Becoming Proficient through Profile Classes: A Longitudinal Study on the Development of Scientific Competencies

Alena Schulte
Bielefeld University, Germany

Claas Wegner
Bielefeld University, Germany

Abstract
The first educational goal within scientific subjects is to acquire a sense of scientific literacy. In science lessons, methods of scientific inquiry provide the tools to achieve this. In this study, we based scientific inquiry on the SDDS-Model according to Klahr (2002). It is divided into three subareas: Search Hypothesis Space, Test Hypothesis, and Evaluate Hypothesis. A multiple-choice test, the NAW-test, was used to examine the extent to which the acquisition of competencies is promoted by attending Profile Classes. In contrast to common practice in Germany, scientific Profile Classes take an interdisciplinary approach to scientific subjects, with the aim to promote the acquisition of scientific competencies. For this purpose, Profile Class students (N=84) at two schools were questioned over the course of a school year at three different test times. Results show that competencies increase over time. A gender difference was not observed.

Keywords
NAW-Test, Profile Classes, Scientific Inquiry, Secondary Schools, Scientific Competencies

Introduction

Scientific literacy is currently one of the goals of science education worldwide. Scientific literacy describes the ability to apply scientific knowledge to acquire new skills. In order to evaluate current and relevant topics in a modern society, the interconnection of everyday phenomena with specialized knowledge is crucial (Organization for Economic Co-operation and Development (OECD), 2019). One possibility to increase scientific literacy is to include scientific methods (generating hypotheses, planning and conducting experiments, and making observations) in class through an inquiry approach. In theory, an inquiry approach is rated as one of the main teaching methods to strengthen scientific literacy (Chi et al., 2019). However, there are many different approaches and definitions of this concept (Kidman & Casinader, 2017). Depending on the theoretical basis, different areas of competence are fostered by scientific inquiry. Based on the model developed by Klahr (2002), scientific inquiry is divided into three core competencies: *Search Hypothesis Space, Test Hypothesis*, and *Evaluate Evidence*.

International studies show that students benefit from an inquiry approach in regards to understanding concepts (see Minner et al., 2010 for a review about inquiry-based science

Citation: Schulte, Alena, & Wegner, Claas. (2021). Becoming Proficient through Profile Classes: A Longitudinal Study on the Development of Scientific Competencies. *Global Education Review, 8* (4), 24-40.

instruction). Crawford (2014) concluded that most science classes are teacher-centered, and fewer teachers apply inquiry methods. Students conduct experiments in strongly guided settings and therefore have few opportunities to choose and carry out their own research projects. This could be caused by insufficient teacher training, since teachers may have little experience in planning and implementing their own science projects (Hodson, 2014). Due to the many curricular requirements, there is also little time in the classroom to implement open learning formats such as scientific inquiry methods, which take more time than a teacher-centered approach.

To counteract this, scientific Profile Classes were developed in Germany and implemented in schools. Additional teaching time should enable students to plan, carry out, and evaluate their own research projects. The extent to which scientific inquiry supports the acquisition of competencies is examined in this study.

Scientific Literacy

The educational goal of scientific literacy goes back to an initiative in the United States from the 1950s that aimed to reform science education. The National Research Council defined scientific literacy as “the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity.” (National Research Council, 1996, p. 22).

Today, this educational goal can be found in both international and national educational standards. Feinstein (2011) pointed out that although scientific literacy is an important and useful educational goal, its usefulness has not yet been empirically tested. Various views and definitions of the term still exist today (Schulte, 2017). Based on a literature review, Roberts (2007) and Roberts et al. (2014) elaborated two visions of scientific literacy: Vision I is based on the perspective of the student as a novice scientist, where the ability to study is the main goal of science education. Students should be enabled to study a scientific subject and pursue a scientific career, subsequently, by engaging in the field of science. In contrast, Vision II is more broad: Science education does not concern individual students, but everyone. All students should have scientific literacy to be able to judge social, political, economic, and ethical problems. The focus is on social participation (Roberts & Bybee, 2014). Sjöstrom et al. (2017) added a third vision that emphasizes solving socio-scientific issues.

