	SE	Infit	Outfit	*M*	r_{it}
			Steps (continued)			
sat_as_4(3)	0.477[a]		0.97	1.03	---	---
sat_as_5(1)	-0.036	0.034	1.00	1.00	---	---
sat_as_5(2)	-0.117	0.045	1.01	1.15	---	---
sat_as_5(3)	0.153[a]		0.99	1.46	---	---
sat_as_6(1)	-0.190	0.033	1.01	1.02	---	---
sat_as_6(2)	0.206	0.050	1.01	1.32	---	---
sat_as_6(3)	-0.016[a]		0.96	1.76	---	---
sat_as_8(1)	-0.331	0.033	0.98	0.96	---	---
sat_as_8(2)	0.331		0.92	0.91	---	---
sat_as_11(1)	-1.262	0.043	1.02	1.03	---	---
sat_as_11(2)	-0.178	0.039	1.00	1.00	---	---
sat_as_11(3)	1.440[a]		1.01	1.13	---	---
sat_as_12(1)	-1.223	0.040	1.03	1.04	---	---
sat_as_12(2)	1.223[a]		1.04	1.04	---	---
sat_as_14(1)	-0.151	0.033	0.99	0.99	---	---
sat_as_14(2)	0.151[a]		1.08	1.09	---	---
sat_as_15(1)	-0.599	0.028	0.99	0.99	---	---
sat_as_15(2)	0.599[a]		1.01	1.00	---	---

Note. Item labels provide information on the scoring applied (di = dichotomously scored within a stem, pc = scored with partial credit within a stem, as = scored across stems). Numbers in parentheses indicate step position.
[a]Parameter estimate is constrained.

grouping variable. The analysis of this form of differential item functioning provided an approach toward the identification of those facets of knowledge about the topic of evaporation and condensation that underwent specifically weak or intense change due to science instruction (cf. Wainer, 2010a). The science achievement test was calibrated according to the partial credit model (Masters, 1982). Thus, the model-based inferential analyses of the general presence of differential item functioning featured a sequence of three nested models for each grouping variable. These were a baseline model, an item interaction model and a full interaction model. The baseline models contained main effects for items and the respective grouping variable as well as an interaction between items and steps that allowed the modeling of varying numbers of steps per item in the partial credit model. In other words, the baseline models constituted applications of the partial credit model with simultaneous control of mean group differences in science achievement. The item interaction models extended the baseline models by inclusion of an interaction between items and the respective grouping variable. Thereby, the item interaction models allowed mean item difficulties to vary between groups. The full interaction models, eventually, refined the item interaction models by replacing the interaction between items and steps with a three-way interaction between items, steps and the respective grouping variable. The full interaction models allowed both mean item difficulties and item-specific step parameters to vary between groups. So, for each grouping variable, the inferential analysis of differential item functioning consisted straightforwardly of comparisons of the baseline model with the item interaction model as well of the item interaction model with the full interaction model (see Figure 5). Additionally, parameter estimates from full interaction models were used to compute logit differences between groups for, in case of dichotomously scored items, item difficulties and, in case of polytomously scored items, difficulties of category boundaries as measures of effect sizes.

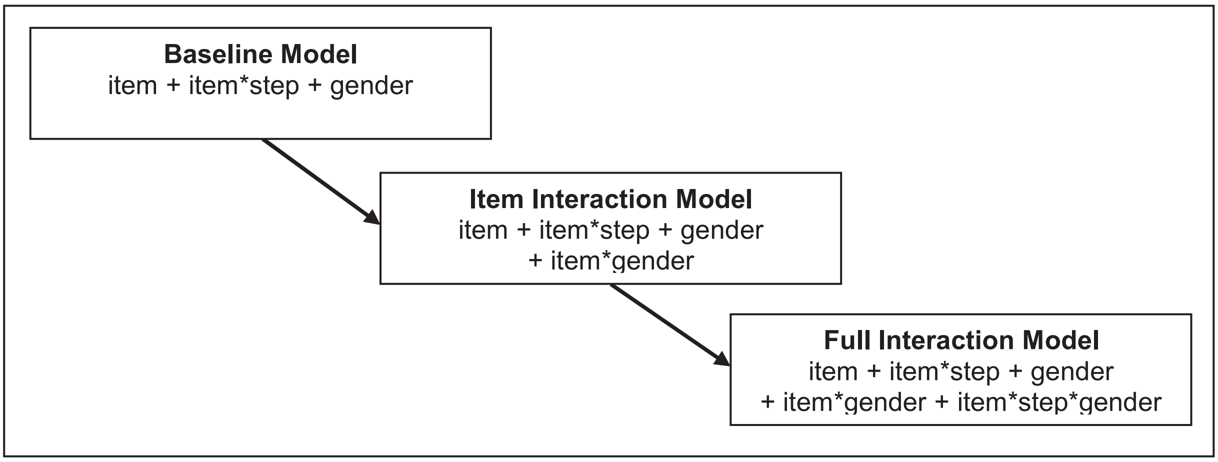

Figure 5. Overview of model comparisons for the inferential analysis of differential item functioning in the science achievement test for the grouping variable of gender.

Due to missing data for the grouping variables as well as for the pre- and post-instructional science achievement test, the explorations of differential item functioning for the individual grouping variables resorted to samples of varying size. The examination of differential item functioning for the grouping variable of measurement occasion included responses from 2519 students for pre-instructional measurement and from 2512 students for post-instructional measurement. The item interaction model fitting the data better than the baseline model, $\chi^2(32) = 1156.93$, $p < .001$, and the full interaction model fitting the data better than the item interaction model, $\chi^2(34) = 162.01$, $p < .001$, for the grouping variable of measurement occasion the full interaction model was found to fit the data relatively best. On the one hand, three items and two category boundaries displayed differential item functioning of intermediate size in favor of pretest responses (see Table 15). For three further category boundaries differential item functioning in favor of pretest responses was large. On the other hand, one item showed intermediate differential item functioning in favor of posttest responses, whereas two items and one category boundary exhibited large differential item functioning in favor of posttest responses. The items and category boundaries showing differential item functioning in favor of pretest responses contained the conceptions that evaporated water turns into oxygen (sat_as_3), that evaporated water concentrates beneath the ceiling (sat_as_7), that coldness turns into condensed water (sat_as_11), that the sun attracts evaporated water (sat_as_14) and that empty clouds collect water (sat_di_15). Moreover, they touched the correct usage of the terms evaporation and vapor (sat_di_1, sat_as_12) and the identification of matter in the gaseous phase (sat_pc_2). In contrast, the items and the category boundary displaying differential item functioning in favor of posttest responses were concerned with the rejection of the conception that water turns into warmth or vice versa during phase changes (sat_as_8) as well as with the recognition that water droplets on cold glass surfaces or fog in the morning result from condensation (sat_di_5, sat_di_9, sat_as_13). Apparently, differential item functioning in favor of pretest responses occurred when extended knowledge about evaporation and condensation, which naturally tends to be absent prior to instruction, could induce deficient responses, such as in case of the conception of the concentration of evaporated water beneath the ceiling. Furthermore, the occurrence of differential item functioning in favor of posttest responses for items covering phenomena of condensation suggested that knowledge about condensation prior to instruction was sparse and that those or similar phenomena typically had been part of the instruction on evaporation and condensation.

For the grouping variable of gender, the exploration of differential item functioning relied on 2674 response sets associated with boys and 2356 response sets associated with girls. The incorporation of the interaction between items and gender improved overall model fit for the item interaction model relative to the baseline model, $\chi^2(32) = 457.53$, $p < .001$. However, the additional inclusion of a three-way interaction between items, steps and gender did not further enhance model fit, $\chi^2(34) = 30.48$, $p = .641$. One item exhibited intermediate differential item functioning in favor of boys. One item displayed large differential item functioning in favor of girls. Thus, in relation to the total number of items and category boundaries of the science achievement test, the consequences of differential item functioning for the grouping variable of gender were minimal and balanced.

With regard to raw score in the science achievement test as the grouping variable, the investigation of differential item functioning took 2516 low-scoring response sets and 2515 high-scoring response sets into consideration. The model fit of the item interaction model exceeded the model fit of the baseline model, $\chi^2(32) = 1761.91$, $p < .001$, while the model fit of the full interaction model outstripped the model fit of the item interaction model, $\chi^2(34) = 201.36$, $p < .001$. In particular, one item and one category boundary showed intermediate differential item functioning in favor of low-scoring students. Moreover, two items and three category boundaries displayed large differential item functioning in favor of low-scoring students. On the contrary, two items and two category boundaries exhibited intermediate differential item functioning in favor of high-scoring students. Furthermore, three items showed large differential item functioning in favor of high-scoring students. So, as items contribute with more weight to the estimation of individual person ability than category boundaries, the effects of differential item functioning were approximately balanced across subgroups, save one item and one category boundary displaying intermediate differential item functioning in favor of high-scoring students. Specifically, low-scoring students fared better than expected on items and category boundaries that referred to the conceptions that water turns into oxygen or vice versa during phase changes (sat_as_3, sat_as_4), that evaporated water concentrates near the ceiling (sat_as_7) and that evaporation is the attraction of water by the sun (sat_as_14). Items and category boundaries that advantaged high-scoring students covered the conceptions that condensed water is sweated out (sat_as_1), that water, during phase changes, turns into or results from smoke (sat_as_2), that heat transforms into water and vice versa (sat_as_8), that evaporated water ceases to exist (sat_as_9), that condensed water is created by light (sat_as_10) and that coldness turns into condensed water (sat_as_11). In essence, high-scoring students did even better than expected in rejecting naïve conceptions, whereas low-scoring students were advantaged when conceptions had to be refuted that contained misguiding connections to the knowledge that water consists of molecules and that heat promotes evaporation.

The inspection of differential item functioning for the grouping variable of socio-economic status used data of 1830 response sets for students of low socio-economic status and of 1843 response sets for students of high socio-economic status. The item interaction model surpassed the baseline model in terms of model fit, $\chi^2(32) = 285.96$, $p < .001$. Nevertheless, the incorporation of a three-way interaction between items, steps and socio-economic status to form the full interaction model did not entail a significant enhancement of overall fit, $\chi^2(34) = 23.22$, $p = .919$. Thus, the item interaction model was disclosed as fitting relatively best to the data. Merely one item exhibited non-negligible differential item functioning, of large size and in favor of students of low socio-economic status.

Response sets of 2655 students without immigrant background and response sets of 965 students with immigrant background were included in the examination of differential item functioning for the grouping variable of immigrant background. Though the inclusion of the interaction between items and immigrant background rendered the fit of the item interaction model better than the fit of the baseline model, $\chi^2(32) = 345.42$, $p < .001$, the full interaction model did not display better fit than the item interaction model, $\chi^2(34) = 23.84$, $p = .903$. Accordingly, the item interaction model was identified as the model with the best relative fit to the data. On the one hand, one item showed intermediate differential item functioning and one item showed large differential item functioning in favor of students without immigrant background. These items were concerned with the student conceptions that water turns into or results from smoke (sat_as_2) and that light creates condensed water (sat_as_10). On the other hand, two items exhibited intermediate differential item functioning in favor of students with immigrant background. These items covered the student conception that evaporated water in closed rooms concentrates near the ceiling (sat_as_7) and the knowledge that cloud formation is not a phenomenon of evaporation (sat_di_7). Nevertheless, for the grouping variable of immigrant background effects of differential item functioning were comparatively small and balanced across groups.

The analysis of differential item functioning for the grouping variable of grade level was based on responses from 2526 virtual fourth-graders and 2505 virtual sixth-graders. The model fit of the item interaction model exceeded the model fit of the baseline model, $\chi^2(32) = 892.80$, $p < .001$. Moreover, the full interaction model possessed better model fit than the item interaction model, $\chi^2(34) = 191.43$, $p < .001$, so that the full interaction model was identified as fitting best to the data. For one item and one category boundary intermediate differential item functioning in favor of fourth-graders was established. For one further item large differential item functioning in favor of fourth-graders was found. In particular, virtual fourth-graders were more likely than expected due to their overall ability to correctly name and identify instances of evaporation (sat_di_1, sat_di_7). Four category boundaries were shown to possess intermediate differential item functioning in favor of sixth-graders. Two category boundaries exhibited large differential item functioning in favor of sixth-graders. Specifically, virtual sixth-graders displayed extraordinary strengths in recognizing that processes of evaporation and condensation involve changes in both visibility and phase of water (sat_as_5, sat_as_6). Nevertheless, effects of differential item functioning tended to cancel out each other across groups.

The exploration of differential item functioning for the grouping variable of school type relied on data of 2526 virtual students from the school type of Grundschule, 1079 virtual students from the school type of Hauptschule and 1426 virtual students from the school type of Gymnasium. With regard to the comparison based on virtual students from the school types of Grundschule and Hauptschule, the item interaction model was found to fit better to the data than the baseline model, $\chi^2(32) = 571.90$, $p < .001$, and, in turn, the full interaction model was found to fit better to the data than the item interaction model, $\chi^2(34) = 85.17$, $p < .001$. Three items and one category boundary exhibited differential item functioning of intermediate size in favor of the school type of Grundschule. On the contrary, one item and four category boundaries displayed intermediate differential item functioning in favor of the school type of Hauptschule. Furthermore, one category boundary showed large differential item functioning in favor of the school type of Hauptschule. So, evaluated in terms of its net effects on the estimation of person ability, there was a minimal tendency for differential item functioning to advantage students attending the school type of Grundschule. Specifically, the item covering the conception that water turns into or results from smoke during phase changes (sat_as_2)

Table 15
Differential Item Functioning in Terms of Logit Differences for the Science Achievement Test

Item	Step	Difference in Difficulty								
		Pre – Post	Boys – Girls	Losco – Hisco	Loses – Hises	Migno – Migyes	Fourth – Sixth	Grund – Haupt	Grund – Gymna	Haupt – Gymna
sat_di_1	1	**-0.478**	0.018	-0.254	-0.122	-0.002	***-0.686***	**-0.610**	***-0.768***	-0.154
sat_di_2	1	0.202	-0.004	-0.178	0.074	-0.172	0.128	0.208	0.036	-0.174
sat_di_3	1	0.412	**-0.464**	0.222	-0.034	0.286	0.272	0.248	0.302	0.052
sat_di_4	1	0.054	-0.118	0.410	0.196	-0.114	0.388	0.234	**0.524**	0.288
sat_pc_1	1	0.221	-0.006	-0.234	-0.003	-0.167	-0.248	0.252	-0.352	-0.244
sat_pc_1	2	0.259	0.030	0.210	0.255	-0.125	0.124	0.108	0.300	**0.552**
sat_pc_2	1	***-0.680***	0.003	0.409	0.359	-0.286	**-0.522**	-0.283	***-0.796***	**-0.519**
sat_pc_2	2	-0.056	-0.063	-0.109	0.117	-0.126	***0.642***	0.315	***0.972***	***0.655***
sat_pc_3	1	-0.119	-0.155	**0.542**	0.205	-0.227	-0.334	**-0.448**	-0.068	0.377
sat_pc_3	2	-0.233	0.163	-0.194	0.171	-0.113	***0.762***	**0.444**	***0.872***	0.423
sat_di_5	1	**0.532**	-0.114	0.224	0.132	-0.128	-0.280	-0.320	-0.238	0.082
sat_di_6	1	0.284	-0.052	0.316	0.012	-0.258	0.182	0.132	0.222	0.090
sat_di_7	1	-0.100	0.288	***-1.320***	***-0.754***	**0.626**	**-0.544**	0.380	***-1.286***	***-1.660***
sat_di_8	1	0.090	0.306	0.104	0.122	0.104	-0.180	-0.248	-0.112	0.138
sat_di_9	1	***0.722***	0.028	0.136	-0.078	0.226	-0.014	-0.176	0.044	0.222
sat_di_10	1	0.058	0.048	0.018	-0.100	0.178	-0.276	**-0.440**	-0.160	0.282
sat_di_11	1	0.100	-0.172	-0.306	-0.184	0.316	-0.018	0.270	-0.292	**-0.560**
sat_di_12	1	-0.316	0.090	-0.212	-0.104	0.138	-0.274	-0.284	-0.274	0.012
sat_di_13	1	-0.010	-0.148	-0.272	-0.260	0.260	0.262	**0.430**	0.096	-0.332
sat_as_1	1	-0.116	0.080	***0.702***	0.340	-0.310	-0.112	-0.412	0.304	***0.716***
sat_as_2	1	0.342	0.082	***1.002***	0.390	***-0.766***	-0.314	**-0.516**	-0.080	**0.438**
sat_as_3	1	***-0.783***	-0.152	***-0.661***	-0.038	0.220	-0.148	-0.013	-0.277	-0.266
sat_as_3	2	-0.307	-0.312	-0.254	-0.050	0.026	-0.010	0.020	-0.020	-0.043
sat_as_3	3	0.118	-0.304	0.327	0.093	-0.132	0.248	-0.007	0.303	0.309
sat_as_4	1	-0.395	-0.133	***-0.673***	-0.374	-0.015	-0.081	0.186	-0.384	**-0.570**
sat_as_4	2	-0.122	-0.136	**-0.471**	-0.155	0.320	-0.225	-0.044	-0.373	-0.330
sat_as_4	3	-0.239	-0.278	0.118	0.012	-0.065	0.150	-0.090	0.242	0.331
sat_as_5	1	-0.075	0.145	-0.069	-0.007	-0.016	0.421	0.287	**0.578**	0.285
sat_as_5	2	-0.050	0.064	-0.407	-0.021	0.292	**0.475**	**0.516**	0.386	-0.138
sat_as_5	3	0.107	-0.005	-0.316	-0.110	0.011	**0.460**	**0.450**	0.368	-0.092
sat_as_6	1	-0.125	-0.078	-0.229	0.076	0.176	**0.633**	***0.726***	**0.539**	-0.193
sat_as_6	2	-0.344	-0.097	-0.344	-0.020	-0.171	0.382	0.375	0.343	-0.039
sat_as_6	3	0.031	-0.167	-0.302	0.136	0.078	**0.472**	0.124	**0.468**	0.334

Table 15 (continued)

Differential Item Functioning in Terms of Logit Differences for the Science Achievement Test

Item	Step	Difference in Difficulty								
		Pre – Post	Boys – Girls	Losco – Hisco	Loses – Hises	Migno – Migyes	Fourth – Sixth	Grund – Haupt	Grund – Gymna	Haupt – Gymna
sat_as_7	1	**-0.472**	***0.718***	***-0.858***	-0.188	**0.428**	-0.082	0.190	-0.344	**-0.530**
sat_as_8	1	-0.101	-0.181	0.340	0.108	-0.081	-0.147	-0.222	0.001	0.225
sat_as_8	2	***0.721***	-0.343	**0.440**	-0.184	0.229	-0.109	-0.042	-0.201	-0.157
sat_as_9	1	0.242	0.072	***1.098***	0.326	-0.238	0.014	-0.274	0.420	***0.696***
sat_as_10	1	-0.316	0.322	**0.592**	0.318	**-0.440**	0.082	-0.238	**0.616**	***0.854***
sat_as_11	1	**-0.474**	0.123	-0.205	-0.185	0.034	-0.340	-0.291	-0.396	-0.105
sat_as_11	2	-0.168	-0.231	-0.348	-0.162	0.244	0.010	0.220	-0.191	-0.409
sat_as_11	3	0.107	-0.276	-0.401	-0.049	0.172	0.090	0.161	-0.003	-0.163
sat_as_12	1	-0.212	-0.317	-0.046	0.247	-0.380	-0.059	0.177	-0.319	**-0.497**
sat_as_12	2	***-0.764***	-0.315	-0.254	0.157	-0.356	0.219	0.127	0.259	0.133
sat_as_13	1	***0.680***	0.120	**0.432**	-0.126	0.088	0.416	0.328	**0.486**	0.156
sat_as_14	1	**-0.638**	0.186	***-1.162***	-0.302	0.172	0.137	**0.486**	-0.237	***-0.724***
sat_as_14	2	-0.118	0.178	-0.334	-0.126	0.304	0.123	0.126	0.105	-0.020
sat_as_15	1	-0.104	-0.187	-0.173	-0.108	0.253	-0.154	0.030	-0.342	-0.372
sat_as_15	2	0.036	0.103	-0.183	-0.028	0.063	0.254	0.354	0.146	-0.208
sat_di_14	1	0.378	-0.086	0.408	-0.080	-0.194	-0.286	-0.280	-0.308	-0.026
sat_di_15	2	**-0.484**	0.046	**-0.474**	-0.030	0.006	0.134	0.148	0.110	-0.040

Note. Calculations of differential item functioning were based on full interaction models comprising main effects for item and the respective grouping variable as well as for the two-way interaction between item and the grouping variable and the three-way interaction between item, step and the grouping variable. Bold print highlights intermediate, bold and italic print signifies large differential item functioning. Pre/Post = pre-/posttest responses; Losco/Hisco = low/high raw score on the scale; Loses/Hises = low/high socio-economic status; Migno/Migyes = non-immigrant/immigrant background; Fourth/Sixth = fourth/sixth grade; Grund = Grundschule; Haupt = Hauptschule; Gymna = Gymnasium.

advantaged students from the school type of Grundschule, whereas category boundaries associated with those items containing the conceptions that evaporation and condensation involve changes in both visibility and phase of water (sat_as_5, sat_as_6) and that the sun attracts evaporated water (sat_as_14) advantaged students from the school type of Hauptschule.

The inferential analysis of differential item functioning for responses associated with the school types of Grundschule and Gymnasium returned the full interaction model as fitting relatively best to the data, with the item interaction model possessing better fit than the baseline model, $\chi^2(32) = 1133.04$, $p < .001$, and the full interaction model displaying better fit than the item interaction model, $\chi^2(34) = 189.18$, $p < .001$. In particular, two items and one category boundary showed large differential item functioning in favor of responses related to the school type of Grundschule. Three items and three category boundaries exhibited differential item functioning of intermediate size in favor of responses associated with the school type of Gymnasium. Additionally, two category boundaries displayed large differential item functioning in favor of responses related to the school type of Gymnasium. Thus, in sum, there was a minimal general tendency for differential item functioning to inflate the estimation of person ability for students from the school type of Gymnasium. Particularly, items and category boundaries associated with the conceptions that light creates condensed water (sat_as_10) and that both visibility and phase of water change during evaporation and condensation (sat_as_5, sat_as_6) advantaged students from the school type of Gymnasium.

Finally, for virtual students from the school types of Hauptschule and Gymnasium, the model fit of the item interaction model exceeded the model fit of the baseline model, $\chi^2(32) = 805.57$, $p < .001$. Likewise, the full interaction model possessed better model fit than the item interaction model, $\chi^2(34) = 123.56$, $p < .001$. Two items and three category boundaries contained intermediate differential item functioning in favor of the school type of Hauptschule. Moreover, one item and one category boundary displayed large differential item functioning in favor of the school type of Hauptschule. In particular, the conceptions that oxygen turns into water during condensation (sat_as_4), that evaporated water concentrates near the ceiling (sat_as_7) and that the sun attracts evaporated water (sat_as_14) were less popular among students from the school type of Hauptschule than expected according to their overall ability. One item and one category boundary showed intermediate differential item functioning in favor of the school type of Gymnasium. In addition, three items and one category boundary possessed large differential item functioning in favor of the school type of Gymnasium. Specifically, the deficient conceptions that objects sweat (sat_as_1), that evaporated water turns into smoke (sat_as_2) or ceases to exist (sat_as_9) and that light generates condensed water (sat_as_10) were less common among students from the school type of Gymnasium than expected due to their estimated overall ability with respect to the topic of evaporation and condensation. Considered in total, effects of differential item functioning exhibited a minimal tendency to advantage students from the school type of Gymnasium.

Summary. Infit values uncovered very good fit of the items to the partial credit model (Bond & Fox, 2007; Masters, 1982; Smith, 2000). The mean of the ability distribution surpassing the mean of the difficulty distribution by more than one standard deviation of the difficulty distribution, many items of the test were comparatively easy to solve for the students. According to the person separation reliability the weighted likelihood estimates of person ability were eminently reliable. Summarizing responses across item stems, the complex scoring procedure obviously did not take all possible sources of local dependency into consideration. For instance, response alternatives nested under a common item stem could be locally dependent. However, additional analyses comparing – for a subset of the items of the science achievement test – the impact of different scoring procedures on the reliability of corresponding estimates of person ability indicated that the separation reliability obtained in the present test analysis might be considered a lower bound for the actual reliability of estimates of person ability generated with the science achievement test (Tröbst, 2010). With respect to the dimensionality of the test, the results of the principal component analysis were difficult to interpret. Though the amount of variance explained by the first component was comparatively low, the sequence of eigenvalues provided evidence in favor of an outstanding first component.

The inspection of differential item functioning for the grouping variable of measurement occasion highlighted a specificity of science learning. In pre-instructional measurement, students solved items for that enriched knowledge about evaporation and condensation could induce deficient answers more often correctly than expected due to their overall topic-specific ability in science. In other words, rudimentarily, the occurrence of differential item functioning in favor of pretest responses captured the induction of misconceptions by science instruction. Items displaying differential item functioning in favor of posttest responses were concerned primarily with phenomena of condensation, a hint that children acquire advanced knowledge about condensation rather through formal instruction than by informal learning or everyday experiences. A similar pattern was observed for the grouping variable of raw score in the science achievement test. Low-scoring students displayed unexpected relative strengths in avoiding response alternatives that contained confusing connections to knowledge of molecules and heat as a promoter of evaporation whereas high-scoring students excelled in refusing naïve conceptions. The consequences of differential item functioning for the estimation of person ability cancelled out each other for the subgroups associated with the grouping variable of grade level. With regard to the grouping variable of school type, the model-based investigations of differential item functioning uncovered in pairwise comparisons a minimal tendency for inflation of ability estimations for students from the school type of Gymnasium relative to students from the school types of Grundschule and Hauptschule. Similarly, a minimal tendency for inflation of ability estimations was found for students from the school type of Grundschule relative to students from the school type of Hauptschule. By and large, the implications of differential item functioning were negligible for the grouping variables of gender, socio-economic status and immigrant background.

Treatment of Missing Data

Questions on socio-economic status in parental questionnaires are prone to suffer from the occurrence of missing data. As a consequence of this, the proper handling of missing data is a crucial issue when parental questionnaires are used to assess family socio-economic status (Ensminger et al., 2000; Maaz, Kreuter, & Watermann, 2006). Specifically, when treated with the conventional method of listwise deletion, high proportions of missing data entail serious setbacks for the credibility of inferential analyses. First, the reduced sample size implicates reduced efficiency of parameter estimations. Second, systematic differences between observed and missing data, in other words, the potential unrepresentativeness of the observed data for the population to which generalization is intended, may cause biased parameter estimations (Allison, 2001; Graham, Cumsille, & Elek-Fisk, 2003; Lüdtke, Robitzsch, Trautwein, & Köller, 2007). Fortunately, under the assumption that data is missing conditionally at random, data augmentation by advanced methods for the treatment of missing data is capable to prevent or, at least, mitigate inefficiency and bias of parameter estimations (Graham, 2009; Lüdtke & Robitzsch, 2010; Rubin, 1976; Spieß, 2009). In essence, full information maximum likelihood estimation and multiple imputation represent the two central advanced methods for the treatment of missing data (Arbuckle, 1996; Enders, 2010; Graham, 2009).

In order to adequately reduce the bias potentially associated with the occurrence of missing data, it is mandatory to formulate an imputation model that is at least as complex as the analysis model. For instance, it may have some intuitive appeal not to include dependent variables in the imputation model when researches fear they could unduly strengthen relations between independent and dependent variables by multiple imputation. In fact, the opposite is true. When dependent variables are omitted from the imputation model, values for missing data are imputed under the assumption that there is no systematic association between independent and dependent variables. In others words, in this case relations of substantial scientific interest are biased towards zero (Allison, 2001; Graham, 2009; Landerman, Land, & Pieper, 1997). The same principle applies to the multilevel structure of data. When clustering of data is not taken into account by the imputation model, the size of intraclass correlations is systematically underestimated in subsequent inferential analyses (Lüdtke & Robitzsch, 2010).

How many imputations of missing values are necessary to ensure accuracy of inferential analyses? With regard to this question, it has been shown that with a small number of imputations estimations of standard errors and of fractions of missing information suffer from imprecision (Bodner, 2008; Lüdtke & Robitzsch, 2010). Especially statistical power for detection of small effects is low with few imputations (Graham, 2009; Graham, Olchowski, & Gilreath, 2007). Thus, in the face of recent simulation studies, the popular recommendation that five imputations are sufficient to obtain exact inferential results must be considered obsolete (Schafer & Olson, 1998). Instead, for instance, Graham and colleagues (2007) have suggested the generation of at least 40 imputations when 50% of the data is missing.

Two general approaches for the imputation of multivariate missing data have evolved (Lüdtke & Robitzsch, 2010; van Buuren & Groothuis-Oudshoorn, 2011). Joint modeling techniques, on the one hand, rely on the specification of a multivariate distribution for the entire data set at hand. For instance, the software package NORM imputes multivariate missing data under the assumption of a multivariate normal distribution (Schafer, 1997). On the other hand, techniques based on chained equations specify the imputation model on a variable-by-variable basis, thereby allowing the formulation of individual equations for the imputation of each incomplete variable (van Buuren, 2007). Therefore, resting on fewer assumptions than the application of joint modeling techniques, the use of chained equations offers both a flexible and a computationally economical approach to the handling of missing data (Lüdtke & Robitzsch, 2010; van Buuren, Brand, Grothuis-Oudshoorn, & Rubin, 2006; van Buuren & Oudshoorn; 1999).

