International Journal of Science Education, 2015
Vol. 37, No. 9, 1343–1363, http://dx.doi.org/10.1080/09500693.2015.1035358

Predicting Students' Skills in the Context of Scientific Inquiry with Cognitive, Motivational, and Sociodemographic Variables

Andreas Nehring[a*†], Kathrin H. Nowak[b], Annette Upmeier zu Belzen[b] and Rüdiger Tiemann[c]
[a]*Department of Educational Sciences, University of Potsdam, Potsdam, Germany;* [b]*Biology Education, Humboldt-Universität zu Berlin, Berlin, Germany;* [c]*Chemistry Education, Humboldt-Universität zu Berlin, Berlin, Germany*

Research on predictors of achievement in science is often targeted on more traditional content-based assessments and single student characteristics. At the same time, the development of skills in the field of scientific inquiry constitutes a focal point of interest for science education. Against this background, the purpose of this study was to investigate to which extent multiple student characteristics contribute to skills of scientific inquiry. Based on a theoretical framework describing nine epistemological acts, we constructed and administered a multiple-choice test that assesses these skills in lower and upper secondary school level ($n = 780$). The test items contained problem-solving situations that occur during chemical investigations in school and had to be solved by choosing an appropriate inquiry procedure. We collected further data on 12 cognitive, motivational, and sociodemographic variables such as conceptual knowledge, enjoyment of chemistry, or language spoken at home. Plausible values were drawn to quantify students' inquiry skills. The results show that students' characteristics predict their inquiry skills to a large extent (55%), whereas 9 out of 12 variables contribute significantly on a multivariate level. The influence of sociodemographic traits such as gender or the social background becomes non-significant after controlling for cognitive and motivational variables. Furthermore, the performance advance of students from upper secondary school level can be explained by controlling for cognitive covariates. We discuss our findings with regard to curricular aspects and raise the question whether the inquiry skills can be considered as an autonomous trait in science education research.

Keywords: *Scientific inquiry; Prediction of achievement in science; Assessment; Secondary school level; Plausible values*

*Corresponding author. Department of Educational Sciences, University of Potsdam, Karl-Liebknecht-Str. 24-25, Potsdam 14476, Germany. Email: andreas.nehring@uni-potsdam.de
†Current address: Institute for Science Education (IDN), Leibniz Universität Hannover, Am Kleinen Felde 30, 30167 Hannover, Germany.

Introduction

Over the past years, scientific inquiry has been a focal point of science education efforts. Since 'science' can be described as 'a body of knowledge and the activities that gave rise to that knowledge' (Zimmermann, 2000, p. 101), the term 'scientific inquiry' relates to the procedures scientists use to gain new knowledge. In the context of science education, it is part of an active and meaningful learning process and combines general science process skills with science content, creativity, and critical thinking to develop scientific knowledge (Lederman et al., 2014). According to Lederman (2004, p. 309), scientific inquiry refers to 'the systematic approaches used by scientists in an effort to answer their question of interest'. With regard to school-related purposes, 'it is useful to conceptualize scientific inquiry as the process by which scientific knowledge is developed' (Lederman, 2004, p. 308). Gott and Duggan (1995, p. 98) state the following: 'As technology advances and more "critical thinkers" are required, science education has a duty to enable students to examine the quality of scientific evidence effectively'. Against this background, scientific inquiry constitutes one of the key concepts in many science education standards (Bernholt, Neumann, & Nentwig, 2012; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005; National Research Council, 2006).

Nevertheless, research shows that students have difficulties in developing scientific inquiry skills and an adequate understanding about the nature of scientific inquiry (Arnold, Kremer, & Mayer, 2014; Chen & Klahr, 1999; Duggan & Gott, 2000; Kuhn, Black, Keselman, & Kaplan, 2000; Lubben, Campbell, Buffler, & Allie, 2001; Schauble, Glaser, Duschl, Schulze, & John, 1995). At the same time, we know from research that learning is a more complex process, which involves cognitive and motivational factors (Alao & Guthrie, 2010; Bransford, Brown, & Cocking, 2000). Thus, different personal traits of the students can contribute to a higher manifestation of scientific inquiry skills and general achievements in school science (Dochy, Segers, & Buehl, 1999; Maerten-Rivera, Myers, Lee, & Penfield, 2010; Shapiro, 2004; Tobias, 1994). Against this background, teachers do not only have to foster inquiry skills by supporting students directly in the learning process (Arnold et al., 2014; Crawford, 2007), they also have to provide rich learning environments that subserve multiple traits of students, which can affect science learning (Bybee & McCrae, 2009). Science education research can contribute to this development by examining which traits predict achievement in science and scientific inquiry skills in particular. Therefore, research includes predictors from cognitive, motivational, and sociodemographic background. The following variables are counted among these predictors:

- prior or conceptual knowledge (Alao & Guthrie, 2010; Hewson & Hewson, 1983; O'Reilly & McNamara, 2007),
- intelligence (Baumert, Lüdtke, Trautwein, & Brunner, 2009; Walpuski, Ropohl, & Sumfleth, 2011),
- cognitive load (Cook, 2006; Paas, Renkl, & Sweller, 2003; Sweller, 1994),

- reading skills (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010; Maerten-Rivera et al., 2010; O'Reilly & McNamara, 2007; Pyburn & Pazicni, 2014; Walpuski et al., 2011),
- interest and motivation (Chamers & Thomas, 1997; Häussler & Hoffmann, 2000; Krapp & Prenzel, 2011; Osborne, Simon, & Collins, 2003),
- self-concept (Häussler & Hoffmann, 2000; Krapp & Prenzel, 2011),
- gender (Chamers & Thomas, 1997; Maerten-Rivera et al., 2010; Osborne et al., 2003; Reid & Skryabina, 2003; Stark & Gray, 2010),
- language spoken at home (van Laere, Aesaert, & van Braak, 2014; Maerten-Rivera et al., 2010) and
- social background (Evans, Kelley, Sikora, & Treiman, 2010; Organisation for Economic Co-operation and Development [OECD], 2006).

Alao and Guthrie (2010) and Maerten-Rivera et al. (2010) argue that research on predictors of science achievement focuses on one or several single factors. As these factors can be interrelated, it is important to examine the unique and joint effects they possibly have on science achievement. Therefore, it is desirable to investigate which factors have an impact on students' skills after controlling for other possible factors. In addition, more traditional content-based assessments are used to investigate possible impacts. Hence, we identified a research gap in the scientific education research area focusing on multiple influences of students' characteristics on their scientific inquiry skills. Filling this gap is the purpose of this study.