Bybee's (2002) definition of scientific literacy is widely cited and used as a basis in numerous school development processes and comparative studies. According to him, scientific literacy involves different levels of scientific understanding. He distinguishes four dimensions:[i]

- *Nominal Scientific Literacy*

  Learners identify scientific ideas or topics, but these associations are still linked to misconceptions or a lack of knowledge.
- *Functional Scientific Literacy*

  Students know scientific terms and use them correctly.
- *Conceptual und Procedural Scientific Literacy*

  Learners are able to connect information and experience and link the different scientific disciplines.

- *Multidimensional Scientific Literacy*
  Learners differentiate science from other disciplines and define its characteristics.

The dimensions can be viewed as levels that can and should be achieved through promotion at school. The model not only focuses on the acquisition of specialized knowledge, but also on process knowledge (Bybee, 2002).

International comparative studies like the Programme for International Student Assessment (PISA) are based on Bybee's scientific literacy concept. Students should be enabled "to engage with science-related issues, and with the ideas of science, as a reflective citizen" (OECD, 2019, p. 100). The concept's aim is for each student to become a scientifically literate person who is willing to take part in reasoned discourse about science and technology. In the latest PISA study of 2018, three sub-competencies of scientific literacy are distinguished: (1) Explaining phenomena scientifically, (2) evaluating and designing scientific enquiry, (3) interpreting data and evidence scientifically (OECD, 2019).

The rapid scientific progress through new findings and technologies requires students to independently access information. According to PISA, this information can be generated from scientific publications as well as from their own research. To draw conclusions from their own actions, students should critically evaluate information that they receive through media, such as social media and newspapers (Schiepe-Tiska et al., 2016).

In Germany, the acquisition of scientific literacy is anchored within the educational standards of the scientific subjects biology, chemistry, and physics, and is based on the definition by Bybee (2002). Educational standards define the knowledge and skills that students should have at the end of a particular school year in order to obtain a qualification (primary and secondary school graduation). The educational standards established by the national government provide the framework for the ordinances issued by the federal states (Kultusministerkonferenz, n.d.). For each scientific subject, there are subject-specific requirements. One goal of all subjects is to apply scientific inquiry to acquire scientific literacy. In secondary school, students acquire competencies in the four areas of knowledge, knowledge acquisition, communication and assessment (see Table 1)

As shown in Table 1, knowledge acquisition includes scientific inquiry skills. The acquisition of scientific methods supports students in achieving scientific literacy. Similar to the lack of clarity to define the term scientific literacy, there are numerous definitions to quantify the competence of scientific inquiry (Chi et al., 2019). Each subject in school has its own set of inquiry methods, which may be referred to as scientific inquiry or geographical inquiry (Kidman & Casinader, 2017).

Scientific Inquiry

Kidman and Casinader (2017) argue that inquiry is a multifaceted concept that takes both teaching and learning perspectives into account. Inquiry can describe domain-specific knowledge, the way students learn, an instructional approach, or the curriculum material (Furtak et al., 2012). In relation to science education, inquiry describes a learning process that generates scientific knowledge through competency, content, and critical thinking (Lederman, 2004; Lederman et al., 2014). The basic idea behind scientific inquiry is that students learn by engaging in scientific activities. Scientific inquiry is comprised of the following: generate scientifically oriented questions, conduct investigations to collect evidence, explain and discuss phenomena (Furtak et al. 2012). Students, therefore, work like scientists in the classroom (Schiefer et al., 2020). Depending on the theoretical basis, different processes/competencies are assigned to scientific inquiry:

- The National Science Education Standards (1996) assign five features to scientific inquiry: Posing questions, constituting, and collecting evidence, formulating explanations, evaluating explanations, communicating and justifying explanations.
- Zachos et al. (2000) divides scientific inquiry in linking theory with evidence, formulating hypotheses, keeping records, using correct or original laboratory materials, identifying cause-and-effect relationships, controlling experiments, and using parsimony in drawing conclusions.
- The model of Nowak et al. (2013) distinguishes scientific inquiry according to three methods (modelling, experimenting and observing, and comparing and arranging) and three scientific reasoning processes (question and hypothesis, plan and performance, and scientific reasoning).
- Kuo et al. (2015) describe four competencies of scientific inquiry: Questioning, experimenting, analyzing, and explaining.
- The scientific competence in the PISA-Studies in 2015 & 2018 (OECD, 2017 & OECD, 2019) is divided into: (1) Explaining phenomena scientifically, (2) evaluating and designing scientific enquiry, (3) interpreting data and evidence scientifically are distinguished.