Table 16
Proportion of Nonresponse and Correlations with Nonresponse for the Variables of the Analysis Models

Variable	% Miss	SES	CFT	SAT1	SCA	INT	SAT2	Grade[a]	Haupt[b]	Gymn[b]
SES	28.96	----	-.12	-.16	-.03	-.08	-.26	.15	.20	-.02
CFT	6.72	-.10	----	-.07	-.05	-.04	-.12	.05	.12	-.05
SAT1	6.01	-.04	-.02	----	-.04	-.02	-.10	.04	.09	-.04
SCA	7.24	-.09	-.07	-.07	----	.02	-.11	.03	.09	-.06
INT	7.31	-.10	-.06	-.07	-.01	----	-.10	.03	.09	-.05
SAT2	6.27	.02	-.01	-.03	-.02	-.01	----	.07	.10	-.01
Grade[a]	0.00	----	----	----	----	----	----	----	----	----
Haupt[b]	0.00	----	----	----	----	----	----	----	----	----
Gymn[b]	0.00	----	----	----	----	----	----	----	----	----

Note. As there was no nonresponse for the variables of grade level and school type, corresponding correlations between nonresponse and variables could not be computed. SES = Socio-economic Status; CFT = Fluid Ability (assessed with CFT 20-R); SAT1 = Topic-specific Prior Knowledge; SCA = Self-concept of Ability; INT = Individual Interest in Science; SAT2 = Post-instructional Science Achievement; Haupt = Hauptschule; Gymn = Gymnasium.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).

Proportion of Missing Data

In the current data set, in line with the apparent susceptibility of parental reports of socio-economic status to the occurrence of missing data (Ensminger et al., 2000), the variable of socio-economic status, operationalized by the HISEI-score for each child, displayed the largest proportion of nonresponses with 28.96% of the data missing (see Table 16). With regard to the central dependent variable and the four potential mediators, the proportion of missing data varied between 6.01% for the measure of topic-specific prior knowledge and 7.31% for the measure of individual interest in science. For the class-level variables of grade level and school type there was no missing data so that nonresponses occurred exclusively on the individual level of the multilevel structure of the data. The occurrence of nonresponses for the variable of socio-economic status was negatively related to the other individual-level variables (see Table 16). In comparison, the negative relation to the central dependent variable, student achievement after instruction, was particularly pronounced, with r = -.26. In other words, students whose parents did not provide information on family socio-economic status were likely to show low ability in the post-instructional achievement test. In contrast, the occurrence of nonresponses for the variable of socio-economic status was positively, r = .20, associated with attendance of the school type of Hauptschule. By and large, this pattern of associations for the occurrence of nonresponses held true for all individual-level variables.

Multiple Imputation

Multiple imputation of missing data for the current data set was performed by means of chained equations with the mice-package (van Buuren & Groothuis-Oudshoorn, 2011) for the statistical computing environment R 2.13.1 (R Development Core Team, 2011). Naturally, the imputation model covered variables included in the analysis models. These were school type as a dummy-coded variable, the HISEI-score for each child as well as the measures for fluid ability, topic-specific prior knowledge, self-concept of ability in science, individual interest in science and post-instructional student achievement. The variable of grade level was not included in the imputation model due to its great overlap with the variable of school type and the higher relevance of the latter variable for the prediction of missing responses. Furthermore, three variables not relevant for the analysis models were incorporated into the imputation model. Besides gender, these included measures for students' perceptions of instructional clarity and for students' perceptions of the use of student experiments from the second measurement occasion, i.e. for the first lesson of the series of lessons. In addition to that, variables associated with the districts or municipalities the participating schools were located in were added to the imputation model from external sources. These variables comprised the number of divorces per 1000 residents in 2008 (Landesbetrieb Technik und Information Nordrhein-Westfalen, 2009), the rate of unemployed workers receiving social welfare according to the legislation of the Sozialgesetzbuch II in April 2008 (Bundesagentur für Arbeit, 2008), the proportion of resident aliens among the population in December 2007 (Ministerium für Generationen, Familie, Frauen und Integration des Landes Nordrhein-Westfalen, 2010, p. 18) and the average annual income of salary and income tax payers in 2007 (Landesbetrieb Technik und Information Nordrhein-Westfalen, 2011).

The three missing responses for the variable of gender were imputed by predictive mean matching. For the other variables, preserving relevant aspects of the multilevel structure of the data, missing responses were imputed under the assumption of a two-level linear model (cf. Lüdtke & Robitzsch, 2010; van Buuren & Groothuis-Oudshoorn, 2011). In particular, random intercepts across classrooms were incorporated into the equations for the substitution of missing responses for the variables of family socio-economic status, fluid ability, topic-specific prior knowledge, self-concept of ability in science, individual interest in science, post-instructional student achievement, the perception of instructional clarity and the perception of the use of student experiments. With regard those relations in the analysis models that were possibly moderated by grade level or school type, random slopes were embodied in the imputation model. On the one hand, these were the relations of socio-economic status with the four potential mediators, namely socio-economic status, fluid ability, prior knowledge, self-concept of ability and individual interest. On the other hand, these were the associations of socio-economic status and the four potential mediators with post-instructional achievement. Moreover, classroom means for the variables of socio-economic status, fluid ability, prior knowledge, self-concept of ability, individual interest, post-instructional achievement, the perception of clarity and the perception of the use of student experiments were generated by passive imputation and used for the imputation of missing responses on other variables. The imputation algorithm was set to run through 20 iterations. A collection of 50 imputed data sets was generated for subsequent inferential analyses. In these data sets, imputed values for the HISEI-score that fell outside the range of permissible values were replaced by the theoretical minimum or maximum of the HISEI-score.

Results

Descriptive Statistics

To establish congruity with the subsequent inferential analyses, computations of descriptive statistics were aggregated across the entire collection of 50 imputed data sets. Therefore, the means and standard deviations retrieved for fluid ability, self-concept of ability in science and individual interest in science for the total sample differed slightly from the corresponding means and standard deviations of the person ability and attitude strength distributions obtained in the initial calibrations (see Table 17). The comparison of students' science achievement at the pre-instructional measurement occasion as an indicator of students' topic-specific prior knowledge (M = 0.63, SD = 0.59) with students' science achievement at the post-instructional measurement occasion as the central dependent variable (M = 1.06, SD = 0.80) revealed that, on average, students had acquired additional topic-specific proficiency in the course of the instructional series on the topic of evaporation and condensation. Moreover, the inspection of the zero-order correlations between the measures gave a first impression of the viability of the assumed mediational paths between family socio-economic status and student achievement. Specifically, the positive bivariate correlation between socio-economic status and post-instructional science achievement, r = .21, hinted at the presence of a social gradient for student achievement that set the scene for a potential mediation by student characteristics. With regard to the association between socio-economic status and the potential mediators, zero-order correlations with cognitive propensities, i.e. fluid ability and prior knowledge, outstripped zero-order correlations with motivational propensities, i.e. self-concept of ability and individual interest in science. In fact, the bivariate correlation between socio-economic status and individual interest in science approached zero, r =.02. Last but not least, positive zero-order correlations between individual propensities and post-instructional science achievement underscored the relevance of all four potential mediators for the prediction of student achievement. Nevertheless, again correlations with cognitive propensities exceeded correlations with motivational propensities. Topic-specific prior knowledge possessed by far the closest relation to post-instructional science achievement, r = .61.

Table 17
Descriptive Statistics and Intercorrelations of the Measures for the Total Sample

Variable	*M*	*SD*	Intercorrelations					
			SES	CFT	SAT1	SCA	INT	SAT2
SES	47.30	20.10	---	.14	.23	.09	.02	.21
CFT	0.58	0.89		---	.41	.10	.01	.35
SAT1	0.63	0.59			---	.17	.04	.61
SCA	0.82	1.66				---	.49	.19
INT	-0.05	1.59					---	.10
SAT2	1.06	0.80						---

Note. Computations were based on aggregation over 50 imputed data sets. SES = Socio-economic Status; CFT = Fluid Ability (assessed with CFT 20-R); SAT1 = Topic-specific Prior Knowledge; SCA = Self-concept of Ability; INT = Individual Interest in Science; SAT2 = Post-instructional Science Achievement.

Further preliminary insights into the structure of the data were gleaned from inspection and comparison of descriptive statistics for the subsamples of fourth- and sixth-graders. With regard to average family socio-economic status, differences between the subsamples of fourth- (M = 47.69, SD = 19.70) and sixth-graders (M = 46.92, SD = 20.46) were negligible (see Table 18). In terms of topic-specific prior knowledge, sixth-graders (M = 0.70, SD = 0.66) tended to outperform fourth-graders (M = 0.57, SD = 0.51). However, counter-intuitively, fourth-graders (M = 1.12, SD = 0.72) surpassed sixth-graders (M = 0.99, SD = 0.87) as far as post-instructional science achievement was concerned. Presumably, this result-ed from the overrepresentation of comparatively low-achieving students from the school type of Hauptschule in the subsample of sixth-graders. Nevertheless, the subsample of sixth-graders possessed superior average fluid ability, in line with the implications of age-related intelligence growth. With respect to the motivational propensities, the picture was reversed. Sixth-graders reported considerably lower self-concept of ability in science and individual interest in science than fourth-graders, a hint at a general developmental decline of academic self-concepts and interests in the transition from elementary to secondary education.

Table 18
Descriptive Statistics for the Measures by Grade Level and School Type

Group	SES		CFT		SAT1		SCA		INT		SAT2	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fourth Grade[a]	47.69	19.70	0.43	0.80	0.57	0.51	1.03	1.66	0.39	1.46	1.12	0.72
Sixth Grade	46.92	20.46	0.72	0.95	0.70	0.66	0.62	1.62	-0.48	1.59	0.99	0.87
Hauptschule	38.93	19.83	0.25	0.86	0.36	0.52	0.55	1.66	-0.43	1.57	0.54	0.67
Gymnasium	53.30	18.62	1.08	0.87	0.96	0.63	0.67	1.59	-0.51	1.61	1.36	0.84

Note. Computations were based on aggregation over 50 imputed data sets. SES = Socio-economic Status; CFT = Fluid Ability (assessed with CFT 20-R); SAT1 = Topic-specific Pri-or Knowledge; SCA = Self-concept of Ability; INT = Individual Interest in Science; SAT2 = Post-instructional Science Achievement.
[a]Naturally, the descriptive statistics reported for the fourth grade were identical to the descrip-tive statistics for the school type of Grundschule.

In agreement with the principle that the exclusive consideration of extreme groups in-flates correlations with respect to the state of affairs in the respective population, the zero-order correlation between socio-economic status and post-instructional science achievement in the subsample of sixth-graders, r = .24, exceeded the corresponding zero-order correlation, r = .16, in the subsample of fourth-graders (see Table 19). Naturally, in the data at hand the inflationary effects of the use of extreme groups were confounded with true developmental variations. However, in similar form, this is the case for the effects of school types in all lon-gitudinal studies of German secondary education that do not feature a comprehensive school type, e.g., the school types of prolonged Grundschule or Gesamtschule, as a reference group.

Alternatively, the disparities between the bivariate correlations for the subsamples of fourth-graders and sixth-graders might be viewed as evidence in favor of an accumulation of primary and secondary effects of social background in German secondary education. With regard to the associations between family socio-economic status and the cognitive propensities as potential mediators, zero-order correlations for the subsample of sixth-graders outstripped corresponding correlations for the subsample of fourth-graders. For the motivational propensities the relation of the bivariate correlations to socio-economic status was reversed, correlations in the subsample of fourth-graders exceeding correlations in the subsample of sixth-graders, although correlations between socio-economic status and individual interest in science were unanimously close to zero. The same pattern of zero-order correlations emerged for the associations between potential mediators and post-instructional science achievement. Bivariate Correlations for the associations between cognitive mediators and science achievement in sixth grade surpassed corresponding correlations in fourth grade, whereas relations between motivational mediators and science achievement were closer in fourth grade than in sixth grade.

Table 19
Intercorrelations of the Measures by Grade Level and School Type

Variable	SES	CFT	SAT1	SCA	INT	SAT2	SES	CFT	SAT1	SCA	INT	SAT2
			Fourth Grade[a]						Sixth Grade			
SES	---	.10	.19	.15	.04	.16	---	.18	.27	.04	-.01	.24
CFT		---	.31	.17	.04	.31		---	.46	.08	.07	.42
SAT1			---	.25	.11	.58			---	.15	.04	.66
SCA				---	.48	.20				---	.48	.17
INT					---	.10					---	.07
SAT2						---						---
			Hauptschule						Gymnasium			
SES	---	.04	.12	.07	-.00	.06	---	.04	.15	-.02	.00	.13
CFT		---	.24	.03	.11	.22		---	.39	.10	.08	.31
SAT1			---	.11	-.02	.52			---	.18	.10	.59
SCA				---	.39	.14				---	.55	.20
INT					---	.02					---	.13
SAT2						---						---

Note. Computations were based on aggregation over 50 imputed data sets. SES = Socio-economic Status; CFT = Fluid Ability (assessed with CFT 20-R); SAT1 = Topic-specific Prior Knowledge; SCA = Self-concept of Ability; INT = Individual Interest in Science; SAT2 = Post-instructional Science Achievement.
[a]Naturally, the descriptive statistics reported for the fourth grade were identical to the descriptive statistics for the school type of Grundschule.

Eventually, descriptive statistics were explored for the subsamples of students from the school types of Hauptschule and Gymnasium. (In this context, remember that the subsample of students from the school type of Grundschule was identical to the subsample of fourth-graders.) Students from the school type of Hauptschule possessed the lowest average family socio-economic status (*M* = 38.93, *SD* = 19.83) of all students, whereas students from the school type of Gymnasium had the highest average family socio-economic (*M* = 53.30, *SD* = 18.62) of all students (see Table 18). The average socio-economic status of students from the school type of Grundschule fell in between (*M* = 47.69, *SD* = 19.70). The same pattern held true for students' post-instructional science achievement as well as for the cognitive propensities of fluid ability and topic-specific prior knowledge. In contrast to this, for the motivational propensities of self-concept of ability in science and individual interest in science, differences between students from the school types of Hauptschule and Gymnasium were small, while the average attitude strength on the motivational propensities in the school types of secondary education was surpassed by the average strength on the motivational propensities in the elementary school type of Grundschule. Thus, at first glance, the pattern of means for the motivational propensities lent only weak support for the conceptualization of the school type of Hauptschule as a self-worth protecting niche for low-achieving students.

The zero-order correlation between family socio-economic status and post-instructional science achievement was lowest in the school type of Hauptschule, *r* = .06, highest in the school type of Grundschule, *r* = .16, and intermediate in the school type of Gymnasium, *r* = .13 (see Table 19). It appears unlikely that restrictions of variances contributed much to this configuration as standard deviations for all measures tended to be only minimally, if at all, smaller for the secondary school types of Hauptschule and Gymnasium than for the elementary school type of Grundschule (see Table 18). Bivariate correlations between socio-economic status and cognitive propensities in the school types of secondary education were similar to each other and congenially smaller than corresponding correlations in the school type of elementary education. A small positive zero-order correlation between socio-economic status and the motivational propensity of self-concept of ability was found for the school type of Hauptschule, *r* = .07. For the school type of Gymnasium, the respective correlation was negative and close to zero, *r* = -.02. Both of these correlations fell short of the corresponding correlation obtained for the subsample of students from the school type of Grundschule, *r* = .15. Furthermore, there were no systematic associations between socio-economic status and individual interest in science in the school types of secondary education. With regard to the size of the zero-order correlations between the potential mediators and post-instructional science achievement, the subsamples of students from the school types of Grundschule and Gymnasium resembled each other. However, in the school type of Hauptschule corresponding correlations tended to be smaller. In particular, the bivariate correlation between individual interest in science and science achievement approached zero, r = .02.

Social Gradient in Post-instructional Science Achievement

In order to gauge the size of the social gradient in science achievement after instruction, four multilevel models were fitted to the data. These comprised a main-effects-only model with grade level as a covariate on the class level (Model 1A), a main-effects-only model with school type as a covariate on the class level (Model 1B), a cross-level interaction model with grade level as a covariate on the class level (Model 1C) and a cross-level interaction model with school type as a covariate on the class level (Model 1D). In particular, the main-effects-only models comprised a main effect of socio-economic status on science achievement on the individual level as well as a random intercept and a main effect of the respective covariate on the class level, hereby modeling mean differences in students' post-instructional science achievement between grade levels or school types. In contrast, besides a random intercept and a main effect of the respective covariate, cross-level interaction models contained a random slope for socio-economic status on the class level. This random slope was regressed on the respective covariate, thereby creating a cross-level interaction between socio-economic status and the respective covariate. In other words, in the cross-level interaction models the effect of socio-economic status on student achievement was allowed to vary across classrooms and it was investigated if the magnitude of this effect was moderated by grade level or school type. In the main-effects-only models the effect of socio-economic status was modeled as constant across all classrooms. So, the moderation of the social gradient in student achievement by grade level or school type could be evaluated by inspection of regression weights for cross-level interactions as well as by comparison of information criteria for corresponding main-effects-only models and cross-level interaction models.

The relevant research question located at the first layer of the structure of research questions, i.e. the detection of a hypothesized small impact of socio-economic status on student achievement, was answered by the statistical significance of the regression weight for the main effect of socio-economic status on student achievement. The associated research questions situated at the second layer of the structure of research questions included a potential amplification of the social gradient in post-instructional student achievement by grade level due to a summation of primary and secondary effects of social origin as well as a possible diminishment of the social gradient in student achievement by school type due to restrictions of variance within the school types of secondary education. The former assumption would be supported by a positive, preferably statistically significant, cross-level interaction between socio-economic status and grade level and superior model fit for the cross-level interaction model with grade level as a covariate on the class level (Model 1C). The latter hypothesis would be corroborated by negative, preferably statistically significant, cross-level interactions between socio-economic status and school type and superior model fit for the cross-level interaction model with school type as a covariate on the class level (Model 1D).

Table 20
Random Coefficient Models for the Social Gradient in Science Achievement

Parameter	Model 1A		Model 1B		Model 1C		Model 1D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.09***	0.02	0.09***	0.02				
Residual Variance	0.71***	0.02	0.71***	0.02	0.71***	0.02	0.71***	0.02
Class Level								
Random Intercept	0.08	0.07	0.08	0.5	0.08	0.07	0.08	0.05
Random Slope SES					0.10**	0.03	0.11***	0.03
SES*Sixth Grade[a]					-0.02	0.05		
SES*Hauptschule[b]							-0.05	0.06
SES*Gymnasium[b]							-0.04	0.06
Sixth Grade[a]	-0.22*	0.10			-0.22*	0.10		
Hauptschule[b]			-0.69***	0.10			-0.70***	0.10
Gymnasium[b]			0.27**	0.10			0.27**	0.10
Residual Variance	0.24***	0.04	0.13***	0.02	0.24***	0.04	0.13***	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	14557.5	15.35	**14501.5**	15.19	14559.7	14.97	14504.0	14.99
BIC	14598.8	15.35	**14548.7**	15.19	14612.7	14.97	14568.8	14.99
aBIC	14576.5	15.35	**14523.3**	15.19	14584.1	14.97	14533.9	14.99

Note. Est. = Estimate; SES = Socio-economic Status.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
$*p < .05$. $**p < .01$. $***p < .001$.

In the main-effects-only models a small social gradient in post-instructional science achievement, $B = 0.09$, $SE = 0.02$, $p < .001$ for both models, was identified; an increase of one standard deviation in family socio-economic status was associated with a gain of approximately one tenth of a standard deviation in science achievement (see Table 20). For the covariate of grade level a negative effect on science achievement was found in the main-effects-only model, $B = -0.22$, $SE = 0.10$, $p = .02$. In sixth grade students' mean science achievement after instruction was lower than in fourth grade. This, at first glance counterintuitive, finding had to be attributed to the overrepresentation of, in relation to the entire population of sixth-graders, low-achieving students from the school type of Hauptschule in the subsample of sixth-graders. Inspection of regression weights for the covariate of school type in the main-effects-only model underscored this interpretation: For students from the school type of Hauptschule mean achievement was significantly lower than for students from the school type of Grundschule, $B = -0.69$, $SE = 0.10$, $p < .001$, whereas mean achievement for students from the school type of Gymnasium was significantly higher than for students from the school type of Grundschule, $B = 0.27$, $SE = 0.10$, $p < .01$. In the cross-level interaction models the mean of the random slope for the effect of socio-economic status on student achievement, $B = 0.10$, $SE = 0.03$, $p < .01$ and $B = 0.11$, $SE = 0.03$, $p < .001$, revealed, in agreement with the main-effects-only models, on the average, a small social gradient in student achievement across classrooms. However, inspection of terms for cross-level interactions uncovered that neither grade level, $B = -0.02$, $SE = 0.05$, $p = .67$, nor school type, $B = -0.05$, $SE = 0.06$, $p = .38$ for the school type of Hauptschule and $B = -0.04$, $SE = 0.06$, $p = .48$ for the school type of Gymnasium, significantly moderated the relation between family socio-economic status and post-instructional science achievement. Accordingly, information criteria were found to be lowest for the main-effects-only model with school type as a covariate on the class level. Additional incorporation of cross-level interactions did not enhance overall model fit.

The analyses of the impact of family socio-economic status on post-instructional science achievement highlighted the presence of a reliable social gradient in science achievement (Models 1A–1D). Despite the emergence of an unsurprising pattern of disparities in mean science achievement between the three school types under investigation (Models 1B and 1D), the current explorations did not uncover an increase of the association between socio-economic status and students' post-instructional science achievement in the transition from elementary to secondary education (Model 1C). In other words, no evidence in favor of an accumulation of primary and secondary effects of social background in sixth grade was found in the data. As the unrepresentativeness of the subsample of sixth-graders due to the exclusive sampling of students from the school types of Hauptschule and Gymnasium inclined the data to progressively reveal the presence of secondary effects of social background, the examinations provided particularly convincing evidence against the relevance of secondary effects of social background for science achievement with regard to the topic of evaporation and condensation in early German secondary education.

Testing the Univariate Mediations

The identification of a mediation proceeds in three steps (Baron & Kenny, 1986). First, an association between a predictor and a dependent variable must be found, i.e., obviously, a relation that can be mediated has to be established in the first place. In case of the current analyses the social gradient for post-instructional science achievement figured as this relation. Second, it must be shown that a potential mediator is associated with the predictor. With regard to the current analyses, it had to be demonstrated that, for instance, fluid ability as a potential mediator was related to socio-economic status. Third, it has to be confirmed that the mediator is associated with the dependent variable and that this association causes a reduction or disappearance of the relation between the predictor and the dependent variable. For this purpose, the dependent variable has to be regressed on the mediator and the predictor simultaneously. The size of the indirect, i.e. mediated, effect of the predictor can be calculated by subtraction of the regression weight for the relation between the predictor and the dependent variable in the third step of the analyses from the regression weight for the corresponding relation in the first step of the analyses. Alternatively, the magnitude of the indirect effect can be computed by multiplication of the regression weight for the relation between the predictor and the mediator in the second step of the analyses with the regression weight for the relation between the mediator and the dependent variable in the third step of the analyses. Accordingly, realization of the third step of the mediation analyses for fluid ability as a potential mediator in the current investigations involved estimation of a regression model predicting post-instructional science achievement simultaneously by fluid ability and socio-economic status (see Figure 6).

Considering each of the four cognitive and motivational propensities separately as a potential mediator of the association between socio-economic status and student achievement, the current analyses basically followed the procedure outlined above. Thus, for each potential mediator the investigations began with the estimation of multilevel regression models delineating the association between and post-instructional science achievement. In other words, initially the social gradients in each of the four potential mediators were estimated. After that, for each potential mediator, multilevel regression models that incorporated family socio-economic status and the potential mediator simultaneously as predictors of science achievement were set up. These models scanned the data for the presence of a substantial relation between the potential mediator and post-instructional science achievement. Moreover, they provided estimates of the strength of the association between socio-economic status and science achievement that could be compared to the social gradient in science achievement in order to capture the occurrence and magnitude of potential mediations by cognitive and motivational propensities. In addition, some of the regression models contained class-level main

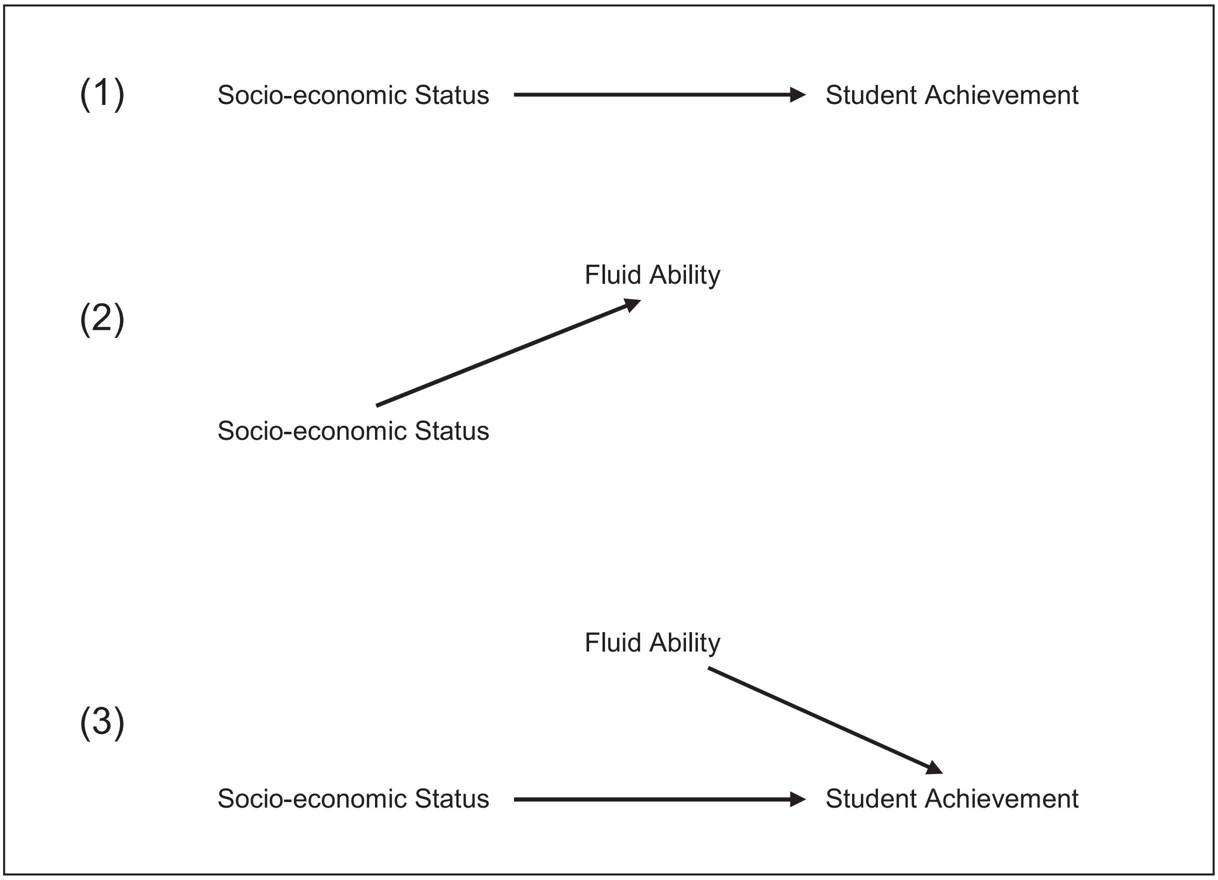

Figure 6. The three steps of a mediation analyses exemplified for fluid ability as a potential mediator of the association between socio-economic status and student achievement.

effects for either grade level or school type. These models were devised to examine the hypotheses located at the first layer of the structure of research questions; assuming constant relations between critical variables across grade levels or school types, they modeled social gradients and mediations in a context of mean differences between grade levels and school types. Other multilevel regression models also included, besides main effects for grade level or school type, cross-level interactions between grade level or school type and critical variables. These models were constructed to elucidate issues located at the second layer of the structure of research questions. Specifically, statistically significant cross-level interactions and superior overall model fit of models covering cross-level interactions – in comparison to models incorporating only main effects – were interpreted as empirical evidence for the presence of relevant moderations of relations between critical variables by grade level or school type.

Mediation by Fluid Ability

In agreement with the general procedure outlined above, the investigation of the role of fluid ability as a potential mediator of the relation between family socio-economic status and post-instructional science achievement commenced with the examination of the association between family socio-economic status and fluid ability. In other words, the social gradient in fluid ability was inspected. For this purpose, the analyses conducted for the social gradient in student achievement were repeated with fluid ability as the dependent variable. On the one hand, two main-effects-only models were fitted to the data, with grade level (Model 2A) and school type (Model 2B), respectively, as covariates on the class level. On the other hand, two cross-level interaction models were estimated, again with grade level (Model 2C) and school type (Model 2D), respectively, as covariates on the class level.