Theoretical Framework

Conceptualizing Scientific Inquiry

According to Anderson (2002), the uses of the term 'inquiry' can be divided into the aspects 'scientific inquiry', 'inquiry learning', and 'inquiry teaching'. Whereas scientific inquiry 'refers to the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work' (National Research Council, 2006, p. 23), inquiry learning 'refers to a learning process in which students are engaged' (Anderson, 2002, p. 2). Inquiry learning emphasizes the active character of learning. Moreover, it is stated that inquiry learning should reflect the nature of scientific inquiry. Thus, epistemological aspects are included and activities in inquiry learning are oriented towards procedures that scientists use to gain new knowledge about the natural world. Accordingly, inquiry teaching constitutes a more process-oriented approach, which relies on authentic questions generated from student experiences as the 'central strategy for teaching science' (Anderson, 2002, p. 2).

This vision of teaching and learning in science classroom gives students a more active role, including the opportunity to determine if questions are appropriate for scientific investigations, to generate investigation designs or to develop a strategy for data collection and organization. Accordingly, the skills that students should acquire refer to these activities. In Germany, the acquisition of these skills has been implemented in the nation-wide educational standards for chemistry, biology, and physics (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005). At the same time, the notion of competence is used to conceptualize the intended learning outcomes (Köller & Parchmann, 2012; Weinert, 2001). Based on a pragmatic point of view, it describes the ability to solve problems arising from specific situational demands on the basis of conceptual knowledge, strategies, and motivational attitudes. With regard to these constructs, however, it is neither entirely clear to which extent these competences can be considered as an autonomous construct (Le Deist & Winterton, 2005) nor how they are structured (Koeppen, Hartig, Klieme, & Leutner, 2008). Thus, the theoretical descriptions of the skills that students develop while learning inquiry remain on a rather general level; therefore, further steps are required to operationalize them in a more detailed manner.

With this in mind, we derived a theoretical framework that provides a clear description of the structure of skills in the context of scientific inquiry (Nehring, Nowak, Upmeier zu Belzen, & Tiemann, 2012; Nowak, Nehring, Tiemann, & Upmeier zu Belzen, 2013). This framework can be applied in chemistry and biology (Nowak et al., 2013), and provides a theoretical foundation for the assessment of these skills and the investigation on relevant predictors. It contains elements of scientific reasoning and inquiry methods.

Scientific Reasoning

Klahr (2000) proposes an integrated view on scientific reasoning with the scientific discovery as dual search (SDDS) framework. This framework conceptualizes scientific discovery as a problem-solving process 'with the ultimate goal of generating, and then appraising, the tenability of a hypothesis about a causal or categorical relationship' (Zimmermann, 2000, p. 107). Thus, scientific reasoning is characterized as a guided search and information gathering task in the *hypothesis space* and the *experiment space*. While the first search procedure involves deriving a hypothesis from current conceptual knowledge, the second aims to find an investigation that is suitable to test the hypothesized predictions. Both procedures are connected together by the *evidence evaluation* process. In this manner, the formulation of hypothesis, the planning of investigations, and the evaluation of investigations are designated as fundamental elements of scientific reasoning. To conclude, we have taken three main processes into consideration in order to theorize scientific reasoning (Nehring, Nowak, Upmeier zu Belzen, & Tiemann, submitted for publication; Nowak et al., 2013):

- *Questions and hypotheses*: Identifying gaps in scientific knowledge, asking scientific questions, and formulating tentative assumptions that can be tested in scientific investigations.
- *Plan and performance*: Identifying the necessary steps and auxiliary equipment to carry out scientific investigations and collecting data relevant to the research question.
- *Analysis and reflection*: Changing the mode of data representation in order to customize the given information and subsequently to evaluate the validity of the investigation.

Inquiry Methods

Inquiry methods refer to the different approaches used to carry out scientific investigations. Inquiry methods differ in the degree of manipulation they require (Nehring et al., submitted for publication). According to Gott and Duggan (1995), these activities include designing experiments, using apparatus, and observation. Observations may involve unaffected objects, whereas experiments include interventions in the object under investigation (Dean & Kuhn, 2007). While students have to choose the appropriate way to identify traits of the phenomena under investigation while observing, they have to change independent variables purposefully to examine the influence of this change while experimenting. Confounding variables have to be controlled (Chen & Klahr, 1999; Lazonder & Egberink, 2014). By doing so, students intervene actively and modify the phenomena under investigation. Hofstein and Lunetta (2004) argue that scientific work in school and in science class is not only about interacting with original materials but also with models. But models are rather used as media instead of being used as research tools. Their role in formulating and testing hypotheses or theories is not sufficiently developed in school contexts (Trier, Krüger, & Upmeier zu Belzen, 2012). Justi (2009) lists activities that are important for using models for inquiry. These include:

> acquiring information about the entity that is being modelled (from empirical observations and/or from previous knowledge), producing a mental model of it, expressing that model in an adequate mode of representation, testing it (through mental and empirical experimentation), and evaluating its scope and limitations. (Justi, 2009, p. 32)

With regard to different amounts of manipulation and scientific work on models, we conclude that the following three inquiry methods are relevant for inquiry in science classrooms (Nehring et al., submitted for publication; Nowak et al., 2013):

- *Observing, comparing, and arranging*: Selecting criteria for observations, making theory-driven observations, collecting data using auxiliary equipment, taking two or more objects into consideration in order to identify differences or to derive classifications.
- *Experimenting*: Identifying dependent, independent, and control variables, actively intervening with research objects, manipulating variables, keeping control variables constant.
- *Using models*: Collecting data, making observations, and carrying out experiments with the help of models, drawing conclusions on the underlying original theory by using substitute objects.