In general, the term scientific inquiry covers all scientific working methods that can be methodically applied in class and used by the students. The vague definition or different design of this concept requires a precise theoretical justification. In this study, the definition of scientific inquiry is based on the *Scientific Discovery as Dual-Search-Modell (SDDS-Model)* by Klahr (2002). Klahr's model was based on empirical studies in the field of psychology, and applied to the practice of science. A group of researchers under Klahr investigated the reasoning of people who solve scientific problems. According to the model, scientific inquiry is divided into three areas: Forming hypotheses, experimenting and evaluating evidence (see Figure 1). If students apply these areas, the quality of learning can be improved (Höffler et al., 2014). The SDDS-Model gives a general framework of human behavior, which can be interpreted in any scientific reasoning task.

In *Search Hypothesis Space,* students use their prior knowledge to generate a hypothesis about a given scientific problem. For example, students investigate the germination of seeds and hypothesize that seeds need water to sprout. The output from *Search Hypothesis Space* is a fully specified hypothesis, which provides the input to *Test Hypothesis*. To test the hypothesis, students plan and conduct experiments. In our example, the students would plan an experiment to test seed germination with and without water. Results of the component *Test Hypothesis* either confirm or deny the hypothesis and are based on the match between the student's prediction and the experimental result. This process is part of

Evaluate Evidence. Students decide whether their experimental results warrant acceptance, rejection, or consideration of the current hypothesis (Klahr, 2002).

Based on the various definitions and perspectives, the focus of this study is on how competent students are in applying the methods of scientific inquiry. For this purpose, the study follows Klahr's assumption that scientific inquiry is divided into three components: *Forming hypotheses, conducting experiments, and drawing conclusions*.

Since scientific inquiry can be divided into the three components according to Klahr's SDDS model, it is crucial to investigate the current state of research concerning student competencies in those three areas.

According to Piekny and Mahler (2013), students develop competencies for *Forming Hypotheses* after primary school. In addition, Osterhaus et al. (2020) note that primary school children (8 years old) can easily distinguish between testing hypotheses and producing an effect. With regard to the three areas of the Klahrs Model, it was found that 7th grade students often have the correct answer in idea formation/generating hypotheses (Klos, 2008). The PISA study in 2015 focused on student scientific competencies. In Germany, the three sub-competencies are equally high and above the OECD average. German students tend to perform better in the competence *explaining phenomena*, but below average in the other two competencies. For the first time in the OECD average and in Germany, boys have significantly higher scientific competence than girls (OECD, 2017). This difference is confirmed in a 2018 study but is no longer significant and is only seen in the competence explaining phenomena (OECD, 2019). In 2015, Finland was the only country where girls achieved better results than boys (OECD, 2017). In a study on the effects of early scientific education in 5th and 6th grades at secondary schools in Germany, Höffler et al. (2014) demonstrated that the knowledge of scientific working methods is increasing for all students and demonstrate that girls show higher scores at both test times.

Walpuski and Schulz (2011) summarize the results of studies on scientific inquiry. Their analysis showed that although students in 7th grade are able to conduct experiments to test their hypothesis under instruction, they find it difficult to conduct experiments independently. Furthermore, they replicate these results with students in 7th grade who are good at forming hypotheses/ideas but find it difficult to plan appropriate experiments and draw conclusions. One reason could be that in science lessons, usually only confirmatory experiments are conducted (Walpuski & Schulz, 2011). In the 2015 PISA study, students performed lower on the competence evaluating and designing scientific enquiry than on the competence explaining phenomena scientifically. Other studies assessing the competence *conducting experiments* show similar results and show that this competence is important in the learning process (OECD, 2017). Piekny and Mahler (2013) point out that elementary school children already understand that experimenting does not always lead to the desired result.

Regarding the area of *drawing conclusions*, the 2015 PISA study shows that students perform the lowest in this competence (OECD, 2017). Chi et al. (2019) investigate disciplinary context effects on student performance in terms of scientific inquiry competencies. Students show particular difficulties in applying this competence and low levels of performance. According to Chi et al. (2019), teachers should offer students more opportunities to use scientific inquiry methods.

Kidman and Casinader (2017) state that students need help in using scientific inquiry, especially in conducting experiments and drawing conclusions.