Table 21
Random Coefficient Models for the Social Gradient in Fluid Ability

Parameter	Model 2A		Model 2B		Model 2C		Model 2D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.07**	0.02	0.06*	0.02				
Residual Variance	0.79***	0.02	0.79***	0.02	0.79***	0.02	0.79***	0.02
Class Level								
Random Intercept	-0.16**	0.06	-0.16***	0.04	-0.16**	0.06	-0.16***	0.04
Random Slope SES					0.08**	0.03	0.08**	0.03
SES*Sixth Grade[a]					-0.02	0.04		
SES*Hauptschule[b]							-0.03	0.06
SES*Gymnasium[b]							-0.08	0.05
Sixth Grade[a]	0.26**	0.08			0.26*	0.09		
Hauptschule[b]			-0.18*	0.08			-0.18*	0.08
Gymnasium[b]			0.70***	0.06			0.72***	0.08
Residual Variance	0.16***	0.03	0.07***	0.02	0.16***	0.03	0.07***	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	14794.7	11.66	**14725.4**	11.72	14797.8	11.81	14727.4	12.03
BIC	14835.9	11.66	**14772.5**	11.72	14850.8	11.81	14792.2	12.03
aBIC	14813.7	11.66	**14747.1**	11.72	14822.3	11.81	14757.3	12.03

Note. Est. = Estimate; SES = Socio-economic Status.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

With respect to the first layer of the structure of research questions, a relevant association between family socio-economic status and students' fluid ability was expected. The actual presence of such a relation would be reflected by a statistically significant regression weight for the main effect of social background on fluid ability. With regard to the second layer of the structure of research questions, the association between socio-economic status and fluid ability was anticipated to amplify in the transition from elementary to secondary education. A statistically significant term for the cross-level interaction between socio-economic status and grade level would indicate the presence of such an intensification of the relation between socio-economic status and fluid ability. There were no further differential hypotheses concerning the impact of school type on this relation.

Regression weights for socio-economic status in the main-effects-only models revealed a small social gradient in fluid ability, $B = 0.07$, $SE = 0.02$, $p < .01$ and $B = 0.06$, $SE = 0.02$, $p = .01$ (see Table 21). Moreover, a positive effect of the covariate of grade level on fluid ability manifested itself in the respective main-effects-only model, $B = 0.26$, $SE = 0.08$, $p = .01$. Mean fluid ability in sixth grade surpassed mean fluid ability in fourth grade. Nonetheless, examination of regression weights for the covariate of school type in the other main-effects-only model pointed out that mean fluid ability in sixth grade differed by school type. Students from the school type of Hauptschule possessed significantly lower mean fluid ability than students from the school type of Grundschule, $B = -0.18$, $SE = 0.08$, $p = .03$. On the contrary, students from the school type of Gymnasium disposed of significantly higher mean fluid ability than students from the school type of Grundschule, $B = 0.70$, $SE = 0.08$, $p < .001$. The mean of the random slope for the effect of family socio-economic status on fluid ability estimated in the cross-level interaction models, $B = 0.08$, $SE = 0.03$, $p < .01$ for both models, was not moderated by the covariates of grade level, $B = -0.02$, $SE = 0.04$, $p = .59$, and school type, $B = -0.03$, $SE = 0.06$, $p = .56$ for the school type of Hauptschule and $B = -0.08$, $SE = 0.05$, $p = .13$ for the school type of Gymnasium. In other words, the strength of the association between socio-economic status and fluid ability did not differ by grade level or school type. In agreement with this, the inclusion of terms for cross-level interactions did not improve overall model fit so that the inspection of information criteria uncovered the main-effects-only model with grade level as the covariate on the class level as fitting best to the data. In summary, the results of the analyses for the social gradient in fluid ability met the second condition for the identification of a mediating relation, the establishment of an association between the predictor, in this case socio-economic status, and the mediator, in this case fluid ability.

As the concluding step of the investigation of fluid ability as a potential mediator of the relation between socio-economic status and science achievement, four multilevel models containing simultaneously socio-economic status and fluid ability as precursors of student achievement were estimated. Relying on the results of the examination of the social gradient in science achievement, in all four models the effect of socio-economic status on science achievement was conceptualized as constant across classrooms. In contrast, the models differed with respect to the inclusion of a moderation of the effect of fluid ability on science achievement by grade level or school type. A main-effects-only model comprising main effects for socio-economic status, fluid ability and grade level (Model 3A) as well as a main-effects-only model incorporating main effects for socio-economic status, fluid ability and school type (Model 3B) were estimated. Furthermore, two cross-level interaction models were fitted to the data. Besides main effects for socio-economic status and grade level (Model 3C) or school type (Model 3D), these models contained a random slope for fluid ability and a term for the cross-level interaction between fluid ability and the respective covariate.

With regard to the first layer of the structure of research questions, it was assumed that students' fluid ability mediated a relevant portion of the effect of socio-economic status on post-instructional science achievement. On the one hand, this entailed expecting a statistically significant regression weight for the main effect of fluid ability on post-instructional achievement. On the other hand, it included anticipating a sizable reduction of the effect of socio-economic status on achievement, relative to the previously estimated pure social gradient in post-instructional achievement. With regard to the second layer of the structure of research questions, the association between students' fluid ability and post-instructional science achievement was expected to attenuate in the transition from elementary to secondary education, both for the entire subsample of sixth-graders and within the school types of Hauptschule and Gymnasium. The presence of such moderations would be indicated by statistically significant terms for the cross-level interactions between fluid ability and grade level as well as fluid ability and school type, respectively. The detection of superior model fit – by particularly low information criteria – for models with terms for cross-level interactions would point into the same direction.

Within the main-effects-only models a moderate effect of fluid ability on post-instructional science achievement emerged, $B = 0.26$, $SE = 0.02$, $p < .001$ and $B = 0.25$, $SE = 0.02$, $p < .001$, respectively; an increase of one standard deviation in fluid ability was accompanied by an improvement of one quarter of a standard deviation in science achievement (see Table 22). Simultaneously, a small, yet statistically significant, effect of socio-economic status on science achievement was found in the main-effects-only models, $B = 0.08$, $SE = 0.02$, $p < .001$ and $B = 0.07$, $SE = 0.02$, $p < .01$, respectively. The mean of the random slope for fluid ability within the cross-level interaction models, $B = 0.26$, $SE = 0.03$, $p < .001$ for both models, closely resembled the estimates of the effect of fluid ability obtained from the main-effects-only models. The term for the cross-level interaction between fluid ability and grade level was found to be statistically insignificant, $B = 0.00$, $SE = 0.04$, $p = .96$. With regard to the cross-level interaction for the covariate of school type, the picture was slightly more complex. The mean of the random slope for fluid ability in the subsample of students from the school type of Grundschule did not differ significantly from the mean of the random slope for fluid ability in the subsamples of students from the school types of Hauptschule, $B = -0.09$, $SE = 0.05$, $p = .11$, and Gymnasium, $B = 0.02$, $SE = 0.05$, $p = .62$. Nevertheless, the difference between the regression weights for the terms of the cross-level interactions for the school types of Hauptschule and Gymnasium exceeded the corresponding standard errors more than two times. In other words, would have the school type of Hauptschule been chosen as the reference category, the term for the cross-level interaction for the school type of Gymnasium would have reached statistical significance. Notably, the background pattern of the effects of the class-level covariates deviated in one aspect from the pattern obtained for the prediction of post-instructional science achievement without additional predictors besides socio-economic status on the individual level: The regression weights for the school type of Gymnasium, $B = 0.09$, $SE = 0.09$, $p = .30$ and $B = 0.07$, $SE = 0.09$, $p = .46$, revealed that there was no substantial difference in post-instructional science achievement between students from the school types of Grundschule and Gymnasium when controlling for the impact of students' differentially distributed fluid ability. The AIC identified the cross-level interaction model with school type as the covariate on the class level as fitting best to the data. However, penalizing the number of model parameters stronger than the AIC, the BIC and the aBIC indicated that the main-effects-only model with school type as the covariate on the class level summarized the data best.

As family socio-economic status remained a statistically significant predictor of students' science achievement in face of the simultaneous incorporation of fluid ability as a precursor of achievement, the preceding analyses indicated that students' fluid ability mediated the relation between socio-economic status and achievement at most partially. A comparison of the social gradients in science achievement obtained without individual-level propensities as concurrent predictors (Models 1A–1B) with the effects of socio-economic status on science achievement in the preceding multilevel models (Models 3A–3D) revealed that approximately one or two tenths of the effect of socio-economic status on achievement were mediated by fluid ability. An increase of one standard deviation in fluid ability entailed a mediated effect of socio-economic status on achievement of one or two percent of one standard deviation in post-instructional science achievement. This partial mediation occurred in the context of an attrition of the effect of fluid ability on achievement in the school type of Hauptschule relative to the school type of Gymnasium though the effect of fluid ability in both school types of secondary education did not differ substantially from the effect of fluid ability in the school type of Grundschule.

Table 22
Random Coefficient Models for Socio-economic Status and Fluid Ability as Predictors of Student Achievement

Parameter	Model 3A		Model 3B		Model 3C		Model 3D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.08***	0.02	0.07**	0.02	0.08***	0.02	0.07***	0.02
CFT	0.26***	0.02	0.25***	0.02				
Residual Variance	0.67***	0.02	0.67***	0.02	0.66***	0.02	0.66***	0.02
Class Level								
Random Intercept	0.12*	0.06	0.12*	0.05	0.12*	0.06	0.12*	0.05
Random Slope CFT					0.26***	0.03	0.26***	0.03
CFT*Sixth Grade[a]					0.00	0.04		
CFT*Hauptschule[b]							-0.09	0.05
CFT*Gymnasium[b]							0.02	0.05
Sixth Grade[a]	-0.29**	0.09			-0.29**	0.09		
Hauptschule[b]			-0.65***	0.09			-0.68***	0.09
Gymnasium[b]			0.09	0.09			0.07	0.09
Residual Variance	0.18***	0.03	0.11***	0.02	0.17***	0.03	0.11*	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	6749.0	19.73	6708.5	19.40	6749.5	19.81	**6706.4**	19.56
BIC	6784.4	19.73	**6749.7**	19.40	6796.6	19.81	6765.3	19.56
aBIC	6765.3	19.73	**6727.5**	19.40	6771.2	19.81	6733.5	19.56

Note. Est. = Estimate; SES = Socio-economic Status; CFT = Fluid Ability (assessed with CFT 20-R).
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
**p* < .05. ***p* < .01. ****p* < .001.

Mediation by Topic-specific Prior Knowledge

Students' topic-specific prior knowledge represented the second cognitive propensity that was examined as a potential mediator of the association between socio-economic status and science achievement. As for the exploration associated with the cognitive propensity of fluid ability, the investigation of the mediating role of prior knowledge started with the estimation of a sequence of four multilevel models dedicated to the assessment of the social gradient in prior knowledge. These were two main-effect-only models, either with grade level (Model 4A) or school type (Model 4B) as a covariate on the class level, and two cross-level interaction models, again either with grade level (Model 4C) or school type (Model 4D) as a covariate on the class level.

With regard to the first layer of the structure of research questions, a close association between family socio-economic status and topic-specific prior knowledge was anticipated. In other words, the observation of a large and statistically significant regression weight for the main effect of social background on prior knowledge was expected. With respect to the second layer of the structure of research questions, the social gradient in prior knowledge was assumed to intensify in the transition from elementary to secondary education due to a summation of primary and secondary effects of social origin. A corroboration of this assumption would be indicated by a statistically significant regression weight for the cross-level interaction between socio-economic status and grade level. Additionally, the detection of superior model fit for the model with a term for the cross-level interaction between socio-economic status and grade level would point into the same direction. There were no further differential hypotheses for the moderation of the social gradient in prior knowledge by school type.

In the main-effects-only models, statistically significant regression weights for socio-economic status indicated the presence of a social gradient in topic-specific prior knowledge, $B = 0.13$, $SE = 0.02$, $p < .001$ and $B = 0.12$, $SE = 0.02$, $p < .001$, which surpassed the social gradient in post-instructional science achievement (see Table 23). Similarly, the means of the random slopes for socio-economic status in the cross-level interaction models gave evidence in favor of a comparatively pronounced social gradient in prior knowledge, $B = 0.14$, $SE = 0.03$, $p < .001$ and $B = 0.15$, $SE = 0.02$, $p < .001$. Apparently, the strength of this social gradient was moderated neither by grade level, $B = -0.02$, $SE = 0.05$, $p = .66$, nor by school type, with $B = -0.05$, $SE = 0.06$, $p = .42$ for the school type of Hauptschule and $B = -0.04$, $SE = 0.05$, $p = .44$ for the school type of Gymnasium. The regression weights for the class-level covariate of grade level, $B = 0.15$, $SE = 0.09$, $p = .10$ for both models, remaining statistically insignificant, there was no substantial difference in average topic-specific prior knowledge between fourth- and sixth-graders. However, average prior knowledge differed significantly between all school types because average prior knowledge in the school type of Hauptschule fell significantly below average prior knowledge in the school type of Grundschule, $B = -0.31$, $SE = 0.09$, $p < .01$ for both models, whereas average prior knowledge in the school type of

Gymnasium significantly exceeded average prior knowledge in the school type of Grundschule, $B = 0.63$, $SE = 0.09$, $p < .001$ for both models. In agreement with both the absence of a moderation of the social gradient in prior knowledge by grade level or school type and the presence of disparities in average prior knowledge between school types, information criteria were found to return the lowest values for the main-effects-only model with school type as the class-level covariate. With respect to the exploration of a potential mediation, the investigation of the social gradient in prior knowledge established the presence of a relation between socio-economic status, the predictor, and topic-specific prior knowledge, the potential mediator.

Table 23
Random Coefficient Models for the Social Gradient in Prior Knowledge

Parameter	Model 4A		Model 4B		Model 4C		Model 4D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.13***	0.02	0.12***	0.02				
Residual Variance	0.72***	0.02	0.72***	0.02	0.71***	0.02	0.71***	0.02
Class Level								
Random Intercept	-0.11†	0.06	-0.11*	0.05	-0.11†	0.06	-0.11*	0.05
Random Slope SES					0.14***	0.03	0.15***	0.03
SES*Sixth Grade[a]					-0.02	0.05		
SES*Hauptschule[b]							-0.05	0.06
SES*Gymnasium[b]							-0.04	0.05
Sixth Grade[a]	0.15	0.09			0.15	0.09		
Hauptschule[b]			-0.31**	0.09			-0.31**	0.09
Gymnasium[b]			0.63***	0.09			0.63***	0.09
Residual Variance	0.21***	0.03	0.11***	0.02	0.21***	0.03	0.11***	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	14579.2	16.48	**14518.9**	16.16	14580.4	16.31	14521.0	15.84
BIC	14620.5	16.48	**14566.0**	16.16	14633.5	16.31	14585.8	15.84
aBIC	14598.2	16.48	**14540.6**	16.16	14604.9	16.31	14550.8	15.84

Note. Est. = Estimate; SES = Socio-economic Status.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

In order to assess the potential mediation of the effects of socio-economic status on science achievement by prior knowledge, four multilevel models featuring family socio-economic status and topic-specific prior knowledge as concurrent predictors of achievement were set up. In particular, the estimation of these models allowed the observation of the impact of topic-specific prior knowledge on the social gradient in post-instructional science achievement. The influence of family socio-economic status on science achievement was modeled as constant across classrooms. In other words, none of the four models contained a random slope for socio-economic status and terms for corresponding cross-level interactions. Nonetheless, models varied with regard to the incorporation of terms for cross-level interactions between prior knowledge and grade level or school type. Specifically, there were two main-effects-only models containing main effects for socio-economic status, prior knowledge and grade level (Model 5A) or school type (Model 5B). Moreover, there were two cross-level interaction models covering main effects for socio-economic status and grade level (Model 5C) or school type (Model 5D) as well as a random slope for prior knowledge and a cross-level interaction between prior knowledge and the respective class-level covariate.

With respect to the first layer of the structure of research questions, it was suspected that students' topic-specific prior knowledge mediated a major portion of the effect of socio-economic status on post-instructional science achievement. This included specific expectations for two regression weights. First, a large and statistically significant regression weight for the main effect of topic-specific prior knowledge on post-instructional achievement was anticipated as prior knowledge was assumed to be a powerful predictor of achievement. Second, relative to the previously estimated sheer social gradient in science achievement, a large diminishment of the regression weight for the main effect of socio-economic status on science achievement was expected, plausibly rendering it statistically insignificant. With regard to the second layer of the structure of research questions, an amplification of the relation between prior knowledge and post-instructional science achievement in the transition from elementary and secondary education was expected as a result of the accumulation of effects of prior knowledge. Furthermore, due to differences in mean prior knowledge, it was assumed that the relation between prior knowledge and achievement would be stronger within the school type of Gymnasium than within the school type of Hauptschule. The actual occurrence of these moderations would be indicated by the observation of statistically significant regression weights for the corresponding cross-level interactions. Moreover, the identification of the model with the best overall model fit via the inspection of information criteria would reveal the most important moderation.

In the main-effects-only models topic-specific prior knowledge proved to be a particularly powerful predictor of post-instructional science achievement. Specifically, the regression weights, $B = 0.55$, $SE = 0.02$, $p < .001$ and $B = 0.54$, $SE = 0.02$, $p < .001$, respectively, indicated that an increment of one standard deviation in topic-specific prior knowledge entailed an enhancement of more than half of a standard deviation in post-instructional science achievement (see Table 24). At the same time, the corresponding regression weights for socio-economic status, $B = 0.03$, $SE = 0.02$, $p = .15$ and $B = 0.02$, $SE = 0.02$, $p = .28$, respectively, failed to attain statistical significance. The means of the random slopes retrieved for the cross-level interaction models, $B = 0.56$, $SE = 0.03$, $p < .001$ for both models, reflected a pronounced general effect of prior knowledge on science achievement as well. This effect was not moderated by grade level, $B = -0.03$, $SE = 0.04$, $p = .51$. With regard to the class-level covariate of school type, a tendency towards a moderation of the effects of prior knowledge was found. Both in the school type of Hauptschule, $B = -0.10$, $SE = 0.06$, $p = .07$, and in the school type of Gymnasium, $B = -0.02$, $SE = 0.05$, $p = .70$, the effect of prior knowledge on science achievement appeared to be weaker than in the school type of Grundschule though only for the former this disparity possessed marginal statistical significance. The pattern of background effects of class-level covariates resembled the corresponding pattern obtained for the investigations of fluid ability as a potential mediator. Under control of students' topic-specific prior knowledge, which was differentially distributed across school types, there was no substantial disparity in post-instructional science achievement between students from the school types of Grundschule and Gymnasium as indicated by the regression weights for the school type of Gymnasium, $B = -0.07$, $SE = 0.07$, $p = .33$ and $B = -0.08$, $SE = 0.08$, $p = .28$, respectively. Unanimously, all three information criteria computed for the models, the AIC, the BIC and the aBIC, identified the cross-level interaction model with school type as a covariate on the class level as summarizing the data most parsimoniously.

The inclusion of topic-specific prior knowledge as a concurrent predictor rendering the effect of socio-economic status on science achievement statistically insignificant, the preceding analyses yielded evidence in favor of a full mediation of the effect of socio-economic status on achievement by prior knowledge. In terms of the reduction of the regression weight for socio-economic status relative to the social gradient in post-instructional science achievement (Models 1A–1B), two thirds to three quarters of the effect of socio-economic status on achievement were mediated by prior knowledge (Models 5A–5D). Accordingly, an increase of one standard deviation in topic-specific prior knowledge implied an increase of approximately six percent of one standard deviation in post-instructional science achievement attributable to the mediation of effects of socio-economic status. Nevertheless, there was a marginally significant tendency for the influence of prior knowledge on science achievement to be weaker in the school type of Hauptschule than in the school type of Grundschule. Corresponding information criteria unambiguously corroborated the relevance of the cross-level interaction between topic-specific prior knowledge and school type.

Table 24
Random Coefficient Models for Socio-economic Status and Prior Knowledge as Predictors of Student Achievement

Parameter	Model 5A		Model 5B		Model 5C		Model 5D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02
SAT1	0.55***	0.02	0.54***	0.02				
Residual Variance	0.51***	0.02	0.51***	0.02	0.49***	0.02	0.49***	0.02
Class Level								
Random Intercept	0.14**	0.05	0.14**	0.04	0.14**	0.05	0.14**	0.04
Random Slope SAT1					0.56***	0.03	0.56***	0.03
SAT1*Sixth Grade[a]					-0.03	0.04		
SAT1*Hauptschule[b]							-0.10[†]	0.06
SAT1*Gymnasium[b]							-0.02	0.05
Sixth Grade[a]	-0.30***	0.07			-0.31***	0.07		
Hauptschule[b]			-0.53***	0.07			-0.56***	0.08
Gymnasium[b]			-0.07	0.07			-0.08	0.08
Residual Variance	0.10***	0.02	0.07***	0.01	0.10***	0.02	0.07***	0.01
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	5989.8	33.68	5966.5	33.15	5971.6	31.61	**5946.6**	31.03
BIC	6025.2	33.68	6007.6	33.15	6018.8	31.61	**6005.6**	31.03
aBIC	6006.1	33.68	5985.4	33.15	5993.4	31.61	**5973.8**	31.03

Note. Est. = Estimate; SES = Socio-economic Status; SAT1 = Topic-specific Prior Knowledge.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
[†]$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Mediation by Self-concept of Ability in Science

The exploration of the potentially mediating function of motivational propensities for the relation between socio-economic status and science achievement began with an investigation of the role of students' self-concept of ability in science. As the first step of this investigation, the social gradient in self-concept of ability was examined. Specifically, this examination comprised the estimation of four multilevel models featuring self-concept of ability in science as the dependent variable. These were the estimations of a main-effects-only model with grade level as a class-level covariate (Model 6A), a main-effects-only model with school type as a class-level covariate (Model 6B), a cross-level interaction model with grade level as a class-level covariate (Model 6C) and, finally, a cross-level interaction model with school type as a class-level covariate (Model 6D).

Table 25
Random Coefficient Models for the Social Gradient in Self-concept of Ability in Science

Parameter	Model 6A		Model 6B		Model 6C		Model 6D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.08**	0.02	0.08**	0.02				
Residual Variance	0.85***	0.02	0.85***	0.02	0.85***	0.03	0.85***	0.03
Class Level								
Random Intercept	0.12*	0.05	0.12*	0.05	0.12*	0.05	0.12*	0.05
Random Slope SES					0.14***	0.03	0.14***	0.03
SES*Sixth Grade[a]					-0.12**	0.05		
SES*Hauptschule[b]							-0.09	0.07
SES*Gymnasium[b]							-0.16**	0.06
Sixth Grade[a]	-0.26**	0.08			-0.26**	0.08		
Hauptschule[b]			-0.28**	0.10			-0.29**	0.10
Gymnasium[b]			-0.24*	0.09			-0.21*	0.09
Residual Variance	0.12***	0.02	0.12***	0.02	0.12***	0.02	0.12***	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	14967.4	13.14	14969.3	13.14	**14960.3**	**12.86**	14961.7	12.42
BIC	**15008.7**	**13.14**	15016.4	13.14	15013.3	12.86	15026.6	12.42
aBIC	14986.4	13.14	14991.0	13.14	**14984.7**	**12.86**	14991.6	12.42

Note. Est. = Estimate; SES = Socio-economic Status.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
**p* < .05. ***p* < .01. ****p* < .001.

With regard to the first layer of the structure of research questions, it was expected to observe a relevant social gradient in students' self-concept of ability in science. Specifically, this entailed the expectation to uncover a statistically significant regression weight for the main effect of family socio-economic status on self-concept of ability. With respect to the second layer of the structure of research questions, a reduction of the strength of the social gradient in self-concept of ability was anticipated. This reduction should be particularly pronounced for the school type of Gymnasium. The confirmation of these hypotheses regarding potential moderations would, on the one hand, entail the observation of negative and statistically significant regression weights for the cross-level interactions between social background and grade level as well as between social background and school type of Gymnasium, respectively. On the other hand, superior overall model fit in terms of smallest values for information criteria would reveal the most relevant moderation.

Regression weights for socio-economic status obtained in the main-effects-only models, $B = 0.08$, $SE = 0.02$, $p < .01$ for both models, indicated a small effect of family socio-economic status on students' self-concept of ability in science (see Table 25). However, inspection of parameter estimates for the cross-level interaction models revealed that this effect was moderated crucially by grade level and school type. The means of the random slopes for socio-economic status, retrieved from the cross-level interaction models, $B = 0.14$, $SE = 0.03$, $p < .001$ for both models, showed that the size of the association between socio-economic status and students' self-concept of ability in the reference category, i.e. the subsample of fourth-graders and, consequentially, students from the school type of Grundschule exceeded the strength of the corresponding effect in the main-effects-only models. Moreover, a statistically significant cross-level interaction between socio-economic status and grade level with a negative regression coefficient, $B = -0.12$, $SE = 0.05$, $p < .01$, indicated that the size of the social gradient in self-concept of ability differed between grade levels, resulting for sixth-grade students in a net effect of socio-economic status of almost zero. In agreement with this, the cross-level interaction terms for the school types of Hauptschule, $B = -0.09$ $SE = 0.07$, $p = .17$, and Gymnasium, $B = -0.16$, $SE = 0.06$, $p < .01$, revealed that the social gradients in self-concept of ability tended to approach zero for the subsamples of students from the school types of Hauptschule and Gymnasium although only for the latter subsample disparity in the social gradient from the reference category of students from the school type of Grundschule possessed statistical significance. These variations of the size of the social gradient in self-concept of ability across grade levels and school types were modeled against a background of mean differences in self-concept of ability between grade levels and school types. There was a negative regression weight for the class-level covariate of grade level, $B = -0.26$, $SE = 0.08$, $p < .01$ for both models. In other words, students' self-concept of ability in science declined in the transition from elementary to secondary education. Moreover, according to the regression weights pertaining to the class-level covariate of school type, this decrease tended to be more pronounced in the school type of Hauptschule, e.g., $B = -0.28$, $SE = 0.10$, $p < .01$ in the main-effects-only model, than in the school type of Gymnasium, e.g., $B = -0.24$ $SE = 0.09$, $p = .01$ in the main-effects-only model. However, an inspection of the associated standard errors clarified that the disparity in the decline of students' self-concept of ability between the school types of secondary education did not bear statistical significance. The identification of the model possessing the relative best model fit according to the three information criteria was not unequivocal. The AIC and the aBIC favored the cross-level interaction model with grade level as a class-level covariate, whereas the BIC marked the main-effects-only model with grade level as a covariate on the class level as the most parsimonious model.

Eventually, family socio-economic status and students' self-concept of ability in science were analyzed as concurrent predictors of post-instructional science achievement in order to gauge the impact of self-concept of ability on the predictive relevance of socio-economic status. These analyses comprised four multilevel models. In line with the results of the investigation of the social gradient in science achievement, these models incorporated socio-economic status in terms of a main effect on the individual level, i.e. as operating with constant fortitude across classrooms. Specifically, the estimated models included two main-effects-only and two cross-level interaction models. The former contained main effects for socio-economic status, self-concept of ability in science and grade level (Model 7A) or school type (Model 7B), while the latter featured main effects for socio-economic status and grade level (Model 7C) or school type (Model 7D) as well as a random slope for self-concept of ability and a term for the cross-level interaction between self-concept of ability and the respective covariate.

Specifically, on the first layer of the structure of research questions, students' self-concept of ability in science was expected to mediate a fraction of the impact of socio-economic status on post-instructional science achievement. Accordingly, it was assumed that a statistically significant regression weight for the main effect of self-concept of ability on science achievement would be retrieved. Moreover, it was anticipated that the size of the regression weight for the main effect of social background on achievement would be diminished considerably, relative to the estimation of the bare social gradient in post-instructional science achievement. On the second layer of the structure of research questions, due to the occurrence of the big-fish-little-pond effect, the association between self-concept of ability in science and post-instructional science achievement was suspected to attenuate in the transition from elementary to secondary education. The detection of a negative and statistically significant cross-level interaction between self-concept of ability and grade level in the prediction of science achievement would support this hypothesis. Likewise, the observation of a superior model fit for the cross-level interaction model with school type as a covariate would argue in favor of the hypothesis of an attenuating relation.

In the main-effects-only models, students' self-concept of ability in science exerted a moderate influence on post-instructional science achievement, $B = 0.19$, $SE = 0.02$, $p < .001$

for both models (see Table 26). In other words, in these models an increment of one standard deviation in self-concept of ability was associated with a growth of post-instructional science achievement of approximately one fifth of a standard deviation. Within the cross-level interaction models the situation was similar. The means of the random slope for self-concept of ability in science hinted at a moderate average effect of self-concept on science achievement, $B = 0.18$, $SE = 0.03$, $p < .001$ for both models, which was moderated neither by grade level, $B = 0.03$, $SE = 0.04$, $p = .44$, nor by school type, $B = -0.02$, $SE = 0.05$, $p = .66$ for the school type of Hauptschule and $B = 0.06$, $SE = 0.05$, $p = .17$ for the school type of Gymnasium. There were statistically significant main effects for family socio-economic status in all four multilevel models, e.g., $B = 0.08$, $SE = 0.02$, $p < .001$ for the main-effects-only model with the covariate of grade level. By and large, the background effects obtained for the class-level covariates mirrored the results found in the analyses of the social gradient of science achievement, i.e. under control of students' self-concept of ability in science there were characteristic mean differences in science achievement between grade levels and school types. On the average, fourth-graders tended to outperform sixth-graders, e.g., $B = -0.17$, $SE = 0.10$, $p = .08$ in the main-effects-only model. Moreover, average science performance in the school type of Hauptschule fell short of average performance in the school type of Grundschule, e.g., $B = -0.64$, $SE = 0.10$, $p < .001$ in the main-effects-only model. At the same time, average science achievement in the school type of Gymnasium exceeded average science achievement in the school type of Grundschule, e.g., $B = 0.31$, $SE = 0.10$, $p < .01$ in the main-effects-only model. The lowest values for all three information criteria were obtained for the main-effects-only model with school type as a covariate on the class level (Model 7B). Thereby it was identified unambiguously as the most parsimonious model of all four multilevel models.