Theoretical Framework for Scientific Inquiry

In order to describe, classify, and assess students' skills in scientific inquiry, we combined these theoretical elements in a framework that consists of two dimensions: *scientific reasoning* and *inquiry methods*. With reference to the SDDS framework (Klahr, 2000), the first dimension covers the steps of a scientific investigation *question and*

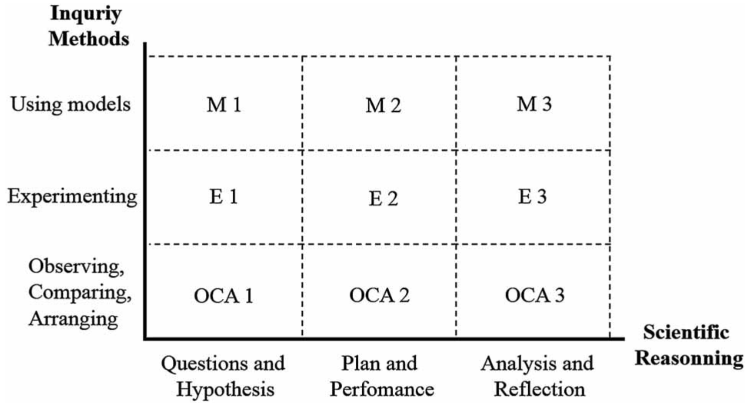

Figure 1. Theoretical framework of students' skills in the field of scientific inquiry (Nehring et al., 2012; Nowak et al., 2013)

hypothesis, *plan and performance*, and *analysis and reflection*. The second dimension takes into account different qualities to carry out scientific investigations: *modelling*, *experimenting*, *observing*, *comparing*, and *arranging*. Combined together, these dimensions create a matrix consisting of nine cells (Figure 1). Each of the cells refers to one epistemological act, which corresponds to one activity that students have to implement for being successful in scientific inquiry.

Based on the notion of competence, this framework operationalizes these activities as problem-solving processes (PISA-Konsortium Deutschland, 2006; Kremer, Specht, Urhahne, & Mayer, 2014) and describes how students should perform while generating new knowledge on the basis of scientific investigations. As a consequence, the term 'inquiry skills', as we use it in this paper, refers to cognitive problem-solving skills in the context of scientific inquiry. Consequently, students possess scientific inquiry skills when they are able to solve problems that occur while carrying out investigations according to the epistemological acts described in the framework.

Aims and Research Questions

With this in mind, the aim of the study was to investigate which of the cognitive, motivational, and sociodemographic characteristics of students contribute to the scientific inquiry skills on the bi- and multivariate level. Therefore, we derived and assessed possible variables influencing these skills together with the skills which we describe on the basis of the theoretical framework. In accordance with the literature (see above), we included cognitive, motivational, and sociodemographic variables. The following research questions are addressed:

(1) To which extent are cognitive and motivational variables related to students' skills in scientific inquiry on the bivariate level?

(2) To which extent can skills in scientific inquiry be predicted by cognitive and motivational variables?
(3) Which variables exert a unique influence on skills in scientific inquiry after controlling for other variables?
(4) To which extent is advancement of students from upper secondary school level in scientific inquiry related to the variables taken into account?

Methods

Variables Taken into Account

In order to answer the stated research questions, we conducted a confirmatory cross-sectional study in secondary schools. Therefore, we assessed the inquiry skills in chemistry as well as the covariates we found to be relevant in the research literature (see above). As we assessed the inquiry skills in the context of chemistry, we set the focus of the variables on chemistry when necessary. The following cognitive covariates were taken into account:

- Conceptual knowledge in chemistry
- Intelligence
- Perceived cognitive load
- Reading skills
- Reading speed

In addition, we collected data concerning the following motivational covariates:

- Enjoyment of chemistry
- Interest in the chemical concepts of the scientific investigations
- Self-concept in chemistry
- Future-oriented motivation to learn chemistry

Furthermore, we included social background variables that may influence the performance of students into our analysis. In this context, large-scale assessments such as Programme for International Student Assessment (PISA) in many cases showed disparities between immigrant and non-immigrant students and students from higher and lower social status (Evans et al., 2010; OECD, 2006). Following the PISA study, we operationalized these variables by the number of books students have at home and the language spoken at home (Frey, Taskinen, et al., 2009). Due to the availability of digital information, however, we consider the book variable to be rather an indicator for the social status. Taken together, this study takes the following sociodemographic variables into account:

- Gender
- Number of books at home
- Other language than the language of instruction spoken at home

In total, we included 12 covariates in this study.

Sample

The data collection took place in 10 secondary schools in Berlin, Germany. The total sample size comprised 780 students, of which 53% were female (see Table 1). It was the decision of teachers to take part in the study. If teachers agreed, all students of a class worked on the items. It took about 90 minutes to fill out the booklets and additional questionnaires.

Instruments

We collected data-describing process skills using a multiple-choice test, which relies on the theoretical inquiry framework. Although we applied this framework in the context of both chemistry and biology (Nowak et al., 2013), we focused only on the chemistry test in order to minimize the influence that different science disciplines might exert. The test consisted of 90 inquiry items. These items contained problem-solving situations that occur during chemical investigations and have to be solved by choosing an appropriate inquiry procedure (Nehring, 2014). Ionic bonding, acid–base reaction, and properties of salts served as the chemical concepts for the items. Table 2 presents an example item.

We applied a multi-matrix design to administer all inquiry items (Frey, Hartig, & Rupp, 2009). Each student worked on one test booklet containing 27 items. Each booklet was linked to another booklet on the basis of 18 anchor items. In total, 30 test booklets were used (Nehring, 2014). Based on a one-parametric logistic Rasch model, all students were located on the same continuous logit scale although they worked on different booklets. Nehring (2014) and Nehring et al. (submitted for publication) show that the expected a posteriori estimation based upon plausible values (EAP/PV) reliability of the test is .69, which corresponds to an acceptable reliability for assessments on group level. The test-based analyses further show that students from upper secondary school level possess more advanced process skills in scientific inquiry (Nehring et al., submitted for publication). Their performance advance amounts to more than a half standard deviation. This corresponds to the values for two schooling years reported in the literature (Bloom, Hill, Black, & Lipsey, 2008; Kuhl, Siegle, & Lenski, 2013).