Regarding interventions, Minner et al. (2010) analyzed the impact of an inquiry approach on students. Their review showed that the students who received inquiry-instructed lessons demonstrated improved learning based on comparisons to their prior knowledge; this finding was also evident when compared to students who were taught using a different method of instruction. Students with instruction based on the investigation cycle (generating questions, designing experiments, collecting data, drawing conclusion, and communicating findings) particularly benefitted. Active thinking and participating in the investigation process, along with hands-on experiences, are associated with increased conceptual learning (Minner et al., 2010). Kuo et al. (2015) postulate that the use of scientific inquiry methods promotes active learning processes among students. In this case, students actively use methods to generate knowledge. Schiefer et al. (2020) show that scientific skills can be learned in elementary school using a scientific inquiry approach.

In summary, based on the international state of research, students perform particularly well in the area of hypothesis formation. Difficulties are encountered by students in almost all studies in the area of drawing conclusions. Furthermore, studies show that students benefit from a scientific-inquiry approach.

Scientific Profile Classes

School profiles are not a uniform concept. It is instead an intentional process with which the school attempts to further develop its existing profile. A distinction is generally made between profiling individual classes and entire schools. Class profiling refers to individual classes that exist along with regular classes. Creating a profile might aim to promote individual students, to compete with other schools and to implement educational policy decisions. Concepts for school profiles exist primarily at the transitions between primary and lower secondary education, and lower and upper secondary education (Altrichter, Heinrich & Soukup-Altrichter, 2011). In Germany, profiles can mostly be found at Gymnasien.[ii] The wide range of subjects offered makes it relatively simple and quick to implement profiles. On the other hand, there is a great deal of competition among high schools in metropolitan areas for student enrollment (Klekovkin et al., 2015). Profile Classes range from concepts for gifted students without an explicit thematic focus (Hackl, 2009) to concepts with a specialized focus; in areas such as music (Haas et al., 2019), science/technology (Spörlein, 2003: Schulte & Wegner, 2020), language/bilingual (Nold et al., 2008), sports (Roth et al., 2017). Out of these areas, a music profile is most often found in Gymnasium. Only a few schools offer a researcher focus (Klekovkin et al., 2015).

Since low scores from students in the first large-scale assessments have been observed (see OECD, 2000 & Martin et al., 2004), numerous initiatives have been taken to improve science teaching in Germany (Schiepe-Tiska et al., 2016). The concepts ranged from individual teaching methods to curricula and school structural changes. Profile Classes include a structural approach to improve science teaching. Scientific subjects are taught separately at most secondary schools in Germany. An interdisciplinary approach to teaching is usually only implemented in 8th and 9th grades in the Gymnasium, when students attend electives.

Scientific Profile Classes on the other hand, take an interdisciplinary approach to scientific subjects from the beginning of lower secondary school, i.e. from the 5th grade onwards. Their aim is to promote interest, self-concept, and competence in science. With additional teaching, students are given the opportunity to work on their own research questions and conduct their own experiments. This self-directed approach is in line with the demand of Chi et al. (2019), to support students in autonomously chosen research projects. In contrast to regular classes, lessons are strongly student-oriented and characterized by open learning settings. Topics are related to the curriculum of the subjects of biology, chemistry, and physics, and therefore follow an interdisciplinary approach. The core curriculum of Profile Classes is designed to enable students to contribute their own ideas and content to the lessons. Students learn to independently develop questions, plan and conduct experiments, and evaluate the results. As a result, lessons in Profile Classes tend to follow the SDDS-Model (see Figure 2).

The focus of teaching in Profile Classes is on scientific inquiry methods, predominantly on generating hypotheses, making and comparing observations, planning and conducting experiments, working with models, evaluating evidence, and drawing conclusions. Applying these methods encourages the students to become self-directed learners and problem-solvers (Concannon et al., 2020). To date, the success of school profiling has been measured in terms of maintaining and/or increasing student enrollment, perceived attractiveness of the school, and instructional development (Klekovkin et al., 2015). Profile classes on the other hand have hardly been evaluated in terms of (long-term) effects on students (Nonte, 2013). In this article, we focus on evaluating profile classes.

Research Questions

Profile Classes are established to promote interest and competence in science, with the focus on acquiring inquiry skills (Schulte & Wegner, 2020). The teaching materials and methods are designed to equally support students in the three core areas of the SDDS-Model developed by Klahr (2002). In Profile Classes, there is a large focus on planning and conducting experiments.