Despite the incorporation of the motivational propensity of students' self-concept of ability in science as a concurrent predictor into the mediation models, the effect of family socio-economic status on post-instructional science achievement remained statistically significant. In other words, at most, a partial mediation of the effects of socio-economic status on science achievement by self-concept of ability in science could be observed. A comparison of the regression weight for socio-economic status in the preceding mediation models (Models 7A–7D) with the social gradient in post-instructional science achievement (Models 1A–1B) revealed that between slightly more than one tenth and slightly more than two tenth of the effect of socio-economic status on achievement were mediated by students' self-concept of ability in science. Thus, an increment of one standard deviation in self-concept of ability in science entailed an increment of one or two percent of one standard deviation in post-instructional science achievement associated with the mediation of effects of family socio-economic status. The effect of socio-economic status on science achievement was not moderated by grade level or school type. However, the partial mediation identified had to be viewed in conjunction with differential social gradients in students' self-concept of ability; though there was a sizeable social gradient in self-concept of ability for the subsample of fourth-graders, socio-economic status and self-concept of ability were unrelated within the subsample of sixth-graders.

Table 26
Random Coefficient Models for Socio-economic Status and Self-concept of Ability in Science as Predictors of Student Achievement

Parameter	Model 7A		Model 7B		Model 7C		Model 7D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.08***	0.02	0.07**	0.02	0.08***	0.02	0.07**	0.02
SCA	0.19***	0.02	0.19***	0.02				
Residual Variance	0.68***	0.02	0.68***	0.02	0.67***	0.02	0.67***	0.02
Class Level								
Random Intercept	0.06	0.07	0.06	0.05	0.06	0.07	0.06	0.05
Random Slope SCA					0.18***	0.03	0.18***	0.03
SCA*Sixth Grade[a]					0.03	0.04		
SCA*Hauptschule[b]							-0.02	0.05
SCA*Gymnasium[b]							0.06	0.05
Sixth Grade[a]	-0.17†	0.10			-0.17†	0.10		
Hauptschule[b]			-0.64***	0.10			-0.65***	0.10
Gymnasium[b]			0.31**	0.10			0.32**	0.10
Residual Variance	0.25***	0.04	0.14***	0.02	0.25***	0.04	0.14***	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	6830.2	16.37	**6776.6**	16.53	6831.5	16.63	6777.6	16.62
BIC	6865.5	16.37	**6817.8**	16.53	6878.6	16.63	6836.6	16.62
aBIC	6846.5	16.37	**6795.6**	16.53	6853.2	16.63	6804.8	16.62

Note. Est. = Estimate; SES = Socio-economic Status; SCA = Self-concept of Ability.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Mediation by Individual Interest in Science

Although the preceding investigation of zero-order correlations between family socio-economic status and individual interest in science had given little indication of a mediation of effects of social background by individual interest, for the sake of congruity, the analyses conducted for the other potential mediators were repeated for the motivational propensity of individual interest in science as well. Specifically, the social gradient in individual interest in science was evaluated by fitting main-effects-only models with grade level (Model 8A) and school type (Model 8B) as covariates on the class-level and cross-level interaction models with grade level (Model 8C) and school type (Model 8D) as covariates on the class-level to the data.

Table 27
Random Coefficient Models for the Social Gradient in Individual Interest in Science

Parameter	Model 8A		Model 8B		Model 8C		Model 8D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.00	0.02	0.00	0.02				
Residual Variance	0.79***	0.02	0.79***	0.02	0.79***	0.02	0.79***	0.02
Class Level								
Random Intercept	0.28***	0.05	0.28***	0.05	0.28***	0.05	0.28***	0.05
Random Slope SES					0.02	0.03	0.02	0.03
SES*Sixth Grade[a]					-0.04	0.05		
SES*Hauptschule[b]							-0.02	0.06
SES*Gymnasium[b]							-0.05	0.06
Sixth Grade[a]	-0.56***	0.08			-0.56***	0.08		
Hauptschule[b]			-0.54***	0.10			-0.54***	0.10
Gymnasium[b]			-0.57***	0.10			-0.56***	0.10
Residual Variance	0.13***	0.02	0.13***	0.02	0.13***	0.02	0.13***	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	**14781.6**	15.10	14783.5	15.02	14783.9	15.03	14786.9	15.07
BIC	**14822.9**	15.10	14830.6	15.02	14836.9	15.03	14851.7	15.07
aBIC	**14800.7**	15.10	14805.2	15.02	14808.3	15.03	14816.8	15.07

Note. Est. = Estimate; SES = Socio-economic Status.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
*p < .05. **p < .01. ***p < .001.

With regard to the first layer of the structure of research questions, it was expected to uncover a relevant social gradient in individual interest in science, i.e. to obtain a statistically significant regression weight for the main effect of family socio-economic status on students' individual interest in science. With respect to the second layer of the structure of research questions, it was speculated that the accumulation of adverse effects of social background with students' age should lead to an intensification of the social gradient in individual interest in science in the transition from elementary to secondary education. Obtaining a statistically significant regression weight for the cross-level interaction between socio-economic status and grade level – in conjunction with a superior model fit for the cross-level interaction model with grade level as a covariate – would flag support for this hypothesis.

Neither in the main-effects-only models, $B = 0.00$, $SE = 0.02$, $p = .96$ and $B = 0.00$, $SE = 0.02$, $p = .93$, respectively, nor in the cross-level interaction models, $B = 0.02$, $SE = 0.03$, $p = .54$ for both models, a systematic relation between family socio-economic status and individual in interest in science was found (see Table 27). Furthermore, there was no evidence in form of statistically significant cross-level interactions in favor of a substantial variation of the effects of socio-economic status on individual interest by grade level, $B = -0.04$, $SE = 0.05$, $p = .43$, or school type, $B = -0.02$, $SE = 0.06$, $p = .78$ for the school type of Hauptschule and $B = -0.05$, $SE = 0.06$, $p = .37$ for the school type of Gymnasium. However, a negative effect for the covariate of grade level indicated large disparities in average individual interest between fourth and sixth grade, e.g., $B = -0.56$, $SE = 0.08$, $p < .001$ in the corresponding main-effects-only model (Model 8A), constituting evidence in favor of a substantial decline of individual interest in science in the transition from elementary to secondary education. Notably, the size of this decline did not differ much by school type, e.g., $B = -0.54$, $SE = 0.10$, $p < .001$ for the school type of Hauptschule and $B = -0.57$, $SE = 0.10$, $p < .001$ for the school type of Gymnasium in the respective main-effects-only model (Model 8B). In agreement with the absence of mean differences in individual interest between school types as well as the absence of variations of effects of socio-economic status on individual interest by grade level and school type, the investigation of information criteria disclosed the main-effects-only model with grade level as the covariate on the class level as possessing the relative best fit to the data. Crucially, the analyses revealed that there was no substantial social gradient in individual interest in science. Thus, a central prerequisite for the establishment of a mediation of effects of socio-economic status by individual interest was not fulfilled. Apart from that, evidence in favor of a general decline of individual interest in science in the transition from elementary to secondary education was found.

Despite the lack of a social gradient for individual interest in science, four additional multilevel models were estimated. These models included socio-economic status and individual interest as concurrent predictors of post-instructional science achievement. Thereby, as for the other mediators, the effect of socio-economic status on science achievement was construed as constant across classrooms. In particular, no cross-level interactions between socio-economic status and grade level or school type were assumed. However, candidate models varied in terms of the incorporation of cross-level interactions for the association between individual interest in science and science achievement. On the one hand, two main-effects-only models containing main effects for socio-economic status, individual interest in science and grade level (Model 9A) or school type (Model 9B) were fitted to the data. On the other hand, two cross-level interaction models covering interactions between individual interest and grade level (Model 9C) or school type (Model 9D) besides the main effects for socio-economic status and the respective class-level covariate were estimated.

With respect to the first layer of the structure of research questions, initially individual interest was expected to mediate a part of the effect of socio-economic status on post-instructional achievement. As already said, this assumption was untenable after the investigation of the absent social gradient in individual interest. Nonetheless, students' individual interest in science was anticipated to predict science achievement. In other words, it was expected to obtain a statistically significant regression weight for the main effect of individual interest on science achievement. With regard to the second layer of the structure of research questions, it was speculated that the association between individual interest and science achievement could amplify in the transition from elementary to secondary education. Specifically, a statistically significant regression weight for the cross-level interaction between individual interest and science achievement would indicate support for this speculation.

Both within the main-effects-only models, $B = 0.08$, $SE = 0.02$, $p < .001$ for both models, and within the cross-level interaction models, $B = 0.08$, $SE = 0.03$, $p < .01$ for both models, a small, yet statistically substantial effect of individual interest in science on post-instructional science achievement was identified (see Table 28). At the same time, a similarly small, but nonetheless statistically significant effect of socio-economic status on science achievement was uncovered both within the main-effects-only models and within the cross-level interaction models, $B = 0.09$, $SE = 0.02$, $p < .001$ for all four models. There was no evidence for a moderation of the effects of individual interest on science achievement by grade level, $B = -0.00$, $SE = 0.04$, $p = .93$. Moreover, the strength of the association between individual interest and science achievement for the school type of Grundschule, which functioned as the reference category, did not differ substantially from the strength of the corresponding relations for the school types of Hauptschule, $B = -0.06$, $SE = 0.05$, $p = .22$, and Gymnasium, $B = 0.04$, $SE = 0.05$, $p = .38$. Nonetheless, the difference between the regression weights for the cross-level interactions between individual interest and the school types of Hauptschule and Gymnasium surpassed the respective standard errors two times. This indicated a statistically significant disparity in the strength of the relation between individual interest and science achievement for the two selected school types of secondary education. In the subsample of students from the school type of Hauptschule the association was considerably weaker, in fact approaching zero, than in the subsample of students from the school type of Gymnasium. However, according to the information criteria of AIC, BIC and aBIC the main-effects-only model with school as a covariate on the class level (Model 9B) displayed the best relative fit to the data. Of course, as there was no association between socio-economic status and individual interest in science in the first place, a reduction of the social gradient in science achievement due to the inclusion of individual interest as predictor of science achievement was not observable in any of the models. The effects of socio-economic status on science achievement were not mediated by individual interest in science.

Table 28
Random Coefficient Models for Socio-economic Status and Individual Interest in Science as Predictors of Student Achievement

Parameter	Model 9A		Model 9B		Model 9C		Model 9D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.09***	0.02	0.09***	0.02	0.09***	0.02	0.09***	0.02
INT	0.08***	0.02	0.08***	0.02				
Residual Variance	0.71***	0.02	0.71***	0.02	0.70***	0.02	0.70***	0.02
Class Level								
Random Intercept	0.06	0.07	0.06	0.05	0.06	0.07	0.06	0.05
Random Slope INT					0.08**	0.03	0.08**	0.03
INT*Sixth Grade[a]					-0.00	0.04		
INT*Hauptschule[b]							-0.06	0.05
INT*Gymnasium[b]							0.04	0.05
Sixth Grade[a]	-0.18†	0.10			-0.18†	0.10		
Hauptschule[b]			-0.65***	0.10			-0.67***	0.10
Gymnasium[b]			0.31**	0.10			0.32**	0.10
Residual Variance	0.24***	0.04	0.13***	0.02	0.24***	0.04	0.13***	0.02
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	6933.8	16.06	**6876.9**	15.92	6937.0	16.27	6877.9	16.60
BIC	6969.1	16.06	**6918.1**	15.92	6984.1	16.27	6936.8	16.60
aBIC	6950.1	16.06	**6895.9**	15.92	6958.7	16.27	6905.0	16.60

Note. Est. = Estimate; SES = Socio-economic Status; INT = Individual Interest in Science.
[a]Reference category is fourth grade. [b]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
$^{\dagger}p < .10$. $*p < .05$. $**p < .01$. $***p < .001$.

Summary of Univariate Mediation Analyses

With respect to the first aim of the study, the inferential analyses established a small social gradient in post-instructional science achievement. An increment of one standard deviation in family socio-economic status entailed an increase of approximately one tenth of a standard deviation in science achievement. In agreement with the hypotheses, the size of the social gradient in science achievement found for the current data set appeared small in comparison to the size of effects of social background on science achievement reported for contemporary large-scale assessments of student competences in Germany (Bonsen et al., 2008; Stubbe et al., 2012).

The zero-order correlations between socio-economic status and post-instructional science achievement obtained for the different subsamples followed the pattern predicted by the assumed moderations of the social gradient in science achievement by grade level and school type. Specifically, for the subsample of sixth-graders, the zero-order correlation between socio-economic status and science achievement exceeded the corresponding correlation for the subsample of fourth-graders. Moreover, within the selective school types of Hauptschule and Gymnasium in secondary education, zero-order correlations between socio-economic status and science achievement were smaller than the respective correlation within the comprehensive school type of Grundschule in elementary education. Despite these descriptive results, the inferential analyses by multilevel regression models yielded no evidence in favor of a moderation of the social gradient in post-instructional science achievement by grade level or school type. Apparently, an accumulation of primary and secondary effects of social origin in the transition from elementary to secondary education was not observed for the current data set.

The second aim of the current investigations was concerned with the role of students' fluid ability as a potential mediator of the association between family socio-economic status and post-instructional science achievement. On the first layer of the structure of research questions, the assumption of a substantial social gradient in students' fluid ability found empirical support. Likewise, a comparatively strong connection between fluid ability and science achievement was uncovered, in line with expectations prior to the analyses. Nevertheless, students' fluid ability mediated only a small portion of the effect of socio-economic status on post-instructional science achievement. On the second layer of the structure of research questions, in the context of a conventional pattern of mean differences in fluid ability between grade levels as well as between school types, the social gradient in fluid ability was moderated neither by grade level nor by school type. In other words, no evidence in favor of an accumulation of adverse effects of socio-economic status on children's fluid ability was obtained. Furthermore, the association between fluid ability and post-instructional science achievement was not moderated by grade level. There were no signs of a general decline of the importance of fluid ability as a predictor of achievement in the transition from elementary to secondary education. However, the association between fluid ability and science achievement within the school type of Hauptschule was significantly smaller than the corresponding association within the school type of Gymnasium. Though, at first glance, it was surmised that this attenuation of the relation between fluid ability and achievement was attributable to a restriction of variance in fluid ability within the school type of Hauptschule, an inspection of the variances in fluid ability in the various subsamples yielded no support for this spontaneous speculation.

Unsurprisingly, in agreement with an extensive literature, students' topic-specific prior knowledge – in comparison to the other cognitive and motivational propensities – was found to be an outstandingly strong predictor of post-instructional science achievement. This finding was accompanied by the detection of a pronounced social gradient in prior knowledge. In fact, the social gradient in topic-specific prior knowledge exceeded the social gradient in post-instructional science achievement. So, in conjunction, on the first layer of the structure of research questions, students' topic-specific prior knowledge was observed to mediate a major portion of the effect of family socio-economic status on science achievement. On the second layer of the structure of research questions, no evidence in favor of a moderation of the social gradient in prior knowledge by grade level or school type was obtained. This result resembled the findings obtained for the social gradient in post-instructional science achievement. Moreover, the association between topic-specific prior knowledge and science achievement did not vary by grade level. However, some indicators – i.e. a marginally significant cross-level interaction term and corresponding information criteria for overall model fit – pointed towards a relevant attrition of the effect of prior knowledge on post-instructional science achievement within the school type of Hauptschule, attended primarily by students of low prior knowledge, relative to the state of affairs for students attending the comprehensive school type of Grundschule. On the contrary, for the school type of Gymnasium, attended primarily by students of comparatively high prior knowledge, no hints in support of an amplification of the association between prior knowledge and post-instructional science achievement were found, relative to the corresponding associations within the other school types. So, all in all, there was merely faint evidence in favor of differential strength of the relation between prior knowledge and post-instructional achievement across school types, hypothesized due to mean differences in prior knowledge, and thus in complexity of the content to be covered in instruction.

With regard to students' self-concept of ability in science as a potential mediator of the relation between socio-economic status and post-instructional science achievement the hypotheses were confirmed to some extent. On the first layer of the structure of research questions, positive associations between socio-economic status and self-concept of ability in science as well as between self-concept of ability in science and post-instructional science achievement were found. This entailed a minor mediation of effects of socio-economic status on science achievement. Furthermore, on the second layer of the structure of research questions, in agreement with prior hypotheses, a dilution of the association between social background and self-concept of ability in science was established for the subsample of sixth-graders. Moreover, also in agreement with prior hypotheses, there was a tendency for this dilution to be particularly pronounced within the presumably least ego-protecting learning environment of the school type of Gymnasium. However, contrary to the hypotheses, there was no evidence in favor of a reduction of the relation between self-concept of ability in science and science achievement in the transition from elementary to secondary education.

In fact, the dilution of the association between socio-economic status and self-concept of ability in science in the transition from elementary to secondary education was so severe that the association approached zero for the subsample of sixth-graders. Thus, the assumption of a mediation of effects of socio-economic status on science achievement by self-concept of ability proved to be viable exclusively for the subsample of fourth-graders, i.e. elementary education; an finding unexpected in its extremity prior to the analyses.

The attempt to establish individual interest in science as a mediator of the relation between family socio-economic status and post-instructional science achievement met with resound failure. Though individual interest proved to be a weak predictor of science achievement, there was no evidence in favor of the existence of a social gradient in individual interest in science, i.e. social background and individual interest in science were unrelated. Moreover, with regard to the second layer of the structure of research questions, the association between socio-economic status and individual interest in science was not moderated by grade level or school type. Likewise, the relation between individual interest in science and science achievement did not vary substantially by grade level or school type. Thus, the differential hypotheses of an intensification of the associations of individual interest with socio-economic status and science achievement in the transition from elementary to secondary education were not confirmed empirically.

Exploring the Multivariate Mediation

After the successive univariate investigations of the mediation of effects of socio-economic status on post-instructional science achievement by cognitive and motivational propensities, three further analyses were conducted to examine the multivariate mediation of effects of socio-economic status on science achievement. These analyses covered the investigation of fluid ability and topic-specific prior knowledge as concurrent mediators, the exploration of topic-specific prior knowledge and self-concept of ability in science as simultaneous mediators, and the examination of a comprehensive mediation incorporating all three relevant cognitive and motivational propensities as concurrent predictors of science achievement. (As it had been found previously to be unrelated to family socio-economic status, students' individual interest in science was not included as a mediator in the comprehensive mediation analysis.)

As outlined above in the theoretical background, fluid ability is a natural precursor of domain-specific knowledge. Therefore, it had to be assumed that students' fluid ability and prior knowledge shared common variance, i.e. that at least a portion of the effect of fluid ability on post-instructional science achievement obtained in the univariate mediation analysis was actually an effect of prior knowledge. Accordingly, the examination of the concurrent mediation of effects of socio-economic status by fluid ability and topic-specific prior knowledge served two purposes. First, it was inspected if students' fluid ability contributed significantly to the prediction of post-instructional science achievement beyond the impact of topic-specific prior knowledge. Second, it was explored if fluid ability mediated an additional portion of the effect of socio-economic status on science achievement.

As self-concept of ability and domain-specific achievement are known to be reciprocally related to each other, it was also obligatory to presume that students' self-concept of ability in science and prior knowledge shared common variance. In other words, it was sensible to assume that a – albeit probably small – portion of the impact of self-concept of ability on science achievement identified in the univariate mediation analyses was actually caused by students' prior knowledge. Thus, the investigation of the simultaneous mediation of effects socio-economic status by students' topic-specific prior knowledge and self-concept of ability in science was also guided by two aims. First, it was examined if students' self-concept of ability had a significant impact on post-instructional science achievement beyond the effects of prior knowledge. In other words, the viability of processes of skill enhancement was tested for data set at hand. Second, it was inquired if self-concept of ability in science mediated a part of the effect of family socio-economic status on science achievement in addition to students' topic-specific prior knowledge. Due to a lack of theoretically sound assumptions about direct relations, and thus directly shared common variance, between fluid ability and self-concept of ability in science, multilevel models with these two propensities as a pair of concurrent mediators for the relation between socio-economic status and science achievement were not estimated.

Finally, to summarize the results of all previous analyses, an investigation of the comprehensive mediation of the effect of socio-economic status on science achievement by cognitive and motivational propensities concluded the empirical explorations. Naturally, these final explorations took the results of the preceding analyses into account. (The same held true for the previous examination of the pairs of concurrent mediators.) In particular, students' individual interest in science was not incorporated as a potential mediator as it had not been identified as a relevant mediator in the univariate mediation analyses. Moreover, with regard to the second layer of the structure of research questions, only cross-level interactions between mediators and school type were inquired in the multivariate mediation analyses. This was done because in the univariate mediation analyses cross-level interactions between mediators and grade level had not been found to be of importance for the prediction of post-instructional science achievement.

Concurrent Mediation by Fluid Ability and Topic-specific Prior Knowledge

Students' fluid ability and topic-specific prior knowledge constituted the two cognitive propensities examined as potential mediators in the previous univariate mediation analyses. Their exploration as concurrent mediators of the association between family socio-economic status and post-instructional science achievement featured the estimation of four multilevel models. In particular, this included the estimation of a main-effects-only model with socio-economic status, fluid ability and prior knowledge as predictors of science achievement on the individual level as well as with school type as a predictor of science achievement on the class level (Model 10A). Moreover, three cross-level interaction models extending the basic main-effects-only model in specific ways were devised: a model containing a cross-level interaction between fluid ability and school type (Model 10B), a model incorporating a cross-level interaction between prior knowledge and school type (Model 10C) and, eventually, a model covering both kinds of cross-level interactions simultaneously (Model 10D). The previous univariate mediation analyses had yielded evidence in favor of the relevance of both the cross-level interaction between fluid ability and school type and the cross-level interaction between prior knowledge and school type.

Table 29

Random Coefficient Models for Socio-economic Status, Fluid Ability and Prior Knowledge as Predictors of Student Achievement with School Type as a Class-level Covariate

Parameter	Model 10A		Model 10B		Model 10C		Model 10D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
CFT	0.11***	0.02			0.11***	0.02		
SAT1	0.51***	0.02	0.51***	0.02				
Residual Variance	0.50***	0.02	0.50***	0.02	0.49***	0.02	0.48***	0.02
Class Level								
Random Intercept	0.15***	0.04	0.16***	0.04	0.15***	0.04	0.16***	0.04
Random Slope CFT			0.13***	0.03			0.12***	0.03
Random Slope SAT1					0.53***	0.03	0.53***	0.03
CFT*Hauptschule[a]			-0.05	0.05			-0.03	0.05
CFT*Gymnasium[a]			-0.02	0.04			-0.02	0.05
SAT1*Hauptschule[a]					-0.09	0.06	-0.08	0.06
SAT1*Gymnasium[a]					-0.02	0.05	-0.01	0.05
Hauptschule[a]	-0.52***	0.07	-0.53***	0.07	-0.55***	0.08	-0.55***	0.08
Gymnasium[a]	-0.13†	0.07	-0.13†	0.07	-0.14†	0.07	-0.13†	0.07
Residual Variance	0.07***	0.01	0.07***	0.01	0.07***	0.01	0.07***	0.01
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	5925.9	33.40	5927.1	33.43	**5905.4**	31.21	5909.2	31.40
BIC	5973.1	33.40	5991.9	33.43	**5970.2**	31.21	5991.7	31.40
aBIC	5947.6	33.40	5957.0	33.43	**5935.3**	31.21	5947.3	31.40

Note. Est. = Estimate; SES = Socio-economic Status; CFT = Fluid Ability (assessed with CFT 20-R); SAT1 = Pre-instructional Student Achievement Test.
[a]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

In all four multilevel models students' fluid ability affected post-instructional science achievement in a statistically significant manner, either in terms of a main effect on the individual level, $B = 0.11$, $SE = 0.02$, $p < .001$ for both models, or in terms of the mean of a random slope on the class level, $B = 0.13$, $SE = 0.02$, $p < .001$ and $B = 0.12$, $SE = 0.02$, $p < .001$, respectively (see Table 29). The same held true for students' topic specific prior knowledge. Both when modeled as a main effect on the individual level, $B = 0.51$, $SE = 0.02$, $p < .001$ for both models, and when modeled as a random slope on the class level, $B = 0.53$, $SE = 0.03$, $p < .001$ for both models, it exerted a statistically significant influence on post-instructional science achievement. Family socio-economic status, however, had no statistically significant effect on science achievement, in none of the multilevel models. Similarly, no evidence in favor of a substantial cross-level interaction between the effect of students' fluid ability and school type was found. Though the regression weights for the cross-level interaction between prior knowledge and school type, B = -0.09, SE = 0.06, p = .10 and B = -0.08, SE = 0.06, p = .15, respectively, for the school type of Hauptschule, and B = -0.02, SE = 0.05, p = .71 and B = -0.01, SE = 0.05, p = .80, respectively, for the school type of Gymnasium, failed to attain statistical significance, all three information criteria calculated for the multilevel models unanimously detected the model with exclusively the cross-level interaction between prior knowledge and school type (Model 10C) as possessing the best relative fit to the data.(Note, in this context, that the Wald test used for determining the statistical significance of individual regression weights is only asymptotically correct.) In other words, despite the missing statistical significance of regression weights, the present investigations indicated that for students from the school type of Hauptschule the effect of prior knowledge on science achievement was somewhat smaller than for students from the other school types. Finally, the regression weights for the main effect of school type on the class level revealed that students from the school type of Hauptschule achieved on the average significantly worse than students from the school type of Grundschule, B = -0.52, SE = 0.07, p < .001, B = -0.53, SE = 0.07, p < .001, B = -0.55, SE = 0.08, p < .001 and B = -0.55, SE = 0.08, p < .001, respectively, while there was also a marginally significant tendency for students from the school type of Gymnasium to perform worse than students from the school type of Grundschule, B = -0.13, SE = 0.07, p = .08, B = -0.13, SE = 0.07, p = .08, B = -0.14, SE = 0.07, p = .07 and B = -0.13, SE = 0.07, p = .07, respectively. So, notably, under control of students' fluid ability and topic-specific prior knowledge, students from the school type of Gymnasium did not outperform students from the school type of Grundschule in the science achievement test.

The incorporation of fluid ability and prior knowledge as concurrent predictors of science achievement in joint models revealed that students' fluid ability exerted a statistically significant influence on post-instructional science achievement beyond topic-specific prior knowledge (Models 10A–10D). However, in comparison to the results obtained in the corresponding univariate mediation analysis (Models 3B & 3D), the strength of this influence was cut approximately in half. In contrast, the magnitude of the effect of prior knowledge on science achievement found in the current multivariate mediation analysis (Models 10A–10D) did not differ considerably from the size of the impact of prior knowledge on science achievement retrieved in the corresponding univariate mediation analysis (Models 5B & 5D). However, the regression weights for family socio-economic status obtained in the present examination of fluid ability and prior knowledge as concurrent predictors (Models 10A–10D) did not fall below the corresponding regressions weights estimated in the univariate mediation analyses for prior knowledge (Models 5B & 5D). In other words, the additional inclusion of students' fluid ability as a predictor of science achievement beyond topic-specific prior knowledge did not entail the mediation of an additional portion of the effects of family socio-economic status on post-instructional science achievement. With respect to the second layer of the structure of research questions, the comparison of the information criteria retrieved for the models of the multivariate mediation by fluid ability and prior knowledge underscored – in spite of the absence of a statistically significant cross-level interaction – the moderation of the effect of prior knowledge on post-instructional science achievement by school type, the impact of topic-specific prior knowledge being comparatively less important for students from the school type of Hauptschule. The relevance of a moderation of the effect of fluid ability on science achievement by school type that had been detected in the previous univariate mediation analyses was not corroborated in the current multivariate mediation analyses.

Concurrent Mediation by Prior Knowledge and Self-concept of Ability

Self-concept of ability and student achievement having been identified in previous research to be reciprocally related to each other, the cognitive propensity of topic-specific prior knowledge and the motivational propensity of self-concept of ability in science were examined as concurrent mediators of the association between family socio-economic status and post-instructional science achievement. The sequence of four multilevel models estimated for this purpose mirrored the set of multilevel models computed for inspection of the concurrent mediation by fluid ability and topic-specific prior knowledge. First, a main-effects-only model with socio-economic status, prior knowledge and self-concept of ability as predictors of science achievement on the individual level as well as with school type as a predictor of achievement on the class level was estimated (Model 11A). Second, this basic model was modified by including a cross-level interaction between prior knowledge and school type (Model 11B). Third, the basic model was extended by incorporating a cross-level interaction between self-concept of ability in science and school type (Model 11C). Fourth, and eventually, the basic model was supplemented simultaneously with both forms of cross-level interactions (Model 11D). The preceding explorations had underlined the importance of the cross-level interaction between topic-specific prior knowledge and school type for overall model fit of estimated multilevel models. On the contrary, there had been no evidence in favor of a substantial cross-level interaction between self-concept of ability in science and school type in the prediction of post-instructional science achievement.