The covariates were assessed on the basis of tests and questionnaires that were standardized or published in the context of international large-scale assessments. We

Table 1. Composition of the sample

Secondary level	School grade	Number of students	Age mean (SD)
Lower secondary level	9	269	14.8 (0.6)
	10	343	15.7 (0.6)
Upper secondary level	11	154	16.7 (0.8)
	12	14	17.5 (0.9)

Table 2. Example item

A neutralization is a chemical reaction in which an acid and a base react. In this process, the pH becomes neutral. Moreover, the temperature of the solution can change. Jan and Laura add the same quantity of acid and base in a test tube. Thereon, they vaporize the solution using a Bunsen burner and observe what happens. Which assumption are they able to prove with the help of their observation?
Tick a box. Laura and Jan are able to investigate, whether … X … a new substance occurs • … acid and base react • … the pH becomes neutral • … the temperature rises

Note: Inquiry method: observing. Scientific reasoning: hypothesis (Nowak et al., 2013; Nehring, 2014).

developed the test for the conceptual knowledge ourselves as it was necessary to match it to the contents that are subject of the scientific investigation in the inquiry test. All questionnaires contained multiple-choice or rating scale items. Table 3 gives on an overview of the instruments used in this study.

All scales, except for the conceptual knowledge in chemistry, show satisfactory reliabilities. Here, the reliability is at $\alpha = .55$. Current developments of tests measuring conceptual knowledge show that this is not a surprising result (Adams & Wieman, 2011; Bretz & Linenberger, 2012; McClary & Bretz, 2012; Wren & Barbera, 2014). Adams and Wieman (2011) argue that conceptual knowledge may constitute a non-coherent latent construct across a multitude of students. Therefore, reliability measures that rely on a highly connected structure, such as Cronbach's α, may produce lower values despite a sufficient item quality. The reliability determined by Cronbach's α indicates, however, that measures on the group level produce reliable values (Lienert & Raatz, 1998). Thus, the test fulfilled the requirements of the study.

Data Analysis

Due to the multi-matrix design, we used uni- and multidimensional item-response models to scale the data collected in the test on inquiry skills (Hartig & Höhler, 2009). These models rely on a probabilistic connection between the ability of a student and the probability to solve an item correctly. Thus, they allow to transform the ordinal raw data into measures on interval scale. In this way, both the items and the students are located on the same continuous logit scale (Bond & Fox, 2007).

In order to quantify the skills of students, we calculated plausible values (Adams & Wu, 2002; Davier, Gonzalez, & Mislevy, 2009). Plausible values are random draws from this estimated posterior distribution of a student's ability. These values do not represent optimal point estimates for individual students as it is the case of marginal

Table 3. Overview of the instruments used in the assessment of the variables

Variable	Type	Source	Reliability
Process skills in scientific inquiry	90 × MC	Self-development	EAP/PV = .69
Conceptual knowledge in chemistry	24 × MC	Self-development	$\alpha = .55$
Intelligence	25 × MC	Heller and Perleth (2000)	$.77 < r < .91$ (test–retest reliability)
Perceived cognitive load	15 × MC	Eckhardt (2010)	$\alpha = .95$
Reading skills and reading speed	23 × MC	Schneider, Schlagmüller, and Ennemoser (2007)	$.84 < r < .87$ (test–retest reliability)
Enjoyment of chemistry	5 × rating scale	Frey, Taskinen, et al. (2009) and PISA-Konsortium Deutschland (2006)	$\alpha = .93$
Interest in the chemical concepts of the scientific investigations	17 × rating scale	Frey, Taskinen, et al. (2009) and PISA-Konsortium Deutschland (2006)	$\alpha = .94$
Self-concept in chemistry	6 × rating scale	Frey, Taskinen, et al. (2009) and PISA (2006)	$\alpha = .93$
Future-oriented motivation to learn chemistry	4 × rating scale	Frey, Taskinen, et al. (2009) and PISA (2006)	$\alpha = .93$
Gender	1 × MC	Self-development	–
Number of books at home	1 × MC	Frey, Taskinen, et al. (2009) and PISA-Konsortium Deutschland (2006)	
Other language than the language of instruction spoken at home	1 × MC	Frey, Taskinen, et al. (2009) and PISA-Konsortium Deutschland (2006)	–

maximum likelihood (MML), expected a posteriori, or Warm's weighted likelihood estimates (WLE) (Wu, 2005). Usually five plausible values are drawn for each student. They are not appropriate for use as individual student scores but lead to an unbiased estimation of the parameters of any model of the population model that is specified (Wu, Adams, Wilson, & Haldane, 2007). The use of individual estimates like MML, EAP, or WLE would lead to an over- or underestimation of the populations' variance and thus to a biased analysis of variable relations. Following Wu (2005), we drew five plausible values. If the relationship of variables were analysed, the variables under investigation would be included in the model that produces the plausible values. 'If the model that produced the plausible values did not include the background variables as regressors, then the regression coefficients produced with the plausible values will be an under-estimate of the true regression coefficients' (Wu, 2005, p. 125). In our case, we included all covariates in a latent regression, which was used to calculate students' abilities. As the logit scale of the Rasch model uses undefined numerical values, we transformed the scale following international comparative studies (OECD, 2013) and set the mean of the plausible values at 500 and the standard deviation at 100 (Nehring et al., submitted for publication).

The analysis of relationships between the variables relied on correlation and hierarchical regression analysis. As the use of plausible values leads to five data sets, all analyses were carried out five times. With regard to this procedure, Davier et al. (2009, p. 8) state: 'There is no need to rely on computational shortcuts by averaging plausible values before calculations: analytical shortcuts such as averaging plausible values produce biased estimates and should be discouraged.' Thus, we compared the results of the five analyses and averaged the values in the last step. In order to estimate the reliability of the drawing of plausible values, we adjusted the standard errors according to the procedure by Rubin (1987) as reported in Baraldi and Enders (2010). This procedure takes into account the within-variance in the data sets as well as between-variance between the data sets. The adjusted standard errors were used to test the relationships between the variables of significance. As we derived presumptions for the relationships between the variables from existing research (see above), we applied one-tailed tests on significance.

As the fourth research question addresses the comparison of two sub-samples (students from upper and lower secondary school level), we used point estimates for individual students to avoid effects of the random drawing in the smaller sub-samples. Thus, WLEs were used as person estimates.

Results

Table 4 presents an overview of the descriptive statistics. The scales cover a whole range of values indicating that there is enough variance in the data to carry out meaningful correlation analysis.

Table 5 presents the results of these correlation analyses on the bivariate level. As we used only complete data sets for the analysis, these results are based on 403 students.