Recent studies on the NAW-test and scientific inquiry show that *forming hypotheses* is well-developed, gender differences are inconsistent (e.g. Klos, 2008; Höffler et al., 2014), that there is an increase in competence over the school years (e.g., Grube & Mayer, 2010; Osterhaus et al., 2020) and that grades have no influence on the test result (Wellnitz et al., 2012). This study examines the following hypotheses:

H1: The three competencies examined – *forming hypotheses*, *planning and conducting experiments*, and *drawing conclusions* – do not differ among students in Profile Classes.

H2: There are no gender differences in test results from students in Profile Classes.

H3: Competencies of Profile Class students determined by the NAW-Test increase over the testing period.

H4: Biology grades do not correlate with NAW-Test results.

H5: Students' scientific interests have a positive influence on NAW-Test test results (in the overall test and the three competencies).

H6: Ability self-concept has a positive influence on NAW-Test test results (in the overall test and the three competencies).

Method

Data Collection and Test Instruments

The sample consists of three 5th grade scientific Profile Classes at two schools (N=84; male: 63 & female: 21). Although the screening process to participate in Profile Classes enforces a balanced gender distribution, more boys apply and thus, participate, in the Profile Classes. Students completed a questionnaire to measure interest and ability self-concept (see Table 2), and they completed the *NAW-test* to evaluate their competencies regarding the three steps of scientific inquiry. In addition, sociodemographic data such as gender, grades, and age were collected. Students completed both instruments at the beginning, in the middle, and at the end of 5th grade (see Figure 3).

The questionnaire consists of 33 closed items, which were answered using a 6-point rating scale (strongly agree – strongly disagree; very good – very bad). The scales of scientific interest and ability self-concept are taken from well-established test instruments (see Table 2).

The *NAW-test* was originally developed by Klos et al. (2008) as German students score below average in international comparative studies. It has been adapted for different age groups and intervention formats (e.g., Walpuski, 2006; Klos, 2008; Mannel, 2011; Koenen, 2014). The version used in this study is based on the NAW-Test by Koenen (2014).

The test instrument is based on the assumption that the experimental-scientific working methods can be simplified to three areas of scientific inquiry methods: *Forming hypotheses*, *Planning and conducting experiments*, and *Drawing conclusions* (Klahr, 2002). The NAW-Test is a paper-pencil test in multiple-choice format; each item consists of an informational text, a question, and four possible answers, one of which is correct. The text provides the students with the essential information to solve the task. The NAW-Test is independent of prior knowledge. Students can achieve a total score of 20. Table 3 shows an example item per area and indicates the score for the individual areas.

Results

A total of 84 students participated in the survey, however, a complete dataset includes 22 students. Reasons for this may be class and school changes, but it may also result from the school shutdown in Spring 2020. Due to different test scores in the competence areas, percentage scores were calculated. The effect strengths of the calculations are given according to Cohen (1988, 1992).

Comparison of competencies in the Profile Classes

Figure 4 shows the students' mean scores in percentage at each test time for each competence (see Figure 4).

An analysis of variance was conducted on the comparison of students' competencies in Profile Classes at each test time:

Test time 1: At the beginning of the school year (N=84)

An analysis of variance (sphericity assumed: Mauchly (2) = .981, p =.454) shows significant differences in the test results of the sub-competencies ($F(2, 166) = 21{,}102$, $p = .000$, partial $\eta^2 = .203$) with an effect strength of .505. The Bonferroni-adjusted post-hoc analysis show that the sub-competence *forming hypotheses* (M=0,41; SD=0,23) differs significantly from the sub-competence *planning and conducting experiments* (M=0,56, SD=0,25) (-0,15.95%-KI [-0,220,-0,083]) and significantly from the sub-competence *drawing conclusion* (M=0,57, SD=0,22) (-0,163.95%-KI [-0,236,-0,091]).

Test time 2: In the middle of the school year (n=46)

An analysis of variance (sphericity assumed: Mauchly (2) = .983, p =.688) shows significant differences in the test results of the sub-competencies ($F(2,90) = 20{,}433$, $p = .000$, partial $\eta^2 = .312$) with an effect strength of .673. The Bonferroni-adjusted post-hoc analysis show that the sub-competence *forming hypotheses* (M=0,54; SD=0,27) differs significantly from the sub-competence *planning and conducting experiments* (M=0,68, SD=0,21) (-0,134.95%-KI [-0,213,-0,055]) and significantly from the sub-competence *drawing conclusion* (M=0,75, SD=0,22) (-0,203.95%-KI [-0,279,-0,127]).