Table 30
Random Coefficient Models for Socio-economic Status, Prior Knowledge and Self-concept of Ability as Predictors of Student Achievement with School Type as a Class-level Covariate

Parameter	Model 11A		Model 11B		Model 11C		Model 11D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Individual Level								
SES	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.02
SAT1	0.52***	0.02			0.52***	0.02		
SCA	0.09***	0.02	0.09***	0.02				
Residual Variance	0.50***	0.02	0.49***	0.02	0.50***	0.02	0.48***	0.02
Class Level								
Random Intercept	0.13**	0.04	0.13**	0.04	0.13**	0.04	0.13**	0.04
Random Slope SAT1			0.54***	0.03			0.54***	0.03
Random Slope SCA					0.09***	0.02	0.08**	0.02
SAT1*Hauptschule[a]			-0.09	0.05			-0.09†	0.06
SAT1*Gymnasium[a]			-0.02	0.05			-0.02	0.05
SCA*Hauptschule[a]					-0.00	0.04	0.02	0.04
SCA*Gymnasium[a]					0.03	0.04	0.04	0.04
Hauptschule[a]	-0.51***	0.08	-0.54***	0.08	-0.51***	0.08	-0.54***	0.08
Gymnasium[a]	-0.04	0.08	-0.05	0.08	-0.03	0.08	-0.04	0.08
Residual Variance	0.08***	0.01	0.08***	0.01	0.08***	0.01	0.08***	0.01
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	5931.5	33.24	**5913.3**	30.81	5936.0	33.16	5917.5	30.86
BIC	5978.6	33.24	**5978.1**	30.81	6000.8	33.16	6000.3	30.86
aBIC	5953.2	33.24	**5943.2**	30.81	5965.9	33.16	5955.55	30.86

Note. Est. = Estimate; SES = Socio-economic Status; SAT1 = Pre-instructional Student Achievement Test; SCA = Self-concept of Ability.
[a]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

As in all preceding analyses, students' topic-specific prior knowledge had a pronounced general effect on post-instructional science achievement: This was true for those multilevel models containing a main effect of prior knowledge on the individual level, B = 0.52, SE = 0.02, $p < .001$ for both instances, as well as for those models featuring a random slope on the class level for the effect of prior knowledge, B = 0.54, SE = 0.03, $p < .001$ for both instances (see Table 30). The regression weights for self-concept of ability on the individual level, B = 0.09, SE = 0.02, $p < .001$ for both instances, as well as the means of the random slopes for self-concept of ability on the class level, $B = 0.09$, $SE = 0.02$, $p < .001$ and $B = 0.08$, $SE = 0.02$, $p < .01$, respectively, indicated that self-concept of ability in science exerted a statistically significant influence in terms of a main effect on post-instructional science achievement. So, self-concept of ability in science contributed in its own right significantly to the prediction of post-instructional science achievement. Specifically, for students of comparable prior knowledge, an increment of one standard deviation in self-concept of ability in science was accompanied by an increase of almost one tenth of a standard deviation in post-instructional science achievement. Thus, as a by-product, the exploration of the concurrent mediation of effects of socio-economic status on post-instructional science achievement by prior knowledge and self-concept of ability in science underscored the relevance of processes of skill enhancement for the student sample and topic under consideration. No evidence in favor of a substantial influence of family socio-economic status on science achievement under simultaneous control of prior knowledge and self-concept of ability was found.

As for the investigation of the concurrent mediation by fluid ability and prior knowledge, all three information criteria pointed to the relevance of the cross-level interaction between topic-specific prior knowledge and school type, prior knowledge being comparatively less important for the prediction of post-instructional science achievement in students from the school type of Hauptschule. Moreover, in agreement with the results of the univariate mediation analyses, the present concurrent mediation analysis yielded no evidence in favor of a statistically significant moderation of the effects of self-concept of ability in science by school type in the prediction of science achievement. For comprehensive interpretation of results, please recall, however, that a substantial cross-level interaction between socio-economic status and grade level had been identified for the social gradient in self-concept of ability in science: Whereas for fourth-graders family socio-economic status was associated significantly with students' self-concept of ability in science, for sixth-graders such a statistically significant relation had not been established.

The inclusion of prior knowledge and self-concept of ability in science as concurrent predictors of post-instructional science achievement in a multivariate mediation analysis established that self-concept of ability exerted a statistically significant influence on science achievement independently of prior knowledge (Models 11A–11D). Nevertheless, the size of the effect of self-concept of ability on science achievement was more than halved when compared to the estimated strength of this association in the corresponding univariate mediation analysis (Models 7B & 7D). In contrast, the sizes of the regression weights and random slopes for the main effect of topic-specific prior knowledge on post-instructional science achievement in the present multivariate mediation analysis (Models 11A–11D) and in the corresponding univariate mediation analysis (Models 5B & 5D) were virtually identical. Furthermore, there were signs that self-concept of ability in science also contributed in its own right to the mediation of effects of socio-economic status: Relative to the state of affairs observed for the estimation of the social gradient in post-instructional science achievement (Models 1A–1D), those investigations adequately modeling the cross-level interaction between prior knowledge and school type entailed a drop of the effect of socio-economic status on science achievement to $B = 0.01$, $SE = 0.02$, $p = .48$ and $B = 0.01$, $SE = 0.02$, $p = .44$ (Models 11B & 11D). This drop exceeded the reduction in the effect of socio-economic status on science achievement due to the exclusive incorporation of the mediator of prior knowledge by a value of 0.01 (Models 5B & 5D). In other words, in addition to the major portion mediated by topic-specific prior knowledge, about one tenth of the original effect of family socio-economic status on science achievement – in absolute terms equaling approximately one percent of a standard deviation in post-instructional science achievement – was mediated specifically by students' self-concept of ability in science.

Comprehensive Mediation by Cognitive and Motivational Propensities

To conclude the investigations, four multilevel path models portraying the comprehensive mediation of effects of socio-economic status on post-instructional science achievement by cognitive and motivational propensities were estimated. The specification of these four models took results of previous analyses into account. As it had not been identified as a mediator in the univariate mediation analyses, students' individual interest in science was not included as a mediator in the four multilevel path models. The four multilevel path models constituted a sequence of models of increasing complexity. The first of these models (Model 12A) covered only main effects, i.e. main effects of family socio-economic status on the mediators of fluid ability, topic-specific prior knowledge and self-concept of ability in science as well as main effects of socio-economic status, fluid ability, topic-specific prior knowledge and self-concept of ability in science on the dependent variable of post-instructional science achievement. Moreover, for all these relations the multilevel structure of the data was taken into consideration and main effects of dummy-coded variables on the class level modeled mean differences between the school types of Hauptschule and Gymnasium on the one hand and the reference category of the school type of Grundschule on the other hand. This basic model was extended in a stepwise fashion by incorporation of an additional cross-level interaction between socio-economic status and school type in the prediction of students' self-concept of ability in science (Model 12B), an additional cross-level interaction between topic-specific prior knowledge and school type (Model 12C) and an additional cross-level interaction between fluid ability and school type (Model 12D). Path models containing grade level as a covariate on the class level were not formulated. Necessarily, the multilevel path models contained the mediators both as dependent variables predicted by socio-economic status and as independent variables for the prediction of post-instructional science achievement. Insofar, the multilevel path models represented a combination of previous models delineating the social gradients in the mediators as well as of previous models investigating the impact of socio-economic status and the mediators on post-instructional science achievement.

Replicating the findings of previous analyses with respect to the second layer of the structure of research questions, a comparison of the information criteria calculated for the four multilevel path models rendered the model containing a term for the cross-level interaction between socio-economic status and school type in the prediction of self-concept of ability and a term for the cross-level interaction between prior knowledge and school type in the prediction of science achievement (Model 12C) as possessing the relative best fit to the data, thereby underlining the general relevance of these two cross-level interactions. In particular, this final path model contained statistically significant social gradients in fluid ability $B = 0.05$, $SE = 0.02$, $p = .02$, and topic-specific prior knowledge, $B = 0.12$, $SE = 0.02$, $p < .001$. The social gradient in self-concept of ability in science was crucially moderated by the class-level covariate of school type. As indicated by the mean of the random slope for the effect of socio-economic status on self-concept of ability in science, it was statistically significant, $B = 0.14$, $SE = 0.03$, $p < .001$, for the reference category of students from the school type of Grundschule. Denoting the difference to the size of the effect within the reference category, the regression weights for the cross-level interaction between socio-economic status and school type, $B = -0.09$, $SE = 0.07$, $p = .16$ for the school type of Hauptschule and $B = -0.17$, $SE = 0.06$, $p < .01$ for the school type of Gymnasium, hinted at the absence of a relevant social gradient in self-concept of ability in science for the school types of secondary education. (Recall that in the univariate mediation analysis for self-concept of ability in science a model with a cross-level interaction between socio-economic status and grade level had been identified as most parsimonious in the prediction of self-concept of ability. For the current investigations school type was selected as the moderator of the relation between socio-economic status and self-concept of ability merely for the sake of uniformity of the class-level covariate used in the comprehensive mediation analysis.)

With regard to the prediction of post-instructional science achievement, both students' fluid ability, $B = 0.09$, $SE = 0.02$, $p < .001$, and students' self-concept of ability in science, B

Table 31

Multilevel Path Models for the Comprehensive Mediation of Effects of Socio-economic Status on Student Achievement with School Type as a Class-level Covariate

Parameter	Model 12A		Model 12B		Model 12C		Model 12D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Dependent Variable: CFT								
Individual Level								
SES	0.05*	0.02	0.05*	0.02	0.05*	0.02	0.05*	0.02
Residual Variance	0.79***	0.02	0.79***	0.02	0.79***	0.02	0.79***	0.02
Class Level								
Random Intercept	-0.16***	0.04	-0.16***	0.04	-0.16***	0.04	-0.16***	0.04
Hauptschule[a]	-0.18*	0.08	-0.18*	0.08	-0.18*	0.08	-0.18*	0.08
Gymnasium[a]	0.70***	0.08	0.70***	0.08	0.70***	0.08	0.70***	0.08
Residual Variance	0.07***	0.01	0.07***	0.01	0.07***	0.01	0.07***	0.01
Dependent Variable: SAT1								
Individual Level								
SES	0.12***	0.02	0.12***	0.02	0.12***	0.02	0.12***	0.02
Residual Variance	0.72***	0.02	0.72***	0.02	0.72***	0.02	0.72***	0.02
Class Level								
Random Intercept	-0.11*	0.05	-0.11*	0.05	-0.11*	0.05	-0.11*	0.05
Hauptschule[a]	-0.31**	0.09	-0.31**	0.09	-0.31**	0.09	-0.31**	0.09
Gymnasium[a]	0.62***	0.09	0.62***	0.09	0.62***	0.09	0.62***	0.09
Residual Variance	0.11***	0.02	0.11***	0.02	0.11***	0.02	0.11***	0.02
Dependent Variable: SCA								
Individual Level								
SES	0.08**	0.02						
Residual Variance	0.85***	0.02	0.85***	0.03	0.85***	0.03	0.85***	0.03
Class Level								
Random Intercept	0.12*	0.05	0.12*	0.05	0.12*	0.05	0.12*	0.05
Random Slope SES			0.14***	0.03	0.14***	0.03	0.14***	0.03
SES*Hauptschule[a]			-0.09	0.07	-0.09	0.07	-0.09	0.07
SES*Gymnasium[a]			-0.17**	0.06	-0.17**	0.06	-0.17**	0.06
Hauptschule[a]	-0.28**	0.10	-0.29**	0.10	-0.29**	0.10	-0.29**	0.10
Gymnasium[a]	-0.24*	0.09	-0.21*	0.09	-0.21*	0.09	-0.21*	0.09
Residual Variance	0.12***	0.02	0.12***	0.02	0.12***	0.02	0.12***	0.02
Dependent Variable: SAT2								
Individual Level								
SES	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02
CFT	0.09***	0.02	0.09***	0.02	0.09***	0.02		
SAT1	0.48***	0.02	0.48***	0.02				
SCA	0.09***	0.02	0.09***	0.02	0.09***	0.02	0.09***	0.02
Residual Variance	0.49***	0.02	0.49***	0.02	0.48***	0.02	0.48***	0.02

Table 31 (continued)
Multilevel Path Models for the Comprehensive Mediation of Effects of Socio-economic Status on Student Achievement with School Type as a Class-level Covariate

Parameter	Model 12A		Model 12B		Model 12C		Model 12D	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
	Dependent Variable: SAT2 (continued)							
Class Level								
Random Intercept	0.08	0.05	0.08	0.05	0.13**	0.04	0.15***	0.04
Random Slope CFT							0.11***	0.03
Random Slope SAT1					0.50***	0.03	0.50***	0.03
CFT*Hauptschule[a]							-0.03	0.05
CFT*Gymnasium[a]							-0.01	0.04
SAT1*Hauptschule[a]					-0.08	0.05	-0.08	0.06
SAT1*Gymnasium[a]					-0.02	0.05	-0.02	0.05
Hauptschule[a]	-0.69***	0.10	-0.69***	0.10	-0.58***	0.08	-0.56***	0.08
Gymnasium[a]	0.27**	0.10	0.27**	0.10	-0.04	0.08	-0.11	0.08
Residual Variance	0.14***	0.02	0.14***	0.02	0.08***	0.01	0.07***	0.01
	Information Criteria							
	M	SD	M	SD	M	SD	M	SD
AIC	27257.2	44.47	27249.5	43.95	**27230.3**	41.55	27234.6	41.72
BIC	27451.7	44.47	27461.7	43.95	**27460.2**	41.55	27482.1	41.72
aBIC	27346.9	44.47	27347.3	43.95	**27336.2**	41.55	27348.7	41.72

Note. Est. = Estimate; SES = Socio-economic Status; CFT = Fluid Ability (assessed with CFT 20-R); SAT1 = Pre-instructional Student Achievement Test; SCA = Self-concept of Ability; SAT2 = Post-instructional Student Achievement Test.
[a]Reference category is Grundschule (i.e. fourth grade).
Lowest values for information criteria are printed in bold.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

= 0.09, SE = 0.02, $p < .001$, contributed substantially. With a mean of the corresponding random slope of B = 0.50, SE = 0.03, $p < .001$, however, students' topic-specific prior knowledge was the most powerful predictor of post-instructional science achievement. Though, in the prediction of science achievement, the regression weights for the cross-level interaction between prior knowledge and school type, B = -0.08, SE = 0.05, p = .12 for the school type of Hauptschule and B = -0.02, SE = 0.05, p = .70 for the school type of Gymnasium, failed to attain statistical significance, the low values of the information criteria as well as the reduced residual variance on the class level for the most parsimonious path model signaled the overall relevance of this cross-level interaction. In the final path model the effect of family socio-economic status on science achievement was statistically insignificant, B = 0.01, SE = 0.02, p = .53. In other words, conjointly the propensities of fluid ability, topic-specific prior knowledge and self-concept of ability in science mediated the entire effect of socio-economic status on science achievement. Two relations in this general pattern of results were moderated by school type. Self-concept of ability in science contributed primarily for students from the school type of Grundschule to the mediation of effects of socio-economic status as the social gradient in self-concept of ability approached zero for the school types of secondary education. Furthermore, topic-specific prior knowledge contributed comparatively less to the mediation of effects of social background for students from the school type of Hauptschule as its effect on post-instructional science achievement tended to be somewhat attenuated within this school type.

Discussion

Central Findings

The univariate mediation analyses uncovered three cognitive and motivational propensities as mediators of the relation between family socio-economic status and post-instructional science achievement: students' fluid ability, students' topic-specific prior knowledge, and students' self-concept of ability in science. The cognitive propensity of prior knowledge entailing a full mediation, i.e. rendering the direct effect of family socio-economic status on students' post-instructional science achievement statistically insignificant, the propensities of fluid ability and self-concept of ability mediated the effect of social background only partially, i.e. they a caused a reduction of the size of the direct effect of socio-economic status on achievement. On the contrary, for the motivational propensity of individual interest in science, as it was unrelated with family socio-economic status, not the slightest sign of a mediation of the effects of social background on science achievement was obtained. The further exploration of the mediating pathways between socio-economic status and post-instructional achievement in multivariate mediation analyses revealed that fluid ability did not contribute measurably beyond students' prior knowledge to the mediation of effects of socio-economic status, whereas the motivational propensity of self-concept of ability mediated a barely quantifiable portion of the effects of social background in its own right.

With regard to the second layer of the structure of research questions, the majority of hypotheses was disconfirmed. The social gradient in post-instructional science achievement was not moderated by grade level or school type. In other words, neither was there evidence in favor of its amplification in the transition from elementary to secondary education due to a summation of primary and secondary effects of social origin nor was there evidence in favor of its diminishment within the school types of secondary education due to restrictions of variance in these school types. What explanations can be offered for the absence of the suspected moderations of the relation between family socio-economic status and student's post-instructional science achievement by grade level or school type? Presumably, the focus on the topic of evaporation and condensation as well as the proximity of the subsamples of fourth- and sixth-graders to the transition from elementary to secondary education hindered the occurrence of a substantial summation of primary and secondary effects of social origin in the current investigations. Moreover, relevant restrictions of variance in post-instructional science achievement within the school types of secondary education relative to the corresponding variance in the comprehensive school type of Grundschule were not observed; in fact, the variance of post-instructional science achievement in the selective school type of Gymnasium exceeded the corresponding variance in the comprehensive school type of Grundschule.

Likewise, the role of students' fluid ability as a mediator of the association between socio-economic status and science achievement was not moderated by grade level or school type. In particular, contrary to the hypotheses, the social gradient in fluid ability did not increase with students' age, i.e. grade level. In other words, there was no evidence in favor of an accumulation of adverse effects of social background on students' fluid ability. Here, the comparably small difference in students' mean age between the subsamples of fourth- and sixth-graders applies as a plausible explanation for the disconfirmation of hypotheses located at the second layer of structure of research questions. Moreover, though the univariate mediation analyses suggested substantially disparate strengths of the relation between fluid ability and post-instructional achievement for the school types of Hauptschule and Gymnasium, the multivariate mediation analyses yielded no evidence in support of a substantial moderation by grade level or school type. This finding contradicted the assumed relative shift from fluid ability to prior knowledge as the dominant predictor of post-instructional achievement with students' age. The topic of evaporation and condensation constitutes a component of the curricula of both elementary and secondary education in the German federal state of North Rhine-Westphalia (Gesellschaft für Didaktik des Sachunterrichts, 2002; Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008a, 2008b, 2011). Nevertheless, elementary school teachers might have a tendency to avoid the provision of instruction on topics of the hard natural sciences (Appleton, 2007). Thus, it might have been suspected that the cumulative acquisition of knowledge with respect to the topic of evaporation and condensation, which represents a necessary prerequisite for the assumed relative shift from fluid ability to prior knowledge as the central predictor of post-instructional achievement, had not occurred in full effect for the sample of the current study. A quick inspection, however, of students' average prior knowledge and post-instructional achievement in the different grade levels and school types under investigation denied this suspicion.

In resemblance to the findings for the social gradient in post-instructional science achievement, the current investigations did not uncover a substantial moderation of the relation between family socio-economic status and topic-specific prior knowledge by grade level or school type. Here, too, no evidence in favor of a summation of primary and secondary effects of social origin in the comparison of fourth- and sixth-graders was found. Again, the restriction of the current analyses to one particular topic, evaporation and condensation, and the proximity of the students subsamples to the actual transition from elementary to secondary education appear as plausible explanations for the absence of substantial moderations.

With regard to the mediation by self-concept of ability in science, a general decoupling of the associations of students' self-concept of ability in science with both socio-economic status and post-instructional science achievement was expected due to the assumed impact of the big-fish-little-pond effect in the transition to secondary education. However, the corresponding analyses did not reveal any substantial moderation of the association between students' self-concept of ability in science and post-instructional achievement. In other words, though the students of the current study displayed a typical decline of self-concept of ability in science in the transition from elementary to secondary education, no evidence in favor of an accompanying diminishment of the relevance of self-concept of ability for the prediction of science achievement was uncovered. As noted repeatedly before, the student subsamples of fourth- and sixth-graders were located relatively closely to the actual transition from elementary to secondary education.

Finally, with regard to the mediation by individual interest in science, a general amplification of the relations of students' individual interest in science with both family socio-economic status and post-instructional science achievement was hypothesized prior to the analyses. The analyses revealed no substantial moderations of the two aforementioned relations. Particularly, no empirical support for an intensification of the association between social background and individual interest in science due to an assumed accumulation of adverse effects of socio-economic status on the formation of academic interest in the course of formal schooling was found. Likewise, the analyses yielded no evidence in favor of a general amplification of the relevance of individual interest in science as a predictor of post-instructional science achievement in the transition from elementary to secondary education. Notably, this absence of systematic moderations for mediational paths involving the motivational propensity of individual interest was embedded in a background of a typical deterioration of students' individual interest in science.

Nonetheless, some hypotheses with regard to the second layer of the structure of research questions were not, or at least not totally, disconfirmed. The impact of prior knowledge on post-instructional science achievement not intensifying generally in the transition from elementary to secondary education, the importance of prior knowledge as a precursor of achievement tended to be smaller for students from the school type of Hauptschule than for students from the school type of Grundschule. As the relevance of prior knowledge as a predictor of science achievement was suspected to increase with students' level of science competence, this observation concurs with the fact that students from the school type of Hauptschule displayed the lowest average prior knowledge and post-instructional achievement of all student groups defined by school type. However, although students from the school type of Gymnasium possessed the relatively highest average topic-specific prior knowledge, the relation between prior knowledge and post-instructional science achievement was not found to amplify within the school type of Gymnasium. Clearly, this finding does not conform to the conception that the relevance of prior knowledge as a predictor of academic achievement increases with students' level of competence and the complexity of the content to be learned. Indeed, at first glance, it invites the speculation that the science achievement test did not cover the entire breadth of the content taught in the classrooms of the school type of Gymnasium. Nevertheless, somewhat contrary to this speculation, there was not the slightest sign of a ceiling effect for the post-instructional performance of students from the school type of Gymnasium in the science achievement test.

Last but not least, a rather comprehensive pattern of empirical support was obtained for the hypothesized moderations of the social gradient in students' self-concept of ability in science. In agreement with a general decoupling of the association between socio-economic status and self-concept of ability in science in the transition from elementary to secondary education, an effect of social background on students' self-concept of ability was disclosed for the subsample of fourth-graders but not for the subsample of sixth-graders. Furthermore, there was a statistically non-significant tendency for this dwindling of the relation between socio-economic status and self-concept of ability in science to be more pronounced within the school type of Gymnasium than within the school type of Hauptschule. This accords with the additional assumption that the school type of Gymnasium represents the least ego-protective learning environment of the school types included in the current study and thus is the most likely of these school types to hamper top-down effects from general self-concept, which in turn is associated presumably with socio-economic status, on domain-specific self-concept of ability in science.

Strengths and Limitations of the Study

The numerous large-scale assessments of student achievement in Germany of the past decade have yielded a detailed picture of social disparities in achievement in various academic domains (e.g., Ehmke & Baumert, 2007; Ehmke & Jude, 2010; Maaz et al., 2009; Walter, 2008). Researchers, policy makers and the public have been surprised by the actual size of these disparities (Geißler, 2004). The present study complements and extends previous research by applying research questions generated mainly in the context of the evaluation of educational systems to an investigation of actual instruction. Moreover, it expands previous research by including psychological propensities as potential mediators of the relation between socio-economic status and student achievement for students before and after the critical transition from elementary to secondary education, with the topic of instruction under investigation held constant across grade levels and school types (see Baumert et al., 2003, for an exploration of the mediating role of psychological propensities for reading competence in 15-year-olds). In addition, two methodological characteristics of the present study deserve specific recognition. First, the relevance of the school type of Hauptschule and the proportion of the student population enrolled in the school type of Hauptschule varies between the federal states of Germany (Ditton, 2013). In fact, some federal states have abolished the school type of Hauptschule or are currently undergoing such a process of abolishment. In the federal state of North Rhine-Westphalia, in which the current study had been conducted, the school type of Hauptschule only serves a minority, i.e. about 15%, of the student population, a minority of comparatively low-achieving students (Autorengruppe Bildungsbericht, 2010). Thus, the oversampling of students from the school type of Hauptschule in the current study improved the precision of the estimation of effects for students from the school type of Hauptschule as compared to the use of a representative sample of students from secondary education. Second, in studies of social gradients in academic achievement, the occurrence of high proportions of missing data on socio-economic status for socially underprivileged families poses a serious threat to the validity of results, especially when parental reports are used to operationalize students' social background. The current study dealt with missing data by means of multiple imputation with a sophisticated imputation model that included variables from external sources and preserved the multilevel structure of the data.

The choice of a measurement model for a test or a scale is not an unambiguous affair, yet it may impact the conclusions about the substantial research questions under consideration (Robitzsch, Dörfler, Pfost, & Artelt, 2011; Rupp & Zumbo, 2003). For the calibration of the tests and scales of the current study, a straightforward decision in favor of the use of the simple Rasch model and its variants was made. This approach ensured a clear mapping of item content to the measured constructs, i.e. the items of a given test or scale contributed with equal weight to estimates of person ability or person attitude strength; in case of two-parameter item response models the contribution of individual items to the estimation of person ability or person attitude strength is determined empirically by the sample under consideration. For the current study, the concurrent calibration of the test and scales was deemed justifiable as all participating students completed the same tests and scales. Within the framework of the Rasch family of item response models, thorough test and scale analyses were performed. By and large, the corresponding retrieval of estimates of person ability and person attitude strength appeared acceptable.

Naturally, with respect to the second layer of the structure of research questions, the present study suffered from the conventional limitations of a cross-sectional field study; it confounded age and cohort effects (e.g., Hua & David, 2008; Schmiedek & Lindenberger, 2007; Zelinski, Kennison, Watts, & Lewis, 2009). However, as the subsamples of fourth- and sixth-graders were separated merely by approximately two years in terms of students' mean age, it appears very unlikely that a confound of age and cohort effects substantially compromised the results of the present investigations. (Obviously, the plausibility of the occurrence of confounded effects of age and cohort increases with the age range covered in a cross-sectional study.) Furthermore, to avoid precipitate hail of longitudinal investigations, note that longitudinal studies also bear limitations, in the sense that they confound age with epoch effects; a limitation not displayed by cross-sectional studies. Apart from that, with regard to the first layer of the structure of research questions, the present analyses relied on a, at least, partially sequential design (cf. Cole & Maxwell, 2003; Mitchell & Maxwell, 2013): The measurement of the potential mediators preceded the assessment of the final criterion, students' post-instructional science achievement. Though family socio-economic status as the primal predictor was assessed in temporal vicinity to the measurement of the mediators, it was based on parental reports, not on students' accounts. In addition, with some justification, socio-economic status can be considered a comparatively stable characteristic of families. Thus, the schedule of data collection of the current study adequately represented the temporal structure of the causal mechanisms supposed to underlie the investigated mediation.

When students' topic-specific prior knowledge was statistically controlled, there was no difference in mean post-instructional science achievement between students from the school type of Grundschule and students from the school type of Gymnasium (see e.g., Model 5D). In other words, sixth-graders with an average age of 12.16 years from the selective school type of Gymnasium did not outstrip fourth-graders with an average of 10.27 years from the comprehensive school type of Grundschule in terms of their learning gains over the course of the series of lessons on the topic of evaporation and condensation. (Recall that the same test instrument was used for the assessment of prior knowledge and for the assessment of post-instructional science achievement.) This counterintuitive observation may raise doubts concerning the validity of the use of the science achievement test for students from the school type of Gymnasium, in particular, and for students from secondary schools, in general. However, thoughtful consideration of the content of the science achievement test is apt to clarify the substantiality of these doubts. A large portion of the items of the science achievement test requested students to endorse correct explanations for phenomena of evaporation and condensation as well as to reject associated misconceptions. Another significant portion of the items called for the knowledge of simple facts about evaporation and condensation. Only few items directly assessed knowledge related to the particle model of matter. Thus, in fact, it appears possible that the science achievement test did not capture the learning progress induced by instruction in the sixth grade of the school type of Gymnasium in its entirety. It is plausible that greater learning gains had occurred for students from the school type of Gymnasium than for students from the school type of Grundschule if instruction in the school type of Gymnasium had really concentrated on explanations for phenomena of evaporation and condensation and associated misconceptions. Nevertheless, there were no floor or ceiling effects in the assessments with the science achievement test, neither for topic-specific prior knowledge nor for post-instructional science achievement.

Students' topic-specific prior knowledge and students' post-instructional science achievement were assessed by the repeated use of the same test instrument. Thus, here students' responses presumably shared variance beyond variance attributable to students' knowledge of the topic of evaporation and condensation, variance pertinent to other features common to the two assessments, such as, for instance, the response format of items. Moreover, unavoidably, the scope of the two assessments was identical. On both occasions, topic-specific knowledge concerning evaporation and condensation was assessed. In contrast, students' motivational propensities were measured in a domain-specific frame, as self-concept of ability in science and as individual interest in science. By definition, fluid ability was assessed as a domain-general capability. These differences in scope possibly restricted the potential overlap in students' measures on prior knowledge and post-instructional achievement, on the one hand, and students' measures on the motivational propensities and fluid ability, on the other hand. So, in sum, the dominance of prior knowledge as mediator of the relation between socio-economic status and science achievement obtained in the present investigations might be considered an upper bound of the true mediating role of prior knowledge. Presumably, the use of distinct tests for the assessment of prior knowledge and post-instructional achievement would have resulted in a more conservative appraisal of the role of prior knowledge as a mediator of the relation between socio-economic status and science achievement. Correspondingly, the use of topic-specific measures for assessment of students' motivational propensities probably would have engendered a more progressive estimation of the relevance of motivational propensities as mediators. Of course, given the clear dominance of prior knowledge as the central predictor of science achievement and as the crucial mediator of the relation between socio-economic status and achievement, such variations in the scope of instruments would not have overthrown the general pattern of associations between relevant variables obtained in the present analyses.