The results show significant correlations for all covariates. Whereas the reading skills and speed are less correlated, the conceptual knowledge, the perceived cognitive load, and the self-concept in chemistry show the strongest relationship on the bivariate level. According to Cohen (1988), the majority of the correlations can be classified as medium or large. Furthermore, two negative correlations can be found. The correlation for the perceived cognitive load shows that high cognitive load goes hand in hand with lower inquiry skills. The correlation for gender indicates that boys perform better in inquiry than girls.

In order to answer the second and third research question, we conducted multiple linear regressions. According to the classification of our variables as cognitive, motivational, or sociodemographic, we constructed three regression models. These models were used to identify which variables influence the inquiry skills after controlling for variables coming from a common theoretical background. In the next step, we combined all variables together into one regression model. This model was used to analyse the total impact of all variables on the inquiry skills.

Table 6 presents the results for the model that includes the cognitive variables 'conceptual knowledge in chemistry', 'intelligence', 'perceived cognitive load', 'reading

Table 4. Descriptive statistics of the analysed variables

Variable	*M*	SD	Minimum	Maximum
Process skills in scientific inquiry	500	100	162	822
Conceptual knowledge in chemistry	13.1	3.9	0	23
Intelligence	10.4	1.3	7.2	13.9
Perceived cognitive load	3.3	0.9	1.0	5.0
Reading skills	50.7	11.0	25	76
Reading speed	842.5	295.3	9	1,727
Enjoyment of chemistry	2.4	0.8	1.0	4.0
Interest in the chemical concepts of the scientific investigations	2.2	0.7	1.0	4.0
Self-concept in chemistry	2.4	0.8	1.0	4.0
Future-oriented motivation to learn chemistry	1.7	0.8	1.0	4.0
	Mdn		Minimum	Maximum
Gender	2		1	2
Number of books at home	26–100		0	More than 500
Other language than the language of instruction spoken at home	0		0	1

Note: The variable 'gender' was coded dichotomously, assigning numerical value 2 for girls and 1 for boys. 'The number of books at home' (26–100) corresponds to the fourth multiple-choice answer out of six.

Table 5. Analysis of the bivariate correlations between inquiry skills and covariates

Variable	Pearson product–moment correlation	Significance
Conceptual knowledge in chemistry	.47	$p < .001$
Intelligence	.39	$p < .001$
Perceived cognitive load	−.50	$p < .001$
Reading skills	.17	$p < .001$
Reading speed	−.15	$p = .004$
Enjoyment of chemistry	.35	$p < .001$
Interest in the chemical concepts of the scientific investigations	.39	$p < .001$
Self-concept in chemistry	.48	$p < .001$
Future-oriented motivation to learn chemistry	.30	$p < .001$
Gender	−.20	$p < .001$
Number of books at home	.15	$p = .009$
Other language than the language of instruction at home	−.15	$p = .006$

Note: The variable 'gender' was coded dichotomously, assigning numerical value 2 for girls and 1 for boys.

skills', and 'reading speed'. These five variables explain 47% of the inquiry skills. All variables contribute significantly to this result. Furthermore, the standardized β-coefficients show that the perceived cognitive load ($\beta = -.37$, $p < .001$), the conceptual knowledge in chemistry ($\beta = .30$, $p < .001$), and the intelligence ($\beta = .21$, $p < .001$) constitute the strongest cognitive predictors for inquiry skills. The reading skills and speed ($\beta = .15$, $p < .001$; $\beta = -.16$, $p < .001$) contribute less to the inquiry skills. Thus, the results of the correlation analysis seem to be confirmed for the cognitive covariates.

The regression model for motivational variables is presented in Table 7. This model explains 25% of the variance in the inquiry data. The results indicate that self-concept ($\beta = .43$, $p < .001$) and interest in the chemical concepts of the scientific investigations ($\beta = .22$, $p < .001$) seem to have the strongest influence on the inquiry skills, whereas the future-oriented motivation does not affect the inquiry skills after controlling for other motivational variables ($\beta = .05$, $p = .255$). Surprisingly, the sign of the regression coefficients ($B = -17.69$) for enjoyment of chemistry is negative although the bivariate correlation was found to be positive. Pandey and Elliott (2010) argue that this effect appears when variables function as suppressor variables. This phenomena is explained and interpreted in the Discussion section.

Table 8 shows the results for the sociodemographic model. This model explains 7% of the variance in the data. All variables of this model contribute to a prediction of the inquiry skills. The standardized β-coefficients also indicate that compared to the

Table 6. Multiple regression for cognitive variables predicting inquiry skills

	B	SE	β	*p*
Constant	343.16	40.52		<.001
Conceptual knowledge in chemistry	7.37	1.00	.30	<.001
Intelligence	16.19	3.03	.21	<.001
Perceived cognitive load	−38.56	4.04	−.37	<.001
Reading skills	1.34	0.35	.15	<.001
Reading speed	−0.05	0.01	−.16	<.001
Adjusted R^2	.47			

Table 7. Multiple regression for motivational variables predicting inquiry skills

	B	SE	β	*p*
Constant	328.09	16.09		<.001
Enjoyment of chemistry	−17.69	9.88	−.15	.055
Interest in the chemical concepts of the scientific investigations	31.77	9.44	.22	.001
Self-concept in chemistry	55.17	7.93	.43	<.000
Future-oriented motivation to learn chemistry	6.13	7.95	.05	.255
Adjusted R^2	.25			

Table 8. Multiple regression for sociodemographic variables predicting inquiry skills

	B	SE	*β*	*p*
Constant	531.34	20.48		<.001
Gender	−39.22	9.73	−.20	<.001
Number of books at home	9.74	3.40	.14	.004
Other language than the language of instruction spoken at home	−26.05	10.73	−.12	.012
Adjusted R^2	.07			

previous analysis the influence of these variables here is less strong. The impact of students' gender cannot be explained by the other sociodemographic variables.

All 12 variables are included into a regression model which is shown in Table 9. Taken all together, these variables explain 54% of the variance in the inquiry data.