Test time 3: At the end of the school year (n=41)

Given a significant p-value in the Mauchly test (Mauchly (2) = .852, p = .044) and due to the small sample size, a Friedman test was calculated, which shows significant differences (Chi-square(2)=8,972, p=.011, n=41). Post-hoc-adjusted Dunn-Bonferroni-tests show that sub-competence *forming hypotheses* (M=0,66; SD=0,30) differs significantly from the sub-competence *planning and conducting experiments* (M=0,76, SD=0,21) (*z*=-0,598, *p*=.020, effect strength:.09).

Gender difference in the Profile Classes

A repeated measures ANOVA was conducted to look at gender as an interaction effect. Overall NAW-Test results (sphericity assumed: Mauchly (2) = .786, *p* = .101; $F(2, 40) = 2{,}918$, *p* = .066, partial η^2 = .127) and the sub-areas of f*orming hypotheses* (sphericity assumed: Mauchly (2) = .877, *p* = .306; $F(2, 38) = 2{,}053$, *p* = .142, partial η^2 = .098), *planning and conducting experiments* (sphericity assumed: Mauchly (2) = .922, *p* = .479; $F(2, 38) = 1{,}191$, *p* = .315, partial η^2 = .059) and d*rawing conclusions* (sphericity assumed: Mauchly (2) = .944, *p* = .577; ($F(2, 40) = 2{,}103$, *p* = .135, partial η^2 = .095) show no significant main effect of gender.

Progression of competence

A repeated measures ANOVA was conducted to examine NAW-test results. The results are reported according to the overall test result (see figure 5) and the three sub-areas (see figure 6).

Overall NAW test results (n=22):

Due to the small sample size and a significant Mauchly-Test, a Friedman test was calculated which shows significant differences (Chi-square(2)=17.084, p=.000, n=22). Post-hoc-adjusted Dunn-Bonferroni-tests show that the first test time (M=0,51; SD=0,16) differs significantly from the third test time (M=0,68, SD=0,25) (*z*=-1,205, *p*=.000, effect strength: .26).

The diagram below illustrates a division of the competence areas (see Figure 6).

Forming Hypotheses **(n=21):**

The Friedman test shows significant differences between test times (Chi-squared(2)=8,027, *p*=.018, n=21). Post-hoc-adjusted Dunn-Bonferroni-tests show that the first test time (M=0,42; SD=0,21) differs significantly from the third test time (M=0,62, SD=0,31) (*z*=-0,810, *p*=.026, effect strength: .18).

***Planning and conducting experiments* (n=21):**

The Friedman test shows significant differences between test times (Chi-squared(2)=12,028, *p*=.002, n=21). Post-hoc-adjusted Dunn-Bonferroni-tests show that the first test time (M=0,56; SD=0,21) differs significantly from the third test time (M=0,79, SD=0,21) (*z*=-0,976, *p*=.005, effect strength: .21).

***Drawing conclusions* (n=22):**

A Friedman test was calculated which shows no significant differences between test times (Chi-Quadrat(2)=2,171, *p*=.338, n=22).

Factors influencing the test results

The influence of biology grade on the overall NAW test result was only observed at the first testing time (t1) in a regression analysis (F(1,78)=4,955, *p*= .029; effect strength: .22; N= 80); also for *forming hypotheses* (F(1,78)=5,021, *p*= .028; effect strength: .22; N= 80) and *drawing conclusions* (F(1,78)=4,055, *p*= .048; effect strength: .20; N= 80).

If the score on the NAW test increases at the first test time point, the grade decreases by -0.83. The corrected R-squared is .048; thus, 4.8% of the total score variance is explained by the grade. The score also decreases in the areas of *forming hypotheses* and *drawing conclusions* (hypotheses: -0.107; conclusions: -0.901). The corrected R-squared for *forming hypotheses* is .048 and for *drawing conclusions* is .037.

Multiple regressions show that the factor scientific interest has no influence on the result, neither in the overall result nor in the sub-areas. The factor ability self-concept is a predictor for the test result of *experiment* at the third test time (F(2,37)=3,946, *p*= .028; effect strength: .39; N= 40). If the score in *experiment* increases, ability self-concept increases by .150; the corrected R-squared is .131, which means that 13% of the total variance of the result is explained by ability self-concept.