In a very general perspective, to fully understand the contribution and relation of the results of a study to the empirical knowledge base of a field, it is necessary to recognize what the implicit and explicit null hypotheses of the analyses of that study were: Is the null hypothesis that instruments function comparably for different student groups, and is thus strong evidence needed to establish differential functioning of instrument across student groups? Or, is differential functioning of instrument across student groups assumed beforehand, and is thus strong evidence needed to demonstrate actual uniformity of functioning of instruments across different student groups? Is the null hypothesis that critical associations between socio-economic status, cognitive and motivational propensities and science achievement are principally the same across grade levels and school types, and do we thus need strong evidence for the actual presence of differences in these associations between grade levels and school types? Or, are the relations among socio-economic status, individual propensities for successful learning and post-instructional achievement are assumed beforehand to differ fundamentally across grade levels and school types, and is thus strong evidence needed to establish the actual uniformity of relevant relations across grade levels and school types? Certainly, the present investigations took the approach of first assuming uniformity of functioning of instruments and uniformity of relations among critical variables across grade levels and school types and subsequently seeking strong evidence in favor of deviations from this assumed uniformity. Accordingly, the approach taken in the current analyses was conservative with respect to the detection of variations of the relations between social background, propensities for learning and academic achievement by grade level or school type.

Implications for Practitioners and Policy Makers

In the present study the major portion of the effects of family socio-economic status on science achievement was mediated by students' prior knowledge, with students' fluid ability and students' self-concept of ability in science being of minor importance as mediators. This finding represents a straightforward counterpart to the observation that social disparities in student achievement increase during summer breaks, i.e. periods without formal schooling (Alexander et al., 2001, 2007; Becker et al., 2008). So, practitioners are probably right to assume that learning processes at home and students' prior domain-specific learning history constitute the main sources for social discrepancies in post-instructional academic achievement. Moreover, as evidenced in the current investigations by a diminishment of the effect of prior knowledge on post-instructional science achievement within the school type of Hauptschule, the contribution of prior knowledge to the prediction of achievement, and thus its relevance as a mediator of effects of socio-economic status, increases with the complexity of the content that has to be acquired. However, in the present study there was no relation between socio-economic status and students' individual interest in science, and thus interest in science was not a mediator of the association between social background and post-instructional achievement. In face of findings of teachers' negative views about the prospects of students of low socio-economic status (Diamond et al., 2004; McLoyd, 1998), for practitioners this certainly is an admonition to recognize socially disadvantaged children's interest in and enthusiasm for learning as an important resource for their potential future development. Apparently, at least for the age group investigated in the present study, late childhood and early adolescence, students of low socio-economic status do not lack interest in and positive emotions towards learning science. They rather fall short from their more affluent peers in terms of fluid ability, prior knowledge and, in consequence, self-confidence in their own abilities. In other words, put as a criticism of practices of teaching, contemporary science instruction is presumably not tailored well to the specific needs of underprivileged children given their restricted cognitive abilities and prior knowledge.

The present investigations also have implications for policy makers aiming at a reduction of social disparities in academic achievement. Recall that students' prior knowledge was found to be the central mediator of the relation between family socio-economic status and post-instructional science achievement. As learning in formal settings generates prior knowledge for subsequent instruction, this finding, in conjunction with evidence in favor of the equalizing power of formal education (Alexander et al., 2001, 2007; Becker et al., 2008; Caro & Lehmann, 2009), certainly advocates preschool education programs and the extension of schooling hours as possible pathways to a diminishment of social inequalities in academic achievement (Anders, 2013; Bellin & Tamke, 2010; Expertenrat Herkunft und Bildungserfolg, 2011; Holtappels, Radisch, Rollett, & Kowoll, 2010). With respect to actual instruction in classrooms, teacher education and teacher professional development need to introduce candidate and in-service teachers to instructional strategies for addressing disparate levels of prior knowledge in order to enable them to deliver instruction that counteracts the widening or proliferation of social inequalities in academic achievement.

The aforementioned recommendations for practitioners and policy makers are formulated with a focus and positive outlook on schooling as an agent for changes in social discrepancies in academic achievement. However, as outlined in the theoretical background for the current analyses, social disparities in children's health and home environments constitute a ubiquitous phenomenon of contemporary western societies (Bradley & Corwyn, 2002; Evans, 2004; McLoyd, 1998). Apparently, theses societies find it acceptable that children's physical and psychosocial environments vary considerably with parental inheritance and occupational success. So, in face of the omnipresence of social inequalities in western societies, it is justified also to ask how reasonable it is to expect formal schooling to eradicate social disparities in academic achievement.

Directions for Future Research

Within the current state of knowledge about the role of formal schooling in the genesis of social disparities in academic achievement there appears to be discrepancy between the results and implications of quantitative studies on the one hand and the results and implications of qualitative studies on the other hand. In agreement with the findings of the present analyses, quantitative studies have tended to identify formal schooling as an equalizer of social disparities, e.g. in research on the effects of the summer break on academic achievement (Alexander et al., 2001, 2007; Becker et al., 2008; Downey et al., 2004). On the contrary, primarily qualitative studies often have found that teachers may hold adverse views of students of disadvantaged socio-economic status and may have low expectations for their academic success (Diamond et al., 2004; Helsper et al., 2009). These findings suggest that teachers and formal schooling actually increase social disparities in academic achievement (Tate, 1995; McLoyd, 1998). In this context, the impact of students' social background – beyond actual academic ability – on teachers' recommendations for the choice of school type in secondary education represents a contribution of teachers and formal schooling to the generation of social discrepancies in student achievement identified by quantitative research (Maaz &

Nagy, 2009; Wagner, Helmke, & Schrader, 2009). Surely, the reconciliation of this, at first glance, contradictory conclusions about the role of formal schooling in the emergence of social disparities in academic achievement is a worthwhile endeavor for future research. Are teachers' attitudes towards and expectations for students of low socio-economic status identified in qualitative studies merely superficial characteristics of formal schooling without any substantial consequences for the formation of social discrepancies in student achievement? Has the equalizing power of formal schooling the potential to increase when teachers' negative views about and expectations for students of low socio-economic status are alleviated?

The investigation of multilevel differential item functioning for extant data sets of large-scale assessments of student achievement can be utilized to further substantiate the claim that formal schooling acts as a great equalizer of competence (or to detect that formal schooling spurs the development of social inequalities in academic achievement): For data sets with intact classrooms the estimation of multilevel differential item functioning uncovers for which items student performance varies substantially between classrooms. Thus, multilevel differential item functioning can be consulted to distinguish items that are particularly sensitive to competence development triggered by instruction from items that measure knowledge acquisition attributable primarily to non-school processes, for instance, learning at home (Muthén, Kao, & Burstein, 1991; Robitzsch, 2009). Obviously, a positive interaction between the student characteristic of social background and the item characteristic of multilevel differential item functioning in the prediction of student performance in an explanatory item response model would indicate an actual reduction of social disparities in academic achievement by formal instruction. In correspondence to this, a negative interaction between students' socio-economic status and the multilevel differential item functioning of items in such an analysis would detect an amplification of social discrepancies by formal instruction (De Boeck & Wilson, 2004; Wilson, De Boeck, & Carstensen, 2008). Possibly, data sets of assessments of student achievement that have aimed for curricular validity are particularly apt for the investigation of this interaction between student and item characteristics (Bonsen, Lintorf, Bos, & Frey, 2008; Wendt, Tarelli, Bos, Frey, & Vennemann, 2012).

References

Achen, C. H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science, 22*, 327–339.

Achenbach, T., Bird, H., Canino G., Phares, V., Gould, M., & Rubio-Stipec, M. (1990). Epidemiological comparisons of Puerto Rican and U.S. mainland children: Parent, teacher and self reports. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 84–93.

Adams, B. N. (1998). *The family: A sociological interpretation*. New York: Harcourt Brace.

Adams, R. J., & Khoo, S. T. (1996). *Quest*. Melbourne, Australia: Australian Council for Educational Research.

Adams, R. J., Wilson, M. R., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1–23.

Adey, P. (2007). The CASE for a general factor in intelligence. In M. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 369–385). New York: Psychology Press.

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley & Sons.

Ainley, M., & Ainley, J. (2011). A cultural perspective on the structure of student interest in science. *International Journal of Science Education, 33*, 51–71.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*, 545–561.

Alexander, K., & Entwisle, D. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*(2, Serial No. 218).

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*, 171–191.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*, 167–180.

Alexander, K. L., Entwisle, D. R., & Thompson, M. (1987). School performance, status relations, and the structure of sentiment: Bringing the teacher back in. *American Sociological Review, 52*, 665–682.

Alexander, K. M., & Schofield, J. W. (2008). Understanding and mitigating stereotype threat's negative influence on immigrant and minority students' academic performance. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 48*, 529–552.

Alexander, P. A. (1992). Domain knowledge: Evolving themes and emerging concerns. *Educational Psychologist, 27*, 33–51.

Alexander, P. A. (1996). The past, present, and future of knowledge research: A re-examination of the role of knowledge in learning and instruction. *Educational Psychologist, 31*, 89–92.

Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The role of subject-matter knowledge and interest in the processing of linear and nonlinear texts. *Review of Educational Research, 64*, 201–252.

Alexander, P. A., Murphy, P. K., Woods, B. S., Duhon, K. E., & Parker, D. (1997). College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning. *Contemporary Educational Psychology, 22*, 125–146.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research, 61*, 315–343.

Allison, P. D. (2001). *Missing data* (Sage University Papers Series on Quantitative Applications in the Social Sciences, Series No. 07-136). Thousand Oaks, CA: Sage.

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology, 66*, 950–967.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261–271.

Anders, Y. (2013). Stichwort: Auswirkungen frühkindlicher institutioneller Betreuung und Bildung. *Zeitschrift für Erziehungswissenschaft, 16*, 237–275.

Andre, T., & Windschitl, M. (2003). Interest, epistemological belief, and intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 173–197). Mahwah, NJ: Erlbaum.

Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573.

Andrich, D. A. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care, 42*(Suppl. 1), 7–16.

Andrich, D. A. (2010). Understanding the response structure and process in the polytomous Rasch model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response models* (pp. 123–152). New York: Routledge.

Appleton, K. (2007). Elementary science teaching. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 493–536). Mahwah, NJ: Erlbaum.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Mahwah, NJ: Erlbaum.

Artelt, C., & Baumert, J. (2004). Zur Vergleichbarkeit von Schülerleistungen bei Leseaufgaben unterschiedlichen sprachlichen Ursprungs. *Zeitschrift für Pädagogische Psychologie, 18*, 171–185.

Asbury, K., Wachs T. D., & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence, 33*, 643–661.

Asendorpf, J. (1990). *Die differentielle Sichtweise in der Psychologie*. Göttingen: Hogrefe.

Atkinson, T., Liem, R., & Liem, J. (1986). Social costs of unemployment: Implications for social support. *Journal of Health and Social Behavior, 27*, 317–331.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.

Autorengruppe Bildungsbericht (Hrsg.). (2010). *Bildung in Deutschland 2010*. Bielefeld: Bertelsmann.

Autorengruppe Bildungsberichterstattung (Hrsg.). (2012). *Bildung in Deutschland 2012*. Bielefeld: Waxmann.

Baker, D. P., & Stevenson, D. L. (1986). Mothers' strategies for children's school achievement: Managing the transition to high school. *Sociology of Education, 59*, 156–166.

Bandalos, D. L., Yates, K., & Thorndike-Christ, T. (1995). Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. *Journal of Educational Psychology, 87*, 611–623.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bar, V., & Galili, I. (1994). Stages of children's views about evaporation. *International Journal of Science Education, 16*, 157–174.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*. Cambridge, UK: Cambridge University Press.

Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology, 92*, 568–582.

Baumert, J., & Köller, O. (1998). Interest research concerning secondary level I: An overview. In L. Hoffmann, A. Krapp, K. A. Renninger & J. Baumert (Eds.), *Interest and learning. Proceedings of the Seeon-Conference on interest and gender* (pp. 241–256). Kiel, Germany: IPN.

Baumert, J., Köller, O., & Schnabel, K. U. (2000). Schulformen als differenzielle Entwicklungsmilieus – eine ungehörige Fragestellung? In Gewerkschaft Erziehung und Wissenschaft (Hrsg.), *Messung sozialer Motivation: Eine Kontroverse* (S. 28–68). Frankfurt/Main: GEW.

Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., et al. (1997). TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske + Budrich.

Baumert, J., Lüdtke, O., Trautwein, U., & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review, 4*, 165–176.

Baumert, J., & Maaz, K. (2006). Das theoretische und methodische Konzept von PISA zur Erfassung sozialer und kultureller Ressourcen der Herkunftsfamilie: Internationale und nationale Rahmenkonzeption. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen* (S. 11–29). Wiesbaden: VS Verlag für Sozialwissenschaften.

Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools? *Child Development, 84*, 1347–1367.

Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, et al. (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Opladen: Leske + Budrich.

Baumert, J., Stanat, P., & Watermann, R. (Hrsg.). (2006a). *Herkunftsbedingte Disparitäten im Bildungswesen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Baumert, J., Stanat, P., & Watermann, R. (2006b). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen* (S. 95–188). Wiesbaden: VS Verlag für Sozialwissenschaften.

Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele et al. (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261–331). Opladen: Leske + Budrich.

Baumert, J., Watermann, R., & Schümer, G. (2003). Disparitäten der Bildungsbeteiligung und des Kompetenzerwerbs. *Zeitschrift für Erziehungswissenschaft, 6*, 46–72.

Becker, M. (2009). *Kognitive Leistungsentwicklung in differenziellen Lernumwelten: Effekte des gegliederten Sekundarschulsystems in Deutschland*. Berlin: Max-Planck-Institut für Bildungsforschung.

Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2006). Leistungszuwachs in Mathematik: Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? *Zeitschrift für Pädagogische Psychologie, 20*, 233–242.

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*, 682–699.

Becker, M., Stanat, P., Baumert, J., & Lehmann, R. (2008). Lernen ohne Schule: Differenzielle Entwicklung der Leseleistungen von Kindern mit und ohne Migrationshintergrund während der Sommerferien. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 48*, 252–276.

Benveniste, L., Carnoy, M., & Rothstein, R. (2003). *All else equal*. New York: Routledge-Farmer.

Bellin, N., & Tamke, F. (2010). Bessere Leistungen durch Teilnahme am offenen Ganztagsbetrieb? *Empirische Pädagogik, 24*, 93–112.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*, 109–160

Blumberg, E. (2008). *Multikriteriale Zielerreichung im naturwissenschaftsbezogenen Sachunterricht der Grundschule. Eine Studie zum Einfluss von Strukturierung in schülerorientierten Lehr-Lernumgebungen auf das Erreichen kognitiver, motivationaler und selbstbezogener Zielsetzungen*. Dissertation, Westfälische Wilhelms-Universität Münster.

Blumenfeld, P. C. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology, 84*, 272–281.

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling, 15*, 651–675.

Böhm, A., & Kuhn, J. (2000). Soziale Ungleichheit und Gesundheit bei Kindern. Ergebnisse von Einschulungsuntersuchungen im Land Brandenburg. *Soziale Arbeit, 49*, 343–346.

Boekaerts, M. (1999). Motivated learning: Studying student-situation transactional units. *European Journal of Psychology of Education, 14*, 41–55.

Bolger, K. E., Patterson C. J., Thompson, W. W., & Kupersmidt, J. B. (1995). Psychosocial adjustment among children experiencing persistent and intermittent family economic hardship. *Child Development, 66*, 1107–1129.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bonsen, M., Frey, K. A., & Bos, W. (2008). Soziale Herkunft. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 141–156). Münster: Waxmann.

Bonsen, M., Lintorf, K., Bos, W., & Frey, A. (2008). TIMSS 2007 Grundschule – Eine Einführung in die Studie. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 19–48). Münster: Waxmann.

Borduin, C. M., & Henggeler, S. W. (1981). Social class, experimental setting, and task characteristics as determinants of mother-child interaction. *Developmental Psychology, 17*, 209–214.

Bornstein, M. C., & Bradley, R. H. (Eds.). (2003). *Socioeconmic status, parenting, and child development*. Mahwah, NJ: Lawrence Erlbaum.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., Voss, A., et al. (2005). *IGLU: Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

Bos, W., Voss, A., Lankes, E.-M., Schwippert, K., Thiel, O., & Valtin, R. (2004). Schullaufbahnempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 191–220). Münster: Waxmann.

Bossaert, G., Doumen, S., Buyse, E., & Verschueren, K. (2011). Predicting children's academic achievement after the transition to first grade: A two-year longitudinal study. *Journal of Applied Developmental Psychology, 32*, 47–57.

Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212*, 1055–1059.

Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in Western society*. New York: Wiley.

Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. *Soziale Welt, 34*(Sonderband 2), 183–198.

Bourdieu, P., & Passeron, J. (1990). *Reproduction in education, society and culture* (2nd ed., R. Nice, Trans.). Thousand Oaks, CA: Sage.

Bradley, R. H., & Corwyn, R. F. (1999). Parenting. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 339–362). New York: Psychology Press.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371–399.

Bradley, R. H., Corwyn, R. F., Burchinal, M., McAdoo, H., & Garcia Coll, C. (2001). The home environments of children in the United States Part II: Relations with behavioral development through age thirteen. *Child Development, 72*, 1868–1886.

Bradley, R. H., Corwyn, R. F., McAdoo, H., & Garcia Coll, C. (2001). The home environments of children in the United States: Part I. Variations by age, ethnicity, and poverty status. *Child Development, 72*, 1844–1867.

Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society, 9*, 275–305.

Bretherton, I. (1993). From dialogue to internal working models: The co-construction of self in relationships. In C. A. Nelson (Ed.), *Minnesota Symposia on Child Psychology: Vol. 26. Memory and affect* (pp. 237–363). Hillsdale, NJ: Erlbaum.

Bretherton, I., & Munholland, K. A. (1999). Internal working models in attachment relationships. In J. Cassidy & P. Shaver (Eds.), *Handbook of attachment* (pp. 89–111). New

York: Guilford Press.

Brett, J. D., Wilkins, J. L. M., Long, M. H., & Wang, F. (2012). Testing a motivational model of achievement: How students' mathematical beliefs and interests are related to their achievement. *European Journal of Psychology of Education, 27*, 1–20.

Brody, G. H., & Flor, D. (1997). Maternal psychological functioning, family processes, and child adjustment in rural, single-parent, African-American families. *Developmental Psychology, 33*, 1000–1011.

Brody, D. J., Pirkle, J., Kramer, R., Flegal, K., Matte, T. D., Gunter, E., et al. (1994). Blood lead levels in the U.S. population: Phase I of the Third National Health and Nutrition Examination Survey (NHANES III, 1988–1991). *Journal of the American Medical Association, 272*, 277–283.

Brooks-Gunn, J., Klebanov, P., & Liaw, E (1995). The learning, physical, and emotional environment of the home in the context of poverty: The Infant Health and Development Program. *Children and Youth Services Review, 17*, 231–250.

Brunner, M. (2008). No g in education? *Learning and Individual Differences, 18*, 152–165.

Brunner, M. & Krauss, S. (2010). Modellierung kognitiver Kompetenzen von Schülern und Lehrkräften mit dem Nested-Faktormodell. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung* (S. 105–126). Münster: Waxmann.

Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence, 36*, 403–421.

Bundesagentur für Arbeit (2008). *Statistik der Grundsicherung für Arbeitsuchende nach dem SGB II. SGB II-Kennzahlen für interregionale Vergleiche*. Nürnberg: BA.

Bullard, R. D., & Wright, B. (1993). Environmental justice for all: Current perspectives on health and research needs. *Toxicology and Industrial Health, 9*, 821–841.

Bybee, R.W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.), *Scientific literacy* (pp. 37–68). Kiel: IPN.

Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, D.C.: American Psychological Association.

Byrnes, J. P., & Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity-propensity analysis. *Contemporary Educational Psychology, 32*, 599–629.

Calsyn, R., & Kenny, D. (1977). Self-concept of ability and perceived evaluations by others: Cause or effect of academic achievement? *Journal of Educational Psychology, 69*, 136–145.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

Campione, J. C., & Brown, A. L. (1978). Toward a theory of intelligence: Contributions from research with retarded children. *Intelligence, 2*, 279–304.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Caro, D. H., & Lehmann, R. (2009). Achievement inequalities in Hamburg schools: How do they change as students get older? *School Effectiveness and School Improvement, 20*, 407–431.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Case, R. (1992). *The mind's staircase*. Hillsdale, NJ: Erlbaum.

Cattell, R. B. (1957). *The IPAT Culture Fair Intelligence Scales*. Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. Amsterdam: Elsevier.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*, 703–722.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "Have-Nots" gain but the "Haves" gain even more. *American Psychologist, 60*, 149–160.

Chamorro-Premuzic, T., Arteche, A., Furnham, A., & Trickot, N. (2009). Assessing pupils' intelligence through self, parental, and teacher estimates. *Educational Psychology, 29*, 83–97.

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology, 67*, 279–291.

Cheadle, J. E. (2009). Parent educational investment and children's general knowledge development. *Social Science Research, 38*, 471–491.

Chen, S.-K., Yeh, Y.-C., Hwang, F.-M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort-multioccasion study. *Learning and Individual Differences, 23*, 172–178.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chi, M. T., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–76). Hillsdale, NJ: Erlbaum.

Chipuer, H. M., Rovine, M. J., & Plomin, R. (1990). LISREL modeling: Genetic and environmental influences on IQ revisited. *Intelligence, 14*, 11–29.

Chiu, M.-S. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology, 104*, 87–107.

Christensen, H., Batterham, P. J., & Mackinnon, A. J. (2013). The getting of wisdom: Fluid intelligence does not drive knowledge acquisition. *Journal of Cognition and Development, 14*, 321–331.

Cianciolo, A. T., & Sternberg, R. J. (2004). *Intelligence: A brief history*. Malden, MA: Blackwell.

Cochran, M., Larner, M., Riley, D., Gunnarson, L., & Henderson, C. (1990). *Extending families: The social networks of parents and their children*. New York: Cambridge University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558–577.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*(Suppl.), 95–120.

Coley, R. J. (2002). *An uneven start: Indicators of inequality in school readiness*. Princeton, NJ: Educational Testing Service.

Collins, W. A. (1990). Parent-child relationships in the transition to adolescence: Continuity and change in interaction, affect, and cognition. In G. R. Adams & T. P. Gullotta (Eds.), *From childhood to adolescence: A transitional period* (pp. 85–106)? Beverly Hills, CA: Sage.

Colom, R., & Flores-Mendoza, C. E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence, 35*, 243–251.

Conger, R. D., Conger, K. J., Elder, G. H., Lorenz, F. O., Simons, R. L., & Whitbeck, L. B. (1993). Family economic stress and adjustment of early adolescent girls. Developmental Psychology, 29, 206–219.

Conger, R. D., & Elder, G. H. (1994). *Families in troubled times*. New York: Aldine de Gruyter.

Connell, J. P. (1985). A new multidimensional measure of children's perception of control. *Child Development, 56*, 1018–1041.

Coopersmith, S. A. (1967). *The antecedents of self-esteem*. San Francisco: Freeman.

Corno, L., Cronbach, L. J., & Kupermintz, H. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mahwah, NJ: Erlbaum.

Costu, B., & Ayas, A. (2005). Evaporation in different liquids: Secondary students' conceptions. *Research in Science & Technological Education, 23*, 75–97.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.

Cronbach, L. J. (1960). *Essential of psychological testing*. Oxford, UK: Harper.

Crouter, A. C., Manke, B. A., & McHale, S. M. (1995). The family context of gender intensification in early adolescence. *Child Development, 66*, 317–329.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: Experiencing flow in work and play*. San Francisco: Jossey-Bass.

Daniels, Z. (2008). *Entwicklung schulischer Interessen im Jugendalter*. (Reihe Pädagogische Psychologie und Entwicklungspsychologie, Bd. 69). Münster: Waxmann.

Davis, O. S. P., Haworth, C. M. A., & Plomin, R. (2009). Dramatic increase in heritability of cognitive development from early to middle childhood. *Psychological Science, 20*, 1301–1308.

Deary, I. J. (1995). Auditory inspection time and intelligence: What is the direction of causation? *Developmental Psychology, 31*, 237–250.

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63*, 453–482.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.

De Boeck, P., & Wilson, M. R. (Eds.) (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

Deci, E. L. (1998). The relation of interest to motivation and human needs: The self-determination theory viewpoint. In L. Hoffmann, A. Krapp, K. A. Renninger & J. Baumert (Eds.), *Interest and learning. Proceedings of the Seeon-Conference on interest and gender* (pp. 146–162). Kiel, Germany: IPN.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality, 19*, 109–134.

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268.

Deci, E. L., & Ryan, R. M. (2002). Self-determination research: Reflections and future directions. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination theory research* (pp. 431–441). Rochester, NY: University of Rochester Press.

Dedrick, R. F., Ferron, J. M., Hess, M. L., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., et al. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79*, 69–102.

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

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature, 388*, 468–471.

Diamond, J. B., Randolph, A., & Spillane, J. P. (2004). Teachers' expectations and sense of responsibility for student learning: The importance of race, class, and organizational habitus. *Anthropology & Education Quarterly, 35*, 75–98.

diSessa, A. (2006). A history of conceptual change research. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 265–281). Cambridge, UK: Cambridge University Press.

Ditton, H. (2005). Der Beitrag von Familie und Schule zur Reproduktion von Bildungsungleichheit. In H. G. Holtappels & K. Höhmann (Hrsg.), *Schulentwicklung und Schulwirksamkeit* (S. 121–130). Weinheim: Juventa.

Ditton, H. (2013). Wer geht auf die Hauptschule? Primäre und sekundäre Effekte der sozialen Herkunft beim Übergang nach der Grundschule. *Zeitschrift für Erziehungswissenschaft, 16*, 731–749.

Ditton, H., & Krüsken, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft, 9*, 348–372.

Ditton, H., & Krüsken, J. (2009). Bildungslaufbahnen im differenzierten Schulsystem – Entwicklungsverläufe von Laufbahnempfehlungen und Bildungsaspirationen in der Grundschulzeit. *Zeitschrift für Erziehungswissenschaft, 12*, 74–102.

Ditton, H., Krüsken, J., & Schauenberg, M. (2005). Bildungsungleichheit – Der Beitrag von Familie und Schule. *Zeitschrift für Erziehungswissenschaft, 8*, 285–304.

Dochy, F. C. (1992). *Assessment of prior knowledge as a determinant for future learning: The use of prior knowledge state tests and knowledge profiles*. Utrecht, The Netherlands: Lemma.

Dochy, F. C. (1994). Prior knowledge and learning. In T. Husen & T. N. Postlethwaite (Eds.), *International Encyclopedia of Education* (2nd ed., pp.4698–4702). Oxford, UK: Pergamon.

Dochy, F. C., & Alexander, P. A. (1995). Mapping prior knowledge: A framework for discussion among researchers. *European Journal of Psychology of Education, 10*, 225–242.

Dochy, F. C., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*, 145–186.

Dodge, K. A., Pettit, G., & Bates, J. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65*, 649–665.

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69*, 613–635.

Drewnowski, A., & Specter, S. E. (2004). Poverty and obesity: The role of energy density and energy costs. *American Journal of Clinical Nutrition, 79*, 6–16.

Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education, 11*, 481–490.

Driver, R., Squires, R., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. New York: Routledge.

Dubow, E., & Ippolito, M. E (1994). Effects of poverty and quality of the home environment on changes in the academic and behavioral adjustment of elementary school-age children. *Journal of Clinical Child Psychology, 23*, 401–412.

Dubow, E., Tisak, J., Causey, D., Hryshko, A., & Reid, G. (1991). A two-year longitudinal study of stressful life events, social support, and social problem-solving skills: Contributions to children's behavioral and academic adjustment. *Child Development, 62*, 583–599.

Duit, R. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 263–282). New York: Pergamon.

Duit, R., & Häußler, P. (1997). Physik und andere naturwissenschaftliche Lernbereiche. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie, Serie Pädagogische Psychologie* (S. 427–460). Göttingen: Hogrefe.

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*, 671–688.

Dumont, H., Trautwein, U., & Lüdtke, O. (2012). Familiärer Hintergrund und die Qualität elterlicher Hausaufgabenhilfe. *Psychologie in Erziehung und Unterricht, 59*, 109–121.

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. (1994). Economic deprivation and early childhood development. *Child Development, 65*, 296–318.

Duncan, G., & Brooks-Gunn, J. (Eds.). (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.

Duncan, G. J., Brooks-Gunn, J., Yeung, W. J., & Smith, J. R. (1998). How much does childhood poverty affect the life changes of children? *American Sociological Review, 63*, 406–423.

Dweck, C. S. (2002). The development of ability conceptions. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 57–88). San Diego, CA: Academic Press.

Eccles-Parsons, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco: Freeman.

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence, 21*, 225–241.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the achiever: The structure of adolescents' academic achievement related-beliefs and self-perceptions. *Personality and Social Psychology Bulletin, 21*, 215–225.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132.

Eccles, J. S., Wigfield, A., Flanagan, C., Miller, C., Reuman, D., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality, 57*, 283–310.

Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64*, 830–847.

Ehmke, T. (2008). Welche Bedeutung haben lernförderliche Prozesse und naturwissenschaftsbezogene Einstellungen im Elternhaus für die Erklärung sozialer Disparitäten in der naturwissenschaftlichen Kompetenz? *Zeitschrift für Erziehungswissenschaft, 10*(Sonderheft 10), 129–148.

Ehmke, T., & Baumert, J. (2007). Soziale Herkunft und Kompetenzerwerb: Vergleiche zwischen PISA 2000, 2003 und 2006. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 309–335). Münster: Waxmann.

Ehmke, T., Hohensee, F., Siegle, T., & Prenzel, M. (2006). Soziale Herkunft, elterliche Unterstützungsprozesse und Kompetenzentwicklung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 225–248). Münster: Waxmann.

Ehmke, T., & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231–254). Münster: Waxmann.

Elder, G. (1974). *Children of the Great Depression*. Chicago: University of Chicago Press.

Elsässer, G. (1998). Gesundheit von Schulanfängern. Auswirkung sozialer Benachteiligung am Beispiel der Brandenburger Schulanfängeruntersuchungen 1994 und 1995. *Das Gesundheitswesen, 60*, 632–637.

Emery, R. E., & Laumann-Billings, L. (1998). An overview of the nature, causes, and consequences of abusive family relationships. *American Psychologist, 53*, 121–135.

Enders, C. (2010). *Applied missing data analysis*. New York: Guilford.

Ensminger, M. E., Forrest, C. B., Riley, A. W., Kang, M., Green, B. F., Starfield, B., et al. (2000). The validity of measures of socioeconomic status of adolescents. *Journal of Adolescent Research, 15*, 392–419.

Entwistle, D. R., Alexander K. L., & Olson, L. S. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review, 59*, 822–838.

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1983). International class mobility and the convergence thesis: England, France and Sweden. *British Journal of Sociology, 34*, 303–343.

Evans, G. W. (2001). Environmental stress and health. In A. Baum, T. Revenson, & J. E. Singer (Eds.), *Handbook of health psychology* (pp. 365–385). Mahwah, NJ: Erlbaum.

Evans, G. W. (2004). The environment of childhood poverty. American Psychologist, 59, 77–92.

Evans, G. W., & English, K. (2002). The environment of poverty: Multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child Development, 73*, 1238–1248.

Ewerhardy, A. (2011). *Zusammenhänge zwischen Verständnisorientierung von naturwissenschaftsbezogenem Sachunterricht und Fortschritten im Verständnis naturwissenschaftlicher Konzepte bei Lernenden der Grundschule*. Elektronisch veröffentlichte Dissertationsschrift, Westfälische Wilhelms-Universität Münster.

Ewerhardy, A., Kleickmann, T., & Möller, K. (2012). Fördert ein konstruktivistisch orientierter naturwissenschaftlicher Sachunterricht mit strukturierenden Anteilen das konzeptuelle Verständnis bei den Lernenden? *Zeitschrift für Grundschulforschung, 5*, 76–88.

Expertenrat Herkunft und Bildungserfolg (Hrsg.). (2011). *Empfehlungen für Bildungs-politische Weichenstellungen in der Perspektive auf das Jahr 2020 (BW 2020)*. Stuttgart: Ministerium für Kultus, Jugend und Sport Baden-Württemberg.

Farah, M. J., & Hackman, D. A. (2012). SES, childhood experience, and the neural bases of cognition. In V. Maholmes & R. B. King (Eds.), *The Oxford handbook of poverty and child development* (pp. 307–318). New York: Oxford University Press.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review, 13*, 1–22.

Feiring, C., & Taska, L. S. (1996). Family self-concept: Ideas on its meaning. In B. Bracken (Ed.), *Handbook of self-concept* (pp. 317–373). New York: Wiley.

Fend, H. (2000). *Entwicklungspsychologie des Jugendalters*. Opladen: Leske + Budrich.

Filipp, S.-H. (2006). Entwicklung von Fähigkeitsselbstkonzepten. *Zeitschrift für Pädagogische Psychologie, 20*, 65–72.

Fischer, K. W., Hand, H. H., Watson, M. W., Van Parys, M., & Tucker, J. (1984). Putting the child into socialization: The development of social categories in preschool children. In L. Katz (Ed.), *Current topics in early childhood education* (Vol. 5, pp. 27–72). Norwood, NJ: Ablex.

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions, 21*, 1095.

Floyd, R. G. (2005). Information-processing approaches to interpretation of contemporary intellectual assessment instruments. In D. P. Flanagan, & P. L. Harrison (Eds.), *Con-*

temporary intellectual assessment: Theories, tests, and issues (pp. 203–233). New York: Guilford.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22*, 200–233.

Fredricks, J., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*, 519–533.

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence, 20*, 507–537.

Frenzel, A. C., Pekrun, R., Dicke, A.-L., & Goetz, T. (2012). Beyond quantitative decline: Conceptual shifts in adolescents' development of interest in mathematics. *Developmental Psychology, 48*, 1069–1082.

Freund, Ph. A., & Holling, H. (2008). Creativity in the classroom: A multilevel analysis investigating the impact of creativity and reasoning ability on GPA. *Creativity Research Journal, 20*, 309–318.

Fricke, K., van Ackeren, I., Kauertz, A., & Fischer, H. E. (2012). Students' perceptions of their teacher's classroom management in elementary and secondary science lessons. In T. Wubbels, J. van Tartwijk, P. den Brok & J. Levy (Eds.), *Interpersonal relationships in education* (pp. 167–185). Rotterdam, The Netherlands: SensePublishers.

Gale, C. R., Batty, G. D., Tynelius P., Deary I. J., & Rasmussen F. (2010). Intelligence in early adulthood and subsequent hospitalization for mental disorders. *Epidemiology, 21*, 70–77.

Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*, 1–56.

Garner, R. (1987). *Metacognition and reading comprehension*. Westport, CT: Ablex.

Garrett, P., Ng'andu, N., & Ferron, J. (1994). Poverty experiences of young children and the quality of their home environments. *Child Development, 65*, 331–345.

Geiser, C. (2010). Datenanalyse mit Mplus. Wiesbaden: VS Verlag für Sozialwissenschaften.

Geißler, R. (2004). Die Illusion der Chancengleichheit im Bildungssystem – von PISA gestört. *Zeitschrift für Soziologie der Erziehung und Sozialisation, 24*, 362–380.

Gesellschaft für Didaktik des Sachunterrichts (Hrsg.). (2002). *Perspektivrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39*, 93–104.

Glutting, J. J., Watkins, M. M., & Youngstrom, E. A. (2003). Multifactored and cross-battery ability assessments: Are they worth the effort? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (pp. 343–377). (2nd Ed.). New York: Guilford Press.

Gniewosz, B. (2010). Die Konstruktion des akademischen Selbstkonzeptes. Eltern und Zensuren als Informationsquellen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42*, 133–142.

Gniewosz, B., & Noack, P. (2012). What you see is what you get: The role of early adolescents' perceptions in the intergenerational transmission of academic values. *Contemporary Educational Psychology, 37*, 70–79.

Goldthorpe, J. H. (1996). Class analysis and the reorientation of class theory: The case of persisting differentials in educational attainment. *British Journal of Sociology, 47*, 481–505.

Gottfredson, L. S. (2002). g: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331−380). New Jersey: LEA.

Gottfredson, L. S. (2003). g, jobs and life. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 293–342). Amsterdam, The Netherlands: Pergamon.

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology, 77*, 631–645.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (1994). Role of parental motivational practices in children's academic intrinsic motivation and achievement. *Journal of Educational Psychology, 86*, 104–113.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93*, 3–13.

Gottfried, A. W. (1985). Measures of socioeconomic status in child development research: Data and recommendations. *Merrill-Palmer Quarterly, 31*, 85–92.

Gottfried, A. W., Gottfried, A. E., Bathurst, K, Guerin, G.W., & Parramore, M. (2003). Socioeconomic status in children's development and family environment: Infancy through adolescence. In M. Bornstein & R. Bradley (Eds.), *Socioeconomic status, parenting, and child development*. Mahwah, NJ: Erlbaum.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 87–114). New York: John Wiley & Sons.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206–213.

Grant, M. D, Kremen, W. S., Jacobson, K. C., Franz, C., Xian, H., Eisen, S. A., et al. (2010) Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood? *Behavior Genetics, 40*, 438–446.

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

Guo, G. (1998). The timing of the influences of cumulative poverty on children's cognitive ability and achievement. *Social Forces, 77*, 257–288.

Häußler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of Research in Science Teaching, 39*, 870–888.

Hauser, R. M., & Huang, M. H. (1997). Verbal ability and socioeconomic success: A trend analysis. *Social Science Research, 26*, 331–376.

Haines, M. M., Stansfeld, S. A., Head, J., & Job, R. F. S. (2002). Multilevel modeling of aircraft noise on performance tests in schools around Heathrow Airport London. *International Journal of Epidemiology and Community Health, 56*, 139–144.

Haller, E. J., & Davis, S. A. (1981). Teacher perceptions, parental social status, and grouping for reading instruction. *Sociology of Education, 54*, 162–174.

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS ONE, 7*(2), 1–16.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52*, 123–142.

Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for IRT parameters using separate versus concurrent estimation in the common item nonequivalent groups equating design. *Applied Psychological Measurement, 26*, 3–24.

Harlen, W. (1998). Teaching for understanding in pre-secondary science. In B. Fraser & K. Tobin (Eds.), *International handbook of science education* (Vol. 2, pp. 183–197). Dordrecht; The Netherlands: Kluwer.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: Brookes.

Harter, S. (1982). The perceived competence scale for children. *Child Development, 53*, 87–97.

Harter, S. (1986). Cognitive-developmental processes in the integration of concepts about emotions and the self. *Social Cognition, 4*, 119–151.

Harter, S. (1990). Causes, correlates and the functional role of global self-worth: A life-span perspective. In R. Sternberg & J. Kolligian Jr. (Eds.), *Competence considered* (pp. 67–

98). New Haven, CT: Yale University Press.

Harter, S. (1999). *The construction of the self*. NewYork: Guilford Press.

Harter, S. (2006). The self. In W. Damon, R. Lerner & N. Eisenberg (Eds.), *Handbook of child psychology* (pp. 505–570). New York: Wiley.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*, 1969–1982.

Hasselhorn, M., & Grube, D. (1997). Entwicklung der Intelligenz und des logischen Denkens: Theoretische Perspektiven und eine perspektivische Theorie. In F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 15–26). München: PVU.

Hasselhorn, M., & Silbereisen, R. (Hrsg.). (2008). *Enzyklopädie der Psychologie, Serie Entwicklungspsychologie: Entwicklungspsychologie des Jugendalters*. Göttingen: Hogrefe.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Havighurst, R. (1972). *Developmental tasks and education* (3rd ed.). New York: Random House.

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

Helmke, A. (2007). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Klett-Kallmeyer.

Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität*. Seelze: Klett-Kallmeyer.

Helmke, A., & Jäger, R. S. (Hrsg.). (2002). *Das Projekt MARKUS – Mathematik-Gesamterhebung Rheinland-Pfalz. Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau: Verlag Empirische Pädagogik.

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

Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie, Serie Pädagogische Psychologie* (S. 71–176). Göttingen: Hogrefe.

Helsper, W., Kramer, R.-T., Hummrich, M., & Busse, S. (2009). *Jugend zwischen Familie und Schule: Eine Studie zu pädagogischen Generationsbeziehungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hess, R., & Holloway, S. (1984). Family and school as educational institutions. In R. D. Parke (Ed.), *Review of child development research* (Vol. 7, pp. 179–222). Chicago, IL: University of Chicago Press.

Higgins, E. T. (1991). Development of self-regulatory and self-evaluative processes: Costs, benefits, and tradeoffs. In M. R. Gunnar & L. A. Sroufe (Eds*.), Minnesota Symposia on Child Development: Vol. 23. Self processes and development* (pp. 125–166). Hillsdale, NJ: Erlbaum.

Hidi, S. (1995). A re-examination of the role of attention in learning from text. *Educational Psychology Review, 7*, 323–350.

Hidi, S. (2000). An interest researcher's perspective: The effects of extrinsic and intrinsic factors on motivation. In S. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 309–339). New York: Academic Press.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*, 111–127.

Hidi, S., Renninger, K. A., & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 89–115). Mahwah, NJ: Erlbaum.

Hofer, M. (2010). Adolescents' development of individual interests: A product of multiple goal regulation? *Educational Psychologist, 45*, 149–166.

Hofer, M., Schmid, S., Fries, S., Dietz, F., Clausen, M., & Reinders, H. (2007). Individual values, motivational conflicts, and learning for school. *Learning and Instruction, 17*, 17-28.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*, 1368–1378.

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting* (pp. 231–252). Mahwah, NJ: Erlbaum.

Hoffmann, L., Häußler, P., & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik*. Kiel: IPN.

Hoffmann, L., Krapp, A., Renninger, K. A., & Baumert, J. (Eds.). (1998). *Interest and learning: Proceedings of the Seeon-Conference on interest and gender*. Kiel, Germany: IPN.

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika, 55*, 577–601.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Holtappels, H., Radisch, F., Rollett, W., & Kowoll, M. E. (2010). Bildungsangebot und Schülerkompetenzen in Ganztagsgrundschulen. In W. Bos, S. Hornberg, K. Arnold, G. Faust, L. Fried, E. Lankes et al. (Hrsg.), *IGLU 2006 – die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens* (S. 165–198). Münster: Waxmann.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf–Gc theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53−91). New York: Guilford Press.

Hua, Z., & David, A. (2008). Study design: Cross-sectional, longitudinal, case, and group. In L. Wei & M. G. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 88–107). Malden, MA: Blackwell Publishing.

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*, 505–528.

Hülsheger, U. R., Maier, G. W. & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment, 15*, 3–18.

Hunt, E. (2011). *Human intelligence*. Cambridge, UK: Cambridge University Press.

Ingersoll, R. (1999). The problem of under qualified teachers in American secondary schools. *Educational Researcher, 28*, 26–37.

Hutten, L. (1980, April). *Some empirical evidence for latent trait model selection*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Jacobs, C., Petermann, F., & Weiß, R. H. (2007). Testinformation: Grundintelligenztest (CFT 20-R). *Diagnostica, 53*, 109–113.

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter? *New Directions for Child and Adolescent Development, 106*, 5–20.

Jacobs, J. E., Davis-Kean, P., Bleeker, M. M., Eccles, J. S., & Malanchuk, O. (2005). I can, but I don't want to: The impact of parents, interests, and activities on gender differences in math. In A. Gallagher & J. Kaufman (Eds.), *Gender differences in mathematics* (pp. 246–263). Cambridge: Cambridge University Press.

Jaeger, M. M. (2011). Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data. *Sociology of Education, 84*, 281–298.

Janssen, R. (2011). Using a differential item functioning approach to investigate measurement invariance. In E. Davidov, P. Schmidt & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 415–432). New York: Routledge.

Jensen, A. R. (1980). Uses of sibling data in educational and psychological research. *American Educational Research Journal, 17*, 153–170.

Jensen, A. R. (1998). *The g-factor: The science of mental ability*. New York: Praeger.

Jensen, A. R. (2002). Psychometric g: Definition and substantiation. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 39–53). Mahwah, NJ: Erlbaum.

Jensen, A. R., & Weng, L.-J. (1994). What is a good g? *Intelligence, 18*, 231–258.

Jensen, E., James, S., Boyce, T., & Hartnett, S. (1983). The Family Routine Inventory: Development and validation. *Social Science and Medicine, 17*, 201–211.

Johnson, P. (1998). Children's understanding of changes of state involving the gas state, Part 2: Evaporation and condensation below boiling point. *International Journal of Science Education, 20*, 695–709.

Jones, K. K., & Byrnes, J. P. (2006). Characteristics of students who benefit from high-quality mathematics instruction. *Contemporary Educational Psychology, 31*, 328–343.

Jungbauer-Gans, M. (2004). Einfluss des sozialen und kulturellen Kapitals auf die Lesekompetenz: Ein Vergleich der PISA 2000-Daten aus Deutschland, Frankreich und der Schweiz. *Zeitschrift für Soziologie, 33*, 375–397.

Jungbauer-Gans, M. (2006). Kulturelles Kapital und Mathematikleistungen: Eine Analyse der PISA 2003-Daten für Deutschland. In W. Georg (Hrsg.), *Soziale Ungleichheit im Bildungssystem: Eine empirisch-theoretische Bestandsaufnahme* (S. 175–198). Konstanz: UVK.

Jussim, L. J., & Eccles, J. (1995). Are teacher expectations biased by students' gender, social class, or ethnicity? In Y.-T. Lee, L. J. Jussim & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 245–271). Washington, DC: APA.

Kammermeyer, G., & Martschinke, S. (2006). Selbstkonzept- und Leistungsentwicklung in der Grundschule – Ergebnisse aus der KILIA-Studie. *Empirische Pädagogik, 20*, 245–259.

Kane, M. (2011). The errors of our ways. *Journal of Educational Measurement, 48*, 12–30.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1*, 152–176.

Karing, C., Matthäi, J., & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie, 25*, 159–172.

Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, 135–153.

Kim, S., & Kolen, M. J. (2006). Robustness to format effects of IRT linking methods for mixed-format tests. *Applied Measurement in Education, 19*, 357–381.

Kim, Y., & Sherraden, M. (2011). Do parental assets matter for children's educational attainment? Evidence from mediation tests. *Children and Youth Services Review, 33*, 969–979.

Kleickmann, T., Brehl, T., Saß, S., Prenzel, M., & Köller, O. (2012). Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In W.

Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 123–169). Münster: Waxmann.

Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Göttingen: Hogrefe.

Knight, G. P., Roosa, M. W., & Umaña-Taylor, A. J. (2009). Sampling, recruiting, and retaining diverse samples. In G. P. Knight, M. W. Roosa & A. J. Umaña-Taylor (Eds.), *Studying ethnic minority and economically disadvantaged populations: Methodological challenges and best practices* (pp. 29–78). Washington, DC: APA.

Köller, O., & Baumert, J. (2001). Leistungsgruppierungen in Sekundarstufe I. Ihre Konsequenzen für die Mathematikleistung und das mathematische Selbstkonzept der Begabung. *Zeitschrift für Pädagogische Psychologie, 15*, 99–110.

Köller, O., Baumert, J., & Schnabel, K. U. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal of Research in Mathematics Education, 32*, 448–470.

Kohn, M. L. (1977). *Social class and conformity*. Chicago: University of Chicago.

Kohn, M. L., & Schooler, C. (1982). Job conditions and personality: A longitudinal assessment of their reciprocal effects. *American Journal of Sociology, 87*, 1257–1283.

Korat, O. (2011). Mothers' and teachers' estimations of first graders' literacy level and its relation to the child's actual performance in different SES groups. *Education & Treatment of Children, 34*, 347–371.

Korenman, S., Miller, J., & Sjaastad, J. (1995). Long-term poverty and child development in the United States: Results from the NLSY. *Children and Youth Services Review, 17*, 127–155.

Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: Vertiefende und weiterführende Metaanalysen. *Psychologische Rundschau, 60*, 82–98.

Krampen, G. (1988). Competence and control orientations as predictors of test anxiety in students: Longitudinal results. *Anxiety Research, 1*, 185–197.

Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht, 44*, 185–201.

Krapp, A. (2002a). An educational-psychological theory of interest and its relation to SDT. In E. L. Deci & R. M. Ryan, *Handbook of self-determination research* (pp. 405–427). Rochester, NY: University of Rochester Press.

Krapp, A. (2002b). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12*, 383–409.

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction, 15*, 381–395.

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning and development. In K. A. Renninger, S. Hidi, & A. Krapp, *The role of interest in learning and development* (pp. 3–25). Hillsdale, NJ: Erlbaum.

Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education, 33*, 27–50.

Kratzmann, J., & Schneider, T. (2009). Soziale Ungleichheiten beim Schulstart. Empirische Untersuchungen zur Bedeutung der sozialen Herkunft und des Kindergartenbesuchs auf den Zeitpunkt der Einschulung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 61, 1–24.

Krause, U., & Stark, R. (2006). *Vorwissen aktivieren*. In H. Mandl & H. Friedrich (Hrsg.), Handbuch Lernstrategien (S. 38–49). Göttingen: Hogrefe.

Kreppold, M. (2012). *Die Auswirkungen des Schulübertritts auf die Entwicklung von Fähigkeitsselbstkonzept, Depression und physischen Beschwerden*. Elektronisch veröffentlichte Dissertationsschrift, Justus-Liebig-Universität Gießen.

Kubinger, K. D., & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica, 53*, 131–143.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*, 148–161.

Kunz, G. C., Drewniak, U., Hatalak, A., & Schön, A. (1992). Zur differentiellen Bedeutung kognitiver, metakognitiver und motivationaler Variablen für das effektive Lernen mit Instruktionstexten und Bildern. In H. Mandl & H. F. Friedrich (Hrsg.), *Lern- und Denkstrategien* (S. 213–229). Göttingen: Hogrefe.

Kvist, A., & Gustafsson, J. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence, 36*, 422–436.

Lam, C. B., McHale, S. M., & Crouter, A. C. (2012). Parent-child shared time from middle childhood to late adolescence: Developmental course and adjustment correlates. *Child Development, 83*, 2089–2103.

Landerman, L. R., Land, K. C., & Pieper, C. F. (1997). An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods & Research, 26*, 3–33.

Landesbetrieb Information und Technik Nordrhein-Westfalen (2009). *Statistische Berichte. Gerichtliche Ehelösungen in Nordrhein-Westfalen 2008*. Düsseldorf: IT.NRW, Geschäftsbereich Statistik.

Landesbetrieb Information und Technik Nordrhein-Westfalen (2011). *Einkommen der Lohn- und Einkommensteuerpflichtigen in NRW 2007* (Tabellenanhang 14.9101 zur Pressemitteilung 160/11: Einkommen 2007 im Rheinland höher als in Westfalen). Düsseldorf: IT.NRW, Geschäftsbereich Statistik.

Lange, K. (2010). *Zusammenhänge zwischen naturwissenschaftsbezogenem fachspezifisch-*

pädagogischem Wissen von Grundschullehrkräften und Fortschritten im Verständnis naturwissenschaftlicher Konzepte bei Grundschülerinnen und -schülern. Elektronisch veröffentlichte Dissertationsschrift, Westfälische Wilhelms-Universität Münster.

Lange, K., Kleickmann, T., Tröbst, S., & Möller, K. (2012). Fachdidaktisches Wissen von Lehrkräften und multiple Ziele im naturwissenschaftlichen Sachunterricht. *Zeitschrift für Erziehungswissenschaft, 15*, 55–76.

Lareau, A. (2003). *Unequal childhoods: Class, race, and family life*. Berkeley, CA: UC Press.

Lavonen, J., & Laaksonen, S. (2009). Context of teaching and learning school science in Finland: Reflections on PISA 2006 results. *Journal of Research in Science Teaching, 46*, 922–946.

Lee, V., & Croninger, R. (1994). The relative importance of home and school in the development of literacy skills for middle-grade students. *American Journal of Education, 102*, 286–329.

Lee, W.-C., & Ban, J.-C. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education, 19*, 23–48.

Lehmann, R. H., Peek, R., & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.

LeTendre, G. K., Hofer, B. K., & Shimizu, H. (2003). What is tracking? Cultural expectations in the United States, Germany, and Japan. *American Educational Research Journal, 40*, 43–89.

Liaw, F., & Brooks-Gunn, J. (1994). Cumulative familial risks and low birth weight children’s cognitive and behavioral development. *Journal of Clinical Child Psychology, 23*, 360–372.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11*, 357–380.

Linver, M. R., Brooks-Gunn, J., & Kohen, D. (2002). Family processes in pathways from income to young children’s development. *Developmental Psychology, 38*, 719–734.

Lipowsky, F. (2006). Auf den Lehrer kommt es an. *Zeitschrift für Pädagogik*, *51*, 47–65.

Lipowsky, F., Kastens, C., Lotz, M., & Faust, G. (2011). Aufgabenbezogene Differenzierung und Entwicklung des verbalen Selbstkonzepts im Anfangsunterricht. *Zeitschrift für Pädagogik, 57*, 868–884.

Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176−193). Cambridge: Cambridge University Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York: Erlbaum Associates.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA:

Addison-Wesley.

Lüdtke, O., & Robitzsch, A. (2010). Umgang mit fehlenden Daten in der empirischen Bildungsforschung. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online. Fachgebiet Methoden der empirischen erziehungswissenschaftlichen Forschung. Quantitative Forschungsmethoden*. Weinheim: Juventa.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau, 58*, 103–117.

Luke, D. A. (2004). *Multilevel modeling* (Sage University Papers Series on Quantitative Applications in the Social Sciences, Series No. 07-143). Thousand Oaks, CA: Sage.

Luster, T., Rhoades, K., & Haas, B. (1989). The relations between parental values and parenting behavior: A test of the Kohn hypothesis. *Journal of Marriage and the Family, 51*, 139–147.

Maaz, K., Baumert, J., & Trautwein, U. (2009). Genese sozialer Ungleichheit im institutionellen Kontext der Schule: Wo entsteht und vergrößert sich soziale Ungleichheit? *Zeitschrift für Erziehungswissenschaft, 12*, 11–46.

Maaz, K., Kreuter, F., & Watermann, R. (2006). Schüler als Informanten? Die Qualität von Schülerangaben zum sozialen Hintergrund. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen* (S. 31–59). Wiesbaden: VS Verlag für Sozialwissenschaften.

Maaz, K., & Nagy, G. (2009). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems: Definition, Spezifikation und Quantifizierung primärer und sekundärer Herkunftseffekte. *Zeitschrift für Erziehungswissenschaft, 12*, 153–182.

Maaz, K., Schroeder, S., & Gresch, C. (2010). Primäre und sekundäre soziale Herkunftseffekte beim Übergang in Sekundarstufe I. Neutralisation sozialer Herkunftseffekte und Konsequenzen auf das Übergangsverhalten. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung* (S. 285–310). Münster: Waxmann.

Maaz, K., Trautwein, U., Gresch, C., Lüdtke, O., & Watermann, R. (2009). Intercoder-Reliabilität bei der Berufscodierung nach der ISCO-88 und Validität des sozioökonomischen Status. *Zeitschrift für Erziehungswissenschaften, 12*, 281–301.

Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*, 99–106.

Maaz, K. & Watermann, R. (2007). Reproduktion oder Mobilität? Zur Wirkung familiärer Prozessmerkmale auf die Studienintention am Ende der gymnasialen Oberstufe. *Zeitschrift für Soziologie der Erziehung und Sozialisation, 27*, 285–303.

Magnusson, K. A., & Duncan, G. J. (2002). Parents in poverty. In M. H. Bornstein (Ed.), *Handbook of parenting* (pp. 95–121). Mahwah, NJ: Erlbaum.

Mandl, H., Gruber, H., & Renkl, A. (1993). Misconceptions and knowledge compartmentalization. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge: The German Wissenspsychologie project* (pp. 161–176). Amsterdam, The Netherlands: Elsevier.

Mapuranga, R., Dorans, N. J., & Middleton, K. (2008). *A review of recent developments in differential item functioning* (ETS Research Rep. No. 08-43). Princeton, NJ: Educational Testing Service.

Marais, I., & Andrich, D. (2008) Formalising dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*, 200–215.

Marks, G. N., Cresswell, J., & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation, 12*, 105–128.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, 129–149.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 81*, 417–430.

Marsh, H. W. (1990a). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82*, 646–656.

Marsh, H. W. (1990b). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*, 623–636.

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. London: British Psychological Society.

Marsh, H. W., & Ayotte, V. (2003). Do multiple dimensions of self-concept become more differentiated with age? The differential distinctiveness hypothesis. *Journal of Educational Psychology, 95*, 687–706.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). San Diego, CA: Academic Press.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163.

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology, 83*, 377–392.

Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort-multioccasion study. *Child Development, 69*, 1030–1053.

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist, 58*, 364–376.

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56–67.

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

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*, 542–552.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*, 107–123.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., et al. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*, 319–350.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397–416.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality, 74*, 403-456.

Marsh, H. W., Xu, M., & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra & J. Sweller (Eds), *APA educational psychology handbook, Vol. 1: Theories, constructs, and critical issues* (pp. 427–458). Washington, DC: American Psychological Association.

Massey, D. (1994). America's apartheid and the urban underclass. *Social Service Review, 68*, 471–487.

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

May, H. (2006). A multilevel Bayesian item response theory method for scaling socioeconomic status in international studies of education. *Journal of Educational and Behavioral Statistics, 31*, 63–79.

May, P. A., & Gossage, J. P. (2001). Estimating the prevalence of fetal alcohol syndrome. *Alcohol Research & Health, 25*, 159–167.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational*

Psychologist, 32, 1–19.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of,multiple intellectual abilities over the life span. *Developmental Psychology, 38*, 115–142.

McGue, M., Bouchard, T. J., Iacono, W. G., & Lykken, D. T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin & G. E. McClearn (Eds.), *Nature, nurture, and psychology* (pp. 59–76). Washington, DC: APA.

McLeod, J., & Shanahan, M. (1993). Poverty, parenting, and children's mental health. *American Sociological Review, 58*, 351–366.

McLoyd, V. C. (1990). The impact of economic hardship on black families and children: Psychological distress, parenting, and socioemotional development. *Child Development, 61*, 311–346.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*, 185–204.