The analysis shows that the influence of all cognitive variables remains significantly high. However, the standardized *β*-coefficients indicate that compared to previous analysis this influence is reduced here. The impact of variables such as interest in the chemical concepts of the scientific investigations ($\beta = .24$, $p < .001$) or self-concept in chemistry ($\beta = .18$, $p < .001$) is approximately as strong as the influence of conceptual knowledge in chemistry ($\beta = .26$, $p < .001$) or perceived cognitive load ($\beta = -.28$, $p < .001$). The motivational variables seem to partially explain the influence of these cognitive variables. The impact of the future-oriented motivation to learn chemistry remains non-significant. Also, the influence of the variables number of books at home and gender can be explained by controlling for other variables. Thus, 9 of the 12 variables significantly predict the inquiry skills, while the cognitive and motivational variables considered in this study seem to have the highest impact on the inquiry skills.

Table 9. Multiple regression for cognitive, motivational, and sociodemographic variables predicting inquiry skills

	B	SE	*β*	*p*
Constant	276.30	43.43		<.001
Conceptual knowledge in chemistry	6.32	0.98	.26	<.001
Intelligence	13.42	2.89	.18	<.001
Perceived cognitive load	−30.10	4.17	−.28	<.001
Reading skills	1.33	0.33	.15	.002
Reading speed	−0.05	0.01	−.16	<.001
Enjoyment of chemistry	−23.99	7.67	−.21	.004
Interest in the chemical concepts of the scientific investigations	33.90	7.35	.24	<.001
Self-concept in chemistry	21.72	6.54	.18	.002
Future-oriented motivation to learn chemistry	7.74	6.12	.06	.142
Gender	0.97	7.27	.01	.460
Number of books at home	1.21	2.42	.02	.330
Other language than the language of instruction spoken at home	−20.17	7.57	−.10	.013
Adjusted R^2	.54			

Table 10. Results of analyses of covariance (ANCOVA) predicting the means of inquiry skills for students from lower and upper secondary school level in comparison to the observed means in the data

Inquiry skills	M_1 (lower secondary school level)	M_2 (upper secondary school level)	$M_2 - M_1$	p
Observed	486	550	64	<.001
Predicted (cognitive variables)	497	514	17	.065
Predicted (motivational variables)	491	540	51	<.001
Predicted (cognitive and motivational variables)	499	512	13	.176

In order to answer the fourth research question, we calculated WLEs as point estimates for individual students and compared the inquiry skills of students from upper and lower secondary school level. On the descriptive level, the data show that the students from lower and upper secondary school level differ by about 64 points in their inquiry skills, which is about .6 standard deviations (Table 10). These differences are significant (t (775) = −7.63, $p < .001$) and correspond to learning gains of about two school years (Bloom et al., 2008). We used ANCOVAs to investigate whether this difference can be explained on the basis of the covariates. Therefore, we included the strongest predictors for inquiry: conceptual knowledge in chemistry, intelligence, perceived cognitive load, interest in the chemical concepts of the scientific investigations, self-concept in chemistry, and enjoyment of chemistry. Following the procedures for the regression analysis, we conducted three analyses and included the cognitive, the motivational, and both types of variables in ANCOVA. Table 9 reports the observed and the predicted means for students from lower and upper secondary school levels.

The results of ANCOVA show that the differences between the students from lower and upper secondary level become non-significant after controlling for the cognitive variables (F (1,578) = 3.42, $p = .065$). The mean differences between both groups are reduced from 64 to 17 points. This effect cannot be replicated using only the affective variables. Although the difference becomes smaller on the numerical level, it remains significant (F (1,639) = 29.35, $p < .001$) and still amounts to 51 points. Finally, we took both types of variables into account. Here, the differences of the means between the two groups of students are not significant (F (1,511) = 1.84, $p < .176$) and thus can be explained by these covariates.

Conclusion

The present study reveals significant impact of students' characteristics on their inquiry skills. These skills, which are described, classified, and assessed on the basis of present theoretical framework, can be predicted by 55% on the basis of the covariates taken into account in this study. This constitutes a proportionally large amount considering that there is still variance generated by the measurement error and by possible teaching activities not considered in our data. These results suggest that students rely to a large extent on different abilities and characteristics while solving inquiry problems.

The cognitive variables in particular (e.g. conceptual knowledge in chemistry, intelligence, and perceived cognitive load) play a key role in predicting the inquiry skills. In total, the cognitive variables predict 47% of the inquiry skills. Together with motivational variables, these variables explain the influence of gender and the number of books at home, indicating that the lower performance of girls and students with lower social status can be traced back to differences in knowledge and cognitive load. From an educational point of view, this constitutes a positive result as these traits can be influenced by instructional activities in class.

However, the analysis indicates that the performance of students with migration background remains lower although characteristic variables are controlled. The spoken languages at home influence the inquiry skills although we controlled for reading skills and reading speed. This result is in line with current research that shows the importance of the home language for science achievement (van Laere et al., 2014). It is possible that another language spoken at home interacts with an understanding of classroom activities (Taboada, 2012).

Furthermore, the analysis underlines the importance of motivational variables for a successful performance in inquiry. Although, only the interest in the chemical concepts of the scientific investigations and the self-concept in chemistry seem to contribute directly to the inquiry skills, the motivational variables predict these skills up to 25%. These results support the importance of efforts to create learning environments in class conducive to a development of interest and self-concept (Häussler & Hoffmann, 2000; Häussler & Hoffmann, 2002; Osborne et al., 2003). However, the data support the assumption that it is not a general interest in the subject that matters but rather a specific interest in the content that is object of the conducted scientific investigation.

As expected, the enjoyment of chemistry is correlated positively to the inquiry skills on a bivariate level; however, the coefficient gets a negative sign in the multivariate linear regression. Pandey and Elliott (2010) describe this phenomenon as negative suppression. The negative suppressor variable absorbs irrelevant variance of the other predictor variables. According to Bortz and Schuster (2010), a negative sign of a suppressor variable does not necessarily mean that the influence is negative. The negative sign is a numerical artefact that contributes to a more accurate prediction of the independent variable. As an interpretation, we suggest that the function of the enjoyment of chemistry is to reinforce students' self-concept and increase their interest in the contents of the scientific investigation. Thus, enjoyment of chemistry can help to perform better in inquiry by strengthening other motivational variables. Here, further analysis using structural equation modelling could be helpful.

Finally, this study shows that students from upper secondary school level perform better in inquiry and that this advance corresponds to the values reported in the literature. However, this advance can be traced back to the cognitive covariates included in our study. Students from upper secondary school level perform better in inquiry as their conceptual knowledge, intelligence, and capacity to deal with complex information are more elaborated. Against this background, we raise the question if and to which extent the skills described in science education standards (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005) can be considered as an autonomous trait.