Discussion

Comparison of competencies in the Profile Classes

Analyses of variance were used to examine whether there were differences between sub-competencies at each test time. The competence *forming hypotheses* differs significantly from *planning and conducting experiments* at all three test times and it also differs from *drawing conclusions* at the first and second test time, with students performing significantly lower in *forming hypotheses* at all test times. The first hypothesis of this study must therefore be rejected (H1). Our findings contradict previous research, which suggests that students perform best at *forming hypotheses* (e.g. Kidman & Casinader, 2017; Schiepe-Tiska et al., 2016; Walpuski & Schulz, 2011). In profile classes, less time is spent on students generating hypotheses compared to the other two competencies. This is due to the fact that the teaching material/methods particularly promote the competence area of *conducting experiments* and *drawing conclusions* by focusing on students conducting their own research projects. Students develop their own questions, plan and conduct experiments, and evaluate the results. A large amount of time is devoted to planning and carrying out experiments in the classes; therefore students spend less time generating their hypothesis. The next big aspect is to present their data, which they must first accurately interpret results and draw conclusions before presenting it at a research conference.

Previous studies have confirmed that the competence *drawing conclusion* is the slowest to develop among students (Chi et al., 2019). In the study by Chi et al. (2019), similar results are shown in regard to the competence "Evaluating and making argumentation," which is similar to *drawing conclusion*. Students experienced the most problems in solving these tasks. Kidman and Casinader (2017) also point out that students need the most support in the area of drawing conclusions. In contrast, the acquisition of this area of competence seems to be promoted very well in the scientific Profile Classes. Students perform significantly better in this competence at the first and second test time.

The digital teaching caused by the coronavirus pandemic could be another reason for the result at the end of the school year, students performing best in *planning and conducting experiments*. Schools were closed between the second and third test dates due to the high number of cases in Germany (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2020 & Robert-Koch-Institute, 2020). Subsequently, classes were taught using digital distance learning. Online teaching made it more difficult to use methods of scientific knowledge acquisition (like hands-on activities or inquiry-learning). An interview study with Profile Class teachers (N=7) report difficulties in getting students to work on their own (Schulte & Wegner, in prep.). These preliminary research results could also have an impact on the examined Profile Classes. Teachers designed lessons for students to conduct experiments at home with few materials, filmed their own experiments in schools to show students, or used virtual resources. It has been shown that students benefit in interest and knowledge acquisition from the use of digital media and virtual labs (Kluge, 2014; Schäfers et al., 2020).

To summarize, the results contradict previous studies. Over time, students in the profile classes perform best in *drawing conclusions* (Chi et al., 2019) and show the greatest difficulty in *forming hypotheses* (Schiepe-Tiska et al., 2016).

Gender difference in the Profile Classes

In this study, gender has no influence on NAW-Test scores. This confirms previous research and our hypotheses (H2) (Höffler et al., 2014), but shows an opposite trend to the PISA results from 2015 and 2018. Although there is a slight tendency in the means for girls to perform better, the small sample size must be considered, along with the fact that the proportion of boys is higher; male students represent 75% of the total sample. Not all students could be questioned at all three test times.

Progression of competence over time

Within one school year, an increase in competence in the NAW-Test and its sub-areas can be observed among Profile Class students, supporting previous research and confirming our hypotheses (H3) (Osterhaus et al., 2020;

Piekny & Maehler, 2013). A pair-wise comparison of the overall results shows that the first test time differs significantly from the second and third time, with the greatest increase between the first and third test time. No significant increase in competence was observed between the second and third time of testing. There was hardly any classroom teaching between those test times. In contrast, between the first test times, the focus in lessons (in-class) was on scientific inquiry. It can therefore be assumed that the teaching in profile classes had a positive effect on the students during this period and on the acquisition of competence in the areas of *forming hypotheses*, *planning and conducting experiments*, and *drawing conclusions*. In contrast to the other competencies, *drawing conclusions* does not increase significantly between the last two test times. Distance learning may have had an impact on the competence acquisition of Profile Class students as methods could hardly be implemented (Schulte & Wegner, in prep.). The significant increase in the area *forming hypotheses* confirms previous research that this skill area can be easily acquired by students (e.g. Walpuski & Schulz, 2011; Schiepe-Tiska et al., 2016).