McLoyd, V. C., Ceballo, R., & Mangelsdorf, S. (1996). The effects of poverty on children's socioemotional development. In J. Noshpitz (Series Ed.) & N. Alessi (Vol. Ed.), *Handbook of child and adolescent psychiatry: Vol. 4. Varieties of development* (pp. 189–206). New York: Wiley.

Mercy, J. A., & Steelman, L. C. (1982). Familial influence on the intellectual attainment of children. *American Sociological Review, 47*, 532–542.

Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika, 57*, 289–311.

Metallidou, P., & Vlachou, A. (2010). Children's self-regulated learning profile in language and mathematics: The role of task value beliefs. *Psychology in the Schools, 47*, 776–788.

Miller, S. A., Manhal, M., & Mee, L. L. (1991). Parental beliefs, parental accuracy, and children's cognitive performance: A search for causal relations. *Developmental Psychology, 27*, 267–276.

Ministerium für Generationen, Familie, Frauen und Integration des Landes Nordrhein-Westfalen (2010). *Zuwanderungsstatistik Nordrhein-Westfalen 2009*. Düsseldorf: MGFFI.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.). (2008a). *Lehrplan Sachunterricht für die Grundschulen des Landes Nordrhein-Westfalen*. Düsseldorf: Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.). (2008b). *Kernlehrplan für das Gymnasium – Sekundarstufe I in Nordrhein-Westfalen Physik*. Frechen: Ritterbach Verlag.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.). (2011). *Kernlehrplan für die Hauptschule in Nordrhein-Westfalen. Lernbereich Naturwissenschaften Biologie, Chemie, Physik*. Düsseldorf: Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*, 424–436.

Mitchell, M. A., & Maxwell, S. E. (2013). A comparison of the cross-sectional and sequential designs when assessing longitudinal mediation. *Multivariate Behavioral Research, 48*, 301–339.

Möller, J., & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau, 55*, 19–27.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*, 1129–1167.

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal, 48*, 1315–1346.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683–729.

Mueller, C. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development, 52*, 13–30.

Muijs, R. D. (1997). Predictors of academic achievement and academic self-concept: A longitudinal perspective. *British Journal of Educational Psychology, 67*, 263–277.

Murphy, P. K., & Alexander, P. A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *Journal of Experimental Education, 70*, 197–214.

Muthén, B. O., Kao, C.-F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement, 28*, 1–22.

Muthén, L. K., & Muthén, B. O. (2011). *Mplus User's Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Myers, D., Baer, W., & Choi, S. (1996). The changing problem of overcrowded housing. *Journal of the American Planning Association, 62*, 66–84.

Nagy, G., Watt, H. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: Different countries, different trajectories? *Journal of Research on Adolescence, 20*, 482–506.

Neisser, U., Boodoo, G., Bouchard, T., Boykin, A., Brody, N., Ceci, S., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–101.

Nelson, K. (1993). Events, narratives, memory: What develops? In C. A. Nelson (Ed.), *Minnesota Symposia on Child Psychology: Vol. 26. Memory and affect* (pp. 1–24). Hillsdale, NJ: Erlbaum.

Nepomnyaschy, L. (2009). Socioeconomic gradients in infant health across race and ethnicity. *Maternal and Child Health Journal, 13*, 720–731.

Nering, M. L., & Ostini, R. (2010). *Handbook of polytomous item response theory models*. New York: Routledge.

Neubauer, A. C. (1997). The mental speed approach to the assessment of intelligence. In J. Kingma & W. Tomic (Eds.), *Advances in cognition and educational practice: Reflections on the concept of intelligence* (pp. 149–174). Greenwich, CT: JAI Press.

Newson, J., & Newson, E. (1976). *Seven-year-olds in the home environment*. New York: Wiley.

Newson, J., & Newson, E. (1977). *Perspectives on school at seven years old*. London: Allen & Unwin.

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology, 58*, 367–408.

Niklas, F., Segerer, R., Schmiedeler, S., & Schneider, W. (2012). Findet sich ein „Matthäus-Effekt“ in der Kompetenzentwicklung von jungen Kindern mit oder ohne Migrationshintergrund? *Frühe Bildung, 1*, 26–33.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., et al. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*, 130–159.

Noble, K. G., Houston, S. M., Kan, E., & Sowell, E. R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Developmental Science, 15*, 516–527.

Nölle, I., Hörstermann, T., Krolak-Schwerdt, S., & Gräsel, C. (2009). Relevante diagnostische Informationen bei der Übergangsempfehlung – die Perspektive der Lehrkräfte. *Unterrichtswissenschaft, 37*, 294–310.

Northstone, K., & Emmett, P. (2005). Multivariate analysis of diet in children at four and seven years of age and associations with socio-demographic characteristics. *European Journal of Clinical Nutrition, 59*, 751–760.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Ohle, A. (2010). *Primary school teachers' content knowledge in physics and its impact on teaching and students' achievement*. Berlin: Logos.

Ohle, A., Fischer, H. E. & Kauertz, A. (2011). Der Einfluss des physikalischen Fachwissens von Primarstufenlehrkräften auf Unterrichtsgestaltung und Schülerleistung. *Zeitschrift für Didaktik der Naturwissenschaften, 17*, 357-388.

Opdenakker, M., & van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal, 27*, 407–432.

Opdenakker, M. C., van Damme, J., de Fraine, B., van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *School Effectiveness and School Improvement, 13*, 399–427.

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching, 20*, 825–838.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*, 1049–1079.

O'Toole, B. I., & Stankov, L. (1992). Ultimate validity of psychological tests. *Personality and Individual Differences, 13*, 699–716.

Paarlberg, K. M., Vingerhoets, A. J. J. M., Passchier, J., Heinen, A. G. J. J., Dekker, G. A., & van Geijn, H. P. (1999). Smoking status in pregnancy is associated with daily stressors and low well-being. *Psychology and Health, 14*, 87–96.

Paek, I. (2002). *Investigation of differential item functioning: Comparisons among approaches, and extension to a multidimensional context*. Unpublished doctoral dissertation, University of California, Berkeley.

Paek, I., & Wilson, M. (2011). Formulating the Rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel-Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement, 71*, 1023–1046.

Parcel, T. L., Dufur, M. J., & Zito, R. C. (2010). Capital at home and at school: A review and synthesis. *Journal of Marriage and Family, 72*, 828–846.

Parcel, T. L., & Menaghan, E. G. (1990). Maternal working conditions and children's verbal facility: Studying the intergenerational transmission of inequality from mothers to young children. *Social Psychology Quarterly, 53*, 132–147.

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8*, 293–316.

Pettit, G. S., Keiley, M. K., Laird, R. D., Bates, J. E., & Dodge, K. A. (2007). Predicting the Developmental Course of mother-reported monitoring across childhood and adolescence from early proactive parenting, child temperament, and parents' worries. *Journal of Family Psychology, 21*, 206–217.

Pezdek, K., Berry, T., & Renno, P. A. (2002). Children's mathematics achievement: The role of parents' perceptions and their involvement in homework. *Journal of Educational Psychology, 94*, 771–777.

Pietsch, M., & Stubbe, T. C. (2007). Inequality in the transition from primary to secondary school: School choices and educational disparities in Germany. *European Educational Research Journal, 6*, 424–445.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667–686.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*, 167–199.

Pinxten, M., de Fraine, B., van Damme, J., & d'Haenens, E. (2010). Causal ordering of academic self-concept and achievement: Effects of type of achievement measure. *British Journal of Educational Psychology, 80*, 689–709.

PISA-Konsortium Deutschland (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.

Plomin, R. (1986). *Development, genetics, and psychology*. Hillsdale, NJ: Erlbaum.

Plomin, R. & DeFries, J. C. (1985). *Origins of individual differences in infancy*. Orlando, FL: Academic Press.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral genetics* (4th ed.). New York: Worth Publishers.

Plomin, R., & Spinath, F. M. (2004). Intelligence: Genetics, genes, and genomics. *Journal of Personality and Social Psychology, 86*, 112–129.

Plucker, J. A., Robinson, N. M., Greenspon, T. S., Feldhusen, J. F., McCoach, D. B., & Subotnik, R. F. (2004). It's not how the pond makes you feel, but rather how high you can jump. *American Psychologist, 59*, 268–269.

Pohlmann, B., & Möller, J. (2009). On the benefit of dimensional comparisons. *Journal of Educational Psychology, 101*, 248–258.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology, 24*, 1–24.

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

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., et al. (2006). *PISA 2003: Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

Prenzel, M., Lankes, E., & Minsel, B. (2000). Interessenentwicklung in Kindergarten und Grundschule: Die ersten Jahren. In U. Schiefele & K.-P. Wild (Hrsg.), *Interesse und Lernmotivation: Untersuchungen zu Entwicklung, Förderung und Wirkung* (S. 11–30). Münster: Waxmann.

Quay, L. C., & Jarrett, O. S. (1986). Teachers' interactions with middle- and lower SES preschool boys and girls. *Journal of Educational Psychology, 78*, 495–498.

R Development Core Team (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rahu, K., Rahu, M., Pullmann, H., & Allik, J. (2010). Effect of birth weight, maternal education and prenatal smoking on offspring intelligence at school age. *Early Human Development, 86*, 493–497.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchial linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: Oxford Psychologists Press.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*, 207–230.

Reckase, M. D. (2009). *Multidimensional item response theory*. Springer: Dordrecht, The Netherlands.

Reeve, C. L., & Hakel, M. D. (2000). Toward an understanding of adult intellectual development: Investigating within-individual convergence of interest and knowledge profiles. *Journal of Applied Psychology, 85*, 897–908.

Renkl, A. (1996). Vorwissen und Schulleistung. In J. Möller & O. Köller (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S.175–190). Weinheim: Beltz PVU.

Renkl, A. (2008). Lehren und Lernen im Kontext der Schule. In A. Renkl (Hrsg.), *Lehrbuch Pädagogische Psychologie* (S. 109–153). Bern, Schweiz: Huber.

Renkl, A. (2009). Wissenserwerb. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 3–26). Berlin: Springer.

Renkl, A. (2011). Aktives Lernen: Von sinnvollen und weniger sinnvollen theoretischen Perspektiven zu einem schillernden Konstrukt. *Unterrichtswissenschaft, 39*, 197–212.

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist, 31*, 115–121.

Renkl, A., & Stern, E. (1994). Die Bedeutung von kognitiven Eingangsvoraussetzungen und schulischen Lerngelegenheiten für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für Pädagogische Psychologie, 8*, 27–39.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation* (pp. 373–404). San Diego, CA: Academic Press.

Renninger, K. A., Ewen, L., & Lasher, A. K. (2002). Individual interest as context in expository text and mathematical word problems. *Learning and Instruction, 12*, 467–491.

Rindermann, H. (2011). Intelligenzwachstum in Kindheit und Jugend. *Psychologie in Erziehung und Unterricht, 58*, 210–224.

Rindermann, H., & Neubauer, A. C. (2004). Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equations models. *Intelligence, 32*, 573–589.

Robitzsch, A. (2009). Methodische Herausforderungen bei der Kalibrierung von Leistungstests. In A. Bremerich-Vos, D. Granzer & O. Köller (Hrsg.), *Bildungsstandards Deutsch und Mathematik* (S. 42–106). Weinheim: Beltz.

Robitzsch, A., Dörfler, T., Pfost, M., & Artelt, C. (2011). Die Bedeutung der Itemauswahl und der Modellwahl für die längsschnittliche Erfassung von Kompetenzen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 43*, 213–227

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*, 83–92.

Rose, N., Pohl, S., Böhme, H., & Steyer, R. (2010). Strukturgleichungsmodelle. In H. Holling & B. Schmitz (Hrsg.), *Handbuch der Psychologie Bd. 13: Handbuch Statistik, Methoden und Evaluation* (S. 600–611). Göttingen: Hogrefe.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley & Sons.

Rupp, A. A., & Zumbo, B. D. (2003). Which model is best? Robustness properties to justify model choice among unidimensional IRT Models under item parameter drift. *The Alberta Journal of Educational Research, 49*, 264–276.

Russell, T., Harlen, W., & Watt, D. (1989). Children's ideas about evaporation. *International Journal of Science Education, 11*, 566–576.

Rutter, M. (1981). Protective factors in children's responses to stress and disadvantage. In M. Kent & J. Rolf (Eds.), *Prevention of psychopathology* (Vol. 3, pp. 49–74). Hanover, NH: University Press.

Ryan, R. M., & Deci, E. L. (2009). Promoting self-determined school engagement: Motivation, learning, and well-being. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 171–196). New York: Taylor Francis.

Ryle, G. (1949). *The concept of mind*. Oxford, UK: Barnes & Noble.

Sandoval, J. (1995). Teaching in subject matter areas: Science. *Annual Review of Psychology, 46*, 355–374.

Scarr, S (1981.) *Race, social class, and individual differences in IQ*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Scarr S., & Weinberg, R. A. (1978). The influence of "family background" on intellectual attainment. *American Sociological Review, 43*, 674–692.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545–571.

Schiefele, H., Krapp, A., Prenzel, M., Heiland, A., & Kasten, H. (1983). *Principles of an educational theory of interest*. Paper presented at the 7th Meeting of the International Society for the Study of Behavioral Development, Munich, Germany.

Schiefele, U. (1990). The influence of topic interest, prior knowledge and cognitive capabilities on text comprehension. In J. M. Pieters, K. Breuer & P. R. J. Simons (Eds.), *Learning environments* (pp. 323–338). Berlin: Springer.

Schiefele, U. (1996). *Motivation und Lernen mit Texten*. Göttingen: Hogrefe.

Schiefele, U. (1998). Individual interest and learning. What we know and what we don't know. In L. Hoffmann, A. Krapp, K. A. Renninger & J. Baumert (Eds.), *Interest and learning. Proceedings of the Seeon-Conference on interest and gender* (pp. 91–104). Kiel, Germany, IPN.

Schiefele, U. (2001). The role of interest in motivation and learning. In J. M. Collis & S. Messick (Eds.), *Intelligence and personality: Bridging the gap in theory and measurement* (pp. 163–194). Mahwah, NJ: Erlbaum.

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 197–223). New York: Taylor Francis.

Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences, 8*, 141–160.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183–212). Hillsdale, NJ: Erlbaum.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15*, 187–211.

Schmiedek, F., & Lindenberger, U. (2007). *Methodische Grundlagen*. In J. Brandstädter & U. Lindenberger (Hrsg.), Entwicklungspsychologie der Lebensspanne (S. 67–96). Stuttgart: Kohlhammer.

Schnabel, K., Alfeld, C., Eccles, J., Köller, O., & Baumert, J. (2002). Parental influence on students' educational choices in the United States and Germany: Different ramifications – same effect? *Journal of Vocational Behavior, 60*, 178–198.

Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung. *Zeitschrift für Erziehungswissenschaft, 14*, 371–396.

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology, 81*, 306–312.

Schnyder, I., Niggli, A., Cathomas, R., Trautwein, U., & Lüdtke, O. (2006). Wer lange lernt,

lernt noch lange nicht viel mehr: Korrelate der Hausaufgabenzeit im Fach Französisch und Effekte auf die Leistungsentwicklung. *Psychologie in Erziehung und Unterricht, 53*, 107–121.

Scott, P. H., Asoko, H., & Driver, R. (1992). Teaching for conceptual change: A review of strategies. In R. Duit, F. Goldberg & H. Niedderer (Eds.), *Research in physics learning: Theoretical issues and empirical studies. Proceedings of an International Workshop in Bremen* (pp. 310–329). Kiel: IPN.

Scott, L. A., Rock, D. A., Pollack, J. M., Ingels, S. J., & Quinn, P. (1995). *Two years later: Cognitive gains and school transitions of NELS:88 eighth graders*. Washington: National Center for Education Statistics.

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish–little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology, 101*, 403–419.

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

Sherman, A. (1994). *Wasting America's future*. Boston: Beacon Press.

Shieh, Y.-Y., & Fouladi, R. T. (2003). The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. *Educational and Psychological Measurement, 63*, 951–985.

Shonkoff, J. P., & Phillips D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academies Press.

Silbereisen, R., & Todt, E. (Eds.). (1994). *Adolescence in context: The interplay of family, school, peers, and work in adjustment*. New York: Springer.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *Journal of Educational Research, 95*, 323–332.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417–453.

Skaalvik, E. M. (1997). Issues in research on self-concept. In M. Maehr & P. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 51–97). Greenwich, CT: JAI Press.

Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology, 58*, 292–307.

Skaalvik, E. M., & Valas, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education, 67*, 135–149.

Skinner, E. A. (1995). *Perceived control, motivation, and coping*. Thousand Oaks, CA: Sage.

Smith, J. R., Brooks-Gunn, J., & Klebanov, P. (1997). Consequences of living in poverty for young children's cognitive and verbal ability and early school achievement. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 132–189). New York: Russell Sage Foundation.

Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement, 1*, 199–218.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Snow, K. (1990). Building memories: The ontogeny of autobiography. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 213–242). Chicago: University of Chicago Press.

Soric, I., & Palekcic, M. (2009). The role of students' interests in self-regulated learning: The relationship between students' interests, learning strategies and causal attributions. *European Journal of Psychology of Education, 24*, 545–565.

Spieß, M. (2009). *Missing-Data Techniken*. Münster: LIT-Verlag.

Spinath, F. M., Ronald, A., Harlaar, N., Price, T. S., & Plomin, R. (2003). Phenotypic "g" early in life: On the etiology of general cognitive ability in a large population sample of twin children aged 2 to 4 years. *Intelligence, 31*, 195–210.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–407.

Stelzl, I. (2005). *Fehler und Fallen der Statistik für Psychologen, Pädagogen und Sozialwissenschaftler*. Münster: Waxmann (Reprint der Ausgabe von 1982, Bern: Huber).

Stern, W. (1911). *Die differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Barth.

Sternberg, R. J., & Grigorenko, E. L. (Eds.). (2002). *The general factor of intelligence: How general is it?* Mahwah, NJ: Erlbaum.

Stevenson, D. L., & Baker, D. P. (1987). The family-school relation and the child's school performance. *Child Development, 58*, 1348–1357.

Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap*. New York: Simon & Schuster.

Stipek, D. & MacIver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60*, 521–538.

Stocké, V. (2009). Adaptivität oder Konformität? Die Bedeutung der Bezugsgruppe und der Leistungsrealität der Kinder für die Entwicklung elterlicher Bildungsaspirationen am Ende der Grundschulzeit. *Zeitschrift für Erziehungswissenschaft, 12*, 257–281.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*, 401–426.

Strunk, U. (1999). *Die Behandlung von Phänomenen aus der unbelebten Natur im Sachun-*

terricht: Die Perspektive der Förderung des Erwerbs von kognitiven und konzeptuellen Fähigkeiten. Bad Iburg: Der Andere Verlag.

Stubbe, T. C., & Bos, W. (2008). Schullaufbahnempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik, 22*, 49–63.

Stubbe, T. C., Tarelli, I., & Wendt, H. (2012). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 231–246). Münster: Waxmann.

Swygert, K. A., McLeod, L. D., & Thissen, D. (2001). Item response theory applied to combinations of multiple-choice and constructed response items - scale scores for patterns of sum scores. In D. Thissen & H. Wainer (Hrsg.), *Test scoring* (pp. 217-250). Hillsdale, NJ: Erlbaum.

Tate, W. F. (1995). School mathematics and African American students: Thinking seriously about opportunity-to-learn standards. *Educational Administration Quarterly, 31*, 24–48.

Taub, G. E., Floyd, R. G., Keith, T. Z., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*, 187–198.

Thissen, D. (2009). On interpreting the parameters for any item response model. *Measurement: Interdisciplinary Research and Perspectives, 7*, 106–110.

Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group mean differences: The concept of item bias. *Psychological Bulletin, 99*, 118–128.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Lawrence Erlbaum.

Tiedemann, J. & Billmann-Mahecha, E. (2010). Wie erfolgreich sind Gymnasiasten ohne Gymnasialempfehlung? Die Kluft zwischen Schullaufbahnempfehlung und Schulformwahl der Eltern. *Zeitschrift für Erziehungswissenschaft, 13*, 649–660.

Todt, E. (1990). Entwicklung des Interesses. In H. Hetzer (Ed.), *Angewandte Entwicklungspsychologie des Kindes- und Jugendalters* (S. 213–264). Heidelberg: Quelle & Meyer.

Trautwein, U., Köller, O., Schmitz, B., & Baumert, J. (2002). Do homework assignments enhance achievement? A multilevel analysis of 7th grade mathematics. *Contemporary Educational Psychology, 27*, 26–50.

Trautwein, U., Lüdtke, O., Kastens, C., & Köller, O. (2006). Effort on homework in grades 5-9: Development, motivational antecedents, and the association with effort on classwork. *Child Development, 77*, 1094–1111.

Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology, 90*, 334–349.

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*, 788–806.

Treiman, D. J. (1977). *Occupational prestige in comparative perspective*. New York: Academic Press.

Tröbst, S. (2010, October). *Violations of local independence: A cause for concern?* Paper presented at the 1st NWU Rasch Workshop, Essen.

Turner, R., & Avison, W. R. (2003). Status variations in stress exposure: Implications for the interpretation of research on race, socioeconomic status, and gender. *Journal of Health and Social Behavior, 44*, 488–505.

Turner, S. L., Steward, J. C., & Lapan, R. T. (2004). Family factors associated with sixth-grade adolescents' math and science career interests. *Career Development Quarterly, 53*, 41–52.

Twenge, J. M., & Campbell, W. K. (2002). Self-esteem and socioeconomic status: A meta-analytic review. *Personality and Social Psychology Review, 6*, 59–71.

Tytler, R. (2000). A comparison of year 1 and year 6 students' conceptions of evaporation and condensation: Dimensions of conceptual progression. *International Journal of Science Education, 22*, 447–467.

Tytler, R., Peterson, S., & Prain, V. (2006). Picturing evaporation: Learning science literacy through a particle representation. *Teaching Science, 52*, 12–17.

Uebersax, J. S. (1987). Diversity of decision-making models and the measurement of interrater agreement. *Psychological Bulletin, 101*, 140–146.

Urhahne, D., Chao, S., Florineth, M., Luttenberger, S., & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student. *British Journal of Educational Psychology, 81*, 161–177.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A systematic review. *Educational Psychologist, 39*, 111–133.

Valtin, R. & Wagner, C. (2004). Der Übergang in die Sekundarstufe I: Psychische Kosten der externen Leistungsdifferenzierung. *Psychologie in Erziehung und Unterricht, 51*, 52–68.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research, 16*, 219–242.

van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation, 76*, 1049–1064.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*, 1–67.

van Buuren, S., & Oudshoorn, K. (1999). *Flexible imputation by MICE* (TNO Report PG/VGZ/99.054). Leiden, The Netherlands: TNO Prevention and Health.

van Ophuysen, S., & Wendt, H. (2009). Zur Veränderung der Mathematikleistung von Klasse 4 bis 6. Welchen Einfluss haben Kompositions- und Unterrichtsmerkmale? *Zeitschrift für Erziehungswissenschaft, 12*, 302–327.

Vock, M., Preckel, F., & Holling, H. (2011). Mental abilities and school achievement: A test of a mediation hypothesis. *Intelligence, 39*, 357–369.

von Stumm, S. (2012). You are what you eat? Meal type, socio-economic status and cognitive ability in childhood. *Intelligence, 40*, 576–583.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45–69.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.

Vosniadou, S., & Ioannides, C. (1998). From conceptual development to science education: A psychological point of view. *International Journal of Science Education, 20*, 1213–1230.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. Learning and Instruction, 15, 317–419.

Wagner, W., Helmke, A., & Schrader, F.-W. (2009). Die Rekonstruktion der Übergangsempfehlung für die Sekundarstufe I und der Wahl des Bildungsgangs auf der Basis des Migrationsstatus, der sozialen Herkunft, der Schulleistung und schulklassenspezifischer Merkmale. *Zeitschrift für Erziehungswissenschaft, 12*, 183–204.

Wainer, H. (2010a). Fourteen conversations about three things. Journal of Educational and *Behavioral Statistics, 35*, 5–25.

Wainer, H. (2010b). Schrödinger's cat and the conception of probability in item response theory. *Chance, 23*(1), 53–56.

Walberg, H. J., & Tsai, S. (1983). Matthew effects in education. *American Educational Research Journal, 20*, 359–373.

Walter, O. (2008). Herkunftsassoziierte Disparitäten im Lesen, der Mathematik und den Naturwissenschaften: Ein Vergleich zwischen PISA 2000, PISA 2003 und PISA 2006. *Zeitschrift für Erziehungswissenschaft, 10*(Sonderheft 10), 149–168.

Walter, O., Senkbeil, M., Rost, J., Carstensen, C. H., & Prenzel, M. (2006). Die Entwicklung der naturwissenschaftlichen Kompetenz von der neunten zur zehnten Klassenstufe: Deskriptive Befunde. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 87–118). Münster: Waxmann.

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 177–210). New York: Macmillan.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*, 427–450.

Watermann, R., & Baumert, J. (2006). Entwicklung eines Strukturmodells zum Zusammenhang zwischen sozialer Herkunft und fachlichen und überfachlichen Kompetenzen: Befunde national und international vergleichender Analysen. In J. Baumert, P. Stanat & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen* (S. 11–29). Wiesbaden: VS Verlag für Sozialwissenschaften.

Watermann, R., Klingebiel, F., & Kurtz, T. (2010). Die motivationale Bewältigung des Grundschulübergangs aus Schüler- und Elternsicht. In K. Maaz, J. Baumert, C. Gresch & N. McElvany (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (S. 355–384). Bonn, Berlin: BMBF.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*, 548–573.

Weiner, B. (2005). Motivation from an attribution perspective and the social psychology of perceived competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 73–84). New York: Guilford Press.

Weinert, F. E., & Hany, E. A. (2003). The stability of individual differences in intellectual development: Empirical evidence, theoretical problems, and new research questions. In R. J. Sternberg & J. Lautrey (Eds.), *Models of intelligence: International perspectives* (pp. 169–181). Washington, DC: APA.

Weinert, F. E., & Helmke, A. (1995). Interclassroom differences in instructional quality and interindividual differences in cognitive development. *Educational Psychologist, 30*, 15–20.

Weinert, F. E., & Helmke, A. (1998). The neglected role of individual differences in theoretical models of cognitive development. *Learning and Instruction, 8*, 309–323.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 (CFT 20-R)*. Göttingen: Hogrefe.

Wendt, H., Tarelli, I., Bos, W., Frey, K., & Vennemann, M. (2012). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2011). In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 27–68). Münster: Waxmann.

Wentzel, K. R. (1998). Parents' aspirations for children's educational attainments: Relations to parental beliefs and social address variables. *Merrill-Palmer Quarterly, 44*, 20–37.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461–481.

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*, 1–35.

Wigfield, A., & Eccles, J. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265–310.

Wigfield, A., Eccles, J. S., Mac Iver, D. J., Reuman, D., & Midgley, C. (1991). Transitions at early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*, 552–565.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). In W. Damon, R. Lerner & N. Eisenberg (Eds.), *Handbook of child psychology* (pp. 933–1002). New York: Wiley.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A., Freedman-Doan, C., et al. (1997). Changes in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*, 451–469.

Wigfield, A., & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in achievement solutions. *Educational Psychologist, 26*, 233–261.

Wilson, M. R. (2005). *Constructing measures*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M. R., De Boeck, P., & Carstensen, C. H. (2008). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 91–120). Göttingen: Hogrefe.

Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology, 61*, 653–678.

Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.

Wodzinski, R. (1996). *Untersuchungen von Lernprozessen beim Lernen Newtonscher Dynamik im Anfangsunterricht*. Münster: LIT.

Wong, M. (1980). Model students? Teachers perceptions and expectations of their Asian and white students. *Sociology of Education, 53*, 236–246.

Wright, B. D. (1984). Despair and hope for educational measurement. *Contemporary Education Review, 3*, 281–288.

Wright, B. D., Linacre, M., Gustafsson, J.-E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.

Wright, B. D., & Masters, G. N. (1981). *The measurement of knowledge and attitude* (Research Memorandum No. 30). Chicago: University of Chicago, MESA Psychometric Laboratory.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.

Wright, B. D., & Stone, M. H. (1999). *Measurement essentials*. Wilmington, DE: Wide Range, Inc.

Wu, M., Adams, R., Wilson, M. R., & Haldane, S. A. (2007). *ConQuest 2.0* [User Manual]. Melbourne, Australia: Australian Council for Educational Research.

Yang, Y., & Gustafsson, J.-E. (2004). Measuring socioeconomic status at individual and collective levels. *Educational Research and Evaluation, 10*, 259–288.

Zelinski, E. M., Kennison, R. F., Watts, A., & Lewis, K. L. (2009). Convergence between cross-sectional and longitudinal studies: Cohort matters. In H. B. Bosworth & C. Hertzog (Eds.), *Aging and cognition: Research methodologies and empirical advances* (pp. 101–118). Washington, DC: APA.

Zumbo, B. D. (2007a). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*, 223–233.

Zumbo, B. D. (2007b). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26: Psychometrics* (pp. 45–79). Amsterdam, The Netherlands: Elsevier Science.

Zumbo, B. D., & Rupp, A. A. (2004). Responsible modeling of measurement data for appropriate inferences: Important advances in reliability and validity theory. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 73–92). Thousand Oaks, CA: Sage.

CURRICULUM VITAE