To answer this question, two explanatory approaches can be identified. First, it is conceivable that the inquiry skills themselves constitute an amalgam of knowledge, intelligence, and motivation as suggested by Weinert (2001) for the notion of competence. Second, it is possible that the classroom activities do not provide enough learning opportunities to develop these skills. Therefore, the covariates have a compensatory function. As the construct inquiry skills are sufficiently developed in the class, students may have to rely on further characteristics to solve inquiry problems. Here, a combination of observational and assessment studies could be fruitful to identify classroom activities that contribute to a development of inquiry skills.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the German Research Foundation (DFG) [grant nos. TI 336/10-1 and UP 12/3-1].

References

Adams, R. J., & Wu, M. L. (Eds.). (2002). *PISA 2000 technical report*. Paris: OECD.

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289–1312. doi:10.1080/09500693.2010.512369

Alao, S., & Guthrie, J. T. (2010). Predicting conceptual understanding with cognitive and motivational variables. *The Journal of Educational Research*, *92*(4), 243–254. doi:10.1080/00220679909597602

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, *13*(1), 1–12.

Arnold, J. C., Kremer, K., & Mayer, J. (2014). Understanding students' experiments—What kind of support do they need in inquiry tasks? *International Journal of Science Education*, *36*(16), 2719–2749. doi:10.1080/09500693.2014.930209

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*(1), 5–37. doi:10.1016/j.jsp.2009.10.001

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*(3), 165–176. doi:10.1016/j.edurev.2009.04.002

Bernholt, S., Neumann, K., & Nentwig, P. (Eds.). (2012). *Making it tangible: Learning outcomes in science education*. Münster: Waxmann.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289–328. doi:10.1080/19345740802400072

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* [Statistics for human and social scientists, 7th revised and extended edition] (7. vollständig überarbeitete und erweiterte Auflage). Berlin: Springer.

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academy Press.

Bretz, S. L., & Linenberger, K. J. (2012). Development of the enzyme–substrate interactions concept inventory. *Biochemistry and Molecular Biology Education*, *40*(4), 229–233. doi:10.1002/bmb.20622

Bybee, R. W., & McCrae, B. (2009). *PISA science 2006: Implications for science teachers and teaching*. Arlington, VA: NSTA Press.

Chamers, S. K., & Thomas, A. (1997). Gender, prior knowledge, interest, and experience in electricity and conceptual change text manipulations in learning about direct current. *Journal of Research in Science Teaching*, *34*(2), 107–123.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*(5), 1098–1120.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, *90*(6), 1073–1091. doi:10.1002/sce.20164

Crawford, B. A. (2007). Learning to teach science as inquiry in the rough and tumble of practice. *Journal of Research in Science Teaching*, *44*(4), 613–642. doi:10.1002/tea.20157

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *102*(3), 687–700. doi:10.1037/a0019452

von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI monograph series: Vol. 2. IERI monograph series: Issues and methodologies in large scale assessments* (pp. 9–36). Hamburg: IEA-ETS Research Institute.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*(3), 384–397. doi:10.1002/sce.20194

Dochy, F., 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*(2), 145–186. doi:10.3102/00346543069002145

Duggan, S., & Gott, R. (2000). Intermediate General National Vocational Qualification (GNVQ) science: A missed opportunity for a focus on procedural understanding? *Research in Science & Technological Education*, *18*(2), 201–214. doi:10.1080/713694978

Eckhardt, M. (2010). *Instruktionale Unterstützung beim Lernen mit Computersimulationen im Fach Biologie* [Instructional support for learning with computer simulations in biology]. Kiel: IPN. Retrieved from http://eldiss.uni-kiel.de/macau/receive/dissertation_diss_00004981

Evans, M., Kelley, J., Sikora, J., & Treiman, D. J. (2010). Family scholarly culture and educational success: Books and schooling in 27 nations. *Research in Social Stratification and Mobility*, *28*(2), 171–197. doi:10.1016/j.rssm.2010.01.002

Frey, A., Hartig, J., & Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice*, *28*(3), 39–53. doi:10.1111/j.1745-3992.2009.00154.x

Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., … Pekrun, R. (Eds.). (2009). *PISA-2006-Skalenhandbuch: Dokumentation der Erhebungsinstrumente* [PISA 2006 handbook of scales. Documentation of assessment instruments]. Münster: Waxmann.

Gott, R., & Duggan, S. (1995). *Investigative work in the science curriculum. Developing science and technology education*. Buckingham: Open University Press.

Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, *35*(2–3), 57–63. doi:10.1016/j.stueduc.2009.10.002

Häussler, P., & Hoffmann, L. (2000). A curricular frame for physics education: Development, comparison with students' interests, and impact on students' achievement and self-concept. *Science Education Review Letters*, *84*(6), 689–705.

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*(9), 870–888. doi:10.1002/tea.10048

Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen (KFT 4-12+ R)* [Cognitive ability test]. Göttingen: Beltz-Testgesellschaft.

Hewson, M. G., & Hewson, P. W. (1983). Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching*, *20*(8), 731–743. doi:10.1002/tea.3660200804

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, *88*(1), 28–54. doi:10.1002/sce.10106

Justi, R. S. (2009). Learning how to model in science classroom: Key teacher's role in supporting the development of students' modelling skills. *educación química*, *20*(1), 32–40.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie/Journal of Psychology*, *216*(2), 61–73. doi:10.1027/0044-3409.216.2.61

Köller, O., & Parchmann, I. (2012). Competencies: The German notation of learning outcomes. In S. Bernholt, K. Neumann, & P. Nentwig (Eds.), *Making it tangible. Learning outcomes in science education* (pp. 151–172). Münster: Waxmann.

Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, *33*(1), 27–50. doi:10.1080/09500693.2010.518645

Kremer, K., Specht, C., Urhahne, D., & Mayer, J. (2014). Relationship in biology between nature of science and scientific inquiry. *Journal of Biological Education (JBE)*, *48*(1), 1–8. doi:10.1080/00219266.2013.788541

Kuhl, P., Siegle, T., & Lenski, A. E. (2013). Soziale Disparitäten [Social disparities]. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle, & C. Pöhlmann (Eds.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* [IQB-inner-German-comparison. Competences at the end of the lower secondary school level] (pp. 275–296). Münster: Waxmann.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*(4), 495–523. doi:10.1207/S1532690XCI1804_3

van Laere, E., Aesaert, K., & van Braak, J. (2014). The role of students' home language in science achievement: A multilevel approach. *International Journal of Science Education*, *36*(16), 2772–2794. doi:10.1080/09500693.2014.936327

Lazonder, A. W., & Egberink, A. (2014). Children's acquisition and use of the control-of-variables strategy: Effects of explicit and implicit instructional guidance. *Instructional Science*, *42*(2), 291–304. doi:10.1007/s11251-013-9284-3

Le Deist, F. D., & Winterton, J. (2005). What is competence? *Human Resource Development International*, *8*(1), 27–46. doi:10.1080/1367886042000338227

Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire. *Journal of Research in Science Teaching*, *51*(1), 65–83. doi:10.1002/tea.21125

Lederman, N. G. (2004). Syntax of nature of science within inquiry and science instruction. In L. B. Flick & N. G. Lederman (Eds.), *Science & technology education library: v. 25. Scientific inquiry and nature of science. Implications for teaching, learning, and teacher education* (pp. 301–317). Dordrecht: Kluwer Academic Publishers.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* [Testconstruction and testanalysis] (6. Aufl., Studienausg). Weinheim: Beltz.

Lubben, F., Campbell, B., Buffler, A., & Allie, S. (2001). Point and set reasoning in practical science measurement by entering university freshmen. *Science Education*, *85*(4), 311–327. doi:10.1002/sce.1012

Maerten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. *Science Education*, *94*(6), 937–962. doi:10.1002/sce.20408

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, *34*(15), 2317–2341. doi:10.1080/09500693.2012.684433

National Research Council. (2006). *National science education standards*. Washington, DC: Author.

Nehring, A. (2014). *Wissenschaftliche Denk- und Arbeitsweisen im Fach Chemie. Eine kompetenzorientierte Modell- und Testentwicklung für den Bereich der Erkenntnisgewinnung: Unveröffentlichtes Dissertationsmanuskript*. [Scientific reasoning and inquiry methods in the context of chemistry. The development of a competence-based test measuring skills in the field of scientific inquiry]. Unpublished inaugural dissertation.

Nehring, A., Nowak, K. H., Upmeier zu Belzen, A., & Tiemann, R. (2012). Doing inquiry in chemistry and biology. The context's influence on the students' cognitive load. *La Chimica nella Scuola*, *XXXIV-3*, 253–258.

Nehring, A., Nowak, K. H., Upmeier zu Belzen, A., & Tiemann, R. (Submitted for publication). Towards a theory-based assessment of inquiry skills in the context of chemistry. *Chemistry Education Research and Practice*.

Nowak, K. H., Nehring, A., Tiemann, R., & Upmeier zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education*, *47*(3), 182–188. doi:10.1080/00219266.2013.822747

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional \'High-Stakes\' measures of high school students' science achievement. *American Educational Research Journal*, *44*(1), 161–196. doi:10.3102/0002831206298171

Organisation for Economic Co-operation and Development. (2006). Background characteristics, mathematics performance and learning environments of immigrant students. In OECD (Ed.), *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003* (pp. 57–164). Paris: Author.

Organisation for Economic Co-operation and Development. (2013). *Pisa 2012 assessment and analytical framework: Mathematics, reading, science, problem solving*. Paris: Author.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079. doi:10.1080/0950069032000032199

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. doi:10.1207/S15326985EP3801_1

Pandey, S., & Elliott, W. (2010). Suppressor variables in social work research: Ways to identify in multiple regression models. *Journal of the Society for Social Work and Research*, *1*(1), 28–40. doi:10.5243/jsswr.2010.2

PISA-Konsortium Deutschland (Ed.). (2006). *PISA 2006 – Die Ergebnisse der dritten internationalen Vergleichsstudie* [PISA 2006 - Results from the third international comparative study]. Münster: Waxmann.

Pyburn, D. T., & Pazicni, S. (2014). Applying the multilevel framework of discourse comprehension to evaluate the text characteristics of general chemistry textbooks. *Journal of Chemical Education*, *91*(6), 778–783. doi:10.1021/ed500006u

Reid, N., & Skryabina, E. A. (2003). Gender and physics. *International Journal of Science Education*, *25*(4), 509–536. doi:10.1080/0950069022000017270

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Whiley.

Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *Journal of the Learning Sciences*, *4*(2), 131–166. doi:10.1207/s15327809jls0402_1

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6-12. Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12* [Test for reading skills and reading speed for grades 6 to 12]. Göttingen: Hogrefe.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss (Jahrgangsstufe 10)* [National education standards for the subject chemistry]. München: Luchterhand.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, *41*(1), 159–189. doi:10.3102/00028312041001159

Stark, R., & Gray, D. (2010). Gender preferences in learning science. *International Journal of Science Education*, *21*(6), 633–643. doi:10.1080/095006999290480

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, *4* (4), 295–312.

Taboada, A. (2012). Relationships of general vocabulary, science vocabulary, and student questioning with science comprehension in students with varying levels of English proficiency. *Instructional Science*, *40*(6), 901–923. doi:10.1007/s11251-011-9196-z

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research*, *64*(1), 37–54.

Trier, U., Krüger, D., & Upmeier zu Belzen, A. (2012). Students' versus scientists' conceptions of models and modelling. In D. Krüger & M. Ekborg (Eds.), *ERIDOB-proceedings: Research in biological education* (pp. 103–115). Berlin: ERIDOB.

Walpuski, M., Ropohl, M., & Sumfleth, E. (2011). Students' knowledge about chemical reactions—Development and analysis of standard-based test items. *Chemistry Education Research and Practice*, *12*(2), 174. doi:10.1039/c1rp90022f

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Sagalnik (Eds.), *Definition and selection of competencies. Theoretical and conceptual foundations* (pp. 45–65). Kirkland, WA: Hogrefe & Huber.

Wren, D., & Barbera, J. (2014). Psychometric analysis of the thermochemistry concept inventory. *Chemistry Education Research and Practice*, *15*(3), 380. doi:10.1039/C3RP00170A

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*(2–3), 114–128. doi:10.1016/j.stueduc.2005.05.005

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell: ACER Press.

Zimmermann, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*, 99–149.