Factors influencing the test results

Biology grades had an influence on NAW test result scores at the first test time and on the results of the areas *forming hypotheses* and *drawing conclusions* (H4). Thereby, the results show a very low effect size and little of the variance of the result is explained based on the grade. This effect should be examined in further studies with larger sample sizes. Previous studies have shown that the NAW test should be independent of grade (see e.g., Klos et al., 2008). Interest has no influence on the overall test result and on the test results of the competence areas (H5), but ability self-concept is a predictor for the test result in *planning and conducting experiments* at the third test time (H6). Digital learning requires self-regulatory skills; students must plan their school days and independently learn. As self-concept positively influences the competence area of experiments, the adapted teaching methods due to distance learning might be beneficial. To further examine this effect, we should investigate the extent to which self-concept in profile classes changes over time, or if students with a higher self-concept answered more items correctly.

Limitations & Conclusion

One of the limitations of this study is the decreasing sample size over time. The sample of N=84 at the first test time of the school year could not be retained at the following two test times. This may be due to the coronavirus pandemic, as after the second test time, the school lockdown started. By the third test time, the questionnaire and NAW-test were converted to a digital format with the hope that this would provide students with the opportunity to continue to take part online. Since some students do not have mobile devices or internet access at home, not all of them could participate. Furthermore, the study should be expanded to survey a comparison group of regular class students. This leads to further research that questions the extent to which the acquisition of competencies of the students with a profile differs from the acquisition of competencies of students without a special scientific promotion. In our study, the examined group will continue to be surveyed in 6th and 7th grade, to analyze change in competence over several years. We could then investigate whether grades or interest have an influence on NAW-test performance, as shown in other studies. This study nevertheless provides crucial information about how scientific competencies in Profile Classes develop over time and how they can be promoted.

References

Altrichter, H., Heinrich, M., & Soukup-Altrichter, K. (2011). *Schulentwicklung durch Schulprofilierung? Zur Veränderung von Koordinationsmechanismen im Schulsystem.* Wiesbaden: VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH, Wiesbaden

Bybee, R. W. (2002). Scientific Literacy – Mythos oder Realität? Gräber, W., Nentwig, P., Koballa, T. Evans, R. (Eds.) *Scientific Literacy – Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. Opladen : Leske + Budrich, 21-43.

Chi, S., Wang, Z., & Liu, X. (2019). Investigating disciplinary context effect on student scientific inquiry competence. *International Journal of Science Education, 41*(18), 2736-2764.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. 2nd edition. New York: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 122*(1), 155-159.

Concannon, J. P., Brown, P. L., Lederman, N. G., & Lederman, J. S. (2020). Investigating the development of secondary students' views about scientific inquiry. *International Journal of Science Education, 42*(6), 906-933.

Crawford, B. A. (2014). From inquiry to scientific practices in the science classroom. Abell, S. K. & Lederman, N. G. (Eds.), *Handbook of research on science education, 2,* 515-542.

Feinstein, N. (2011). Salvaging science literacy. *Science Education, 95*(1), 168-185.

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science-teaching: A meta-analysis. *Review of Educational Research, 82*(3), 300-329.

Grube, C. & Mayer, J. (2010). Wissenschaftsmethodische Kompetenzen in der Sekundarstufe I: eine Untersuchung zur Entwicklung des wissenschaftlichen Denkens. Harms, U. & Mackensen-Friedrichs, I. (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik. Band 4,* Innsbruck: StudienVerlag, 155–168.

Haas, M., Nonte, S., Krieg, M. & Stubbe, T. C. (2019). Unterrichtsqualität in Musikklassen. Befunde aus der quasi-experimentellen Studie ProBiNi. Weidner, V., Rolle, C. (Eds.), *Praxen und Diskurse aus Sicht musikpädagogischer Forschung,* Münster: Waxmann. 137–154

Hackl, A. (2009). Förderkonzepte hochbegabter Gymnasiastinnen und Gymnasiasten am Beispiel des Deutschhaus-Gymnasium Würzburg. Bosse, D. (Ed.), *Gymnasiale Bildung zwischen Kompetenzorientierung und Kulturarbeit*. VS Verlag für Sozialwissenschaften: Wiesbaden, 187-198.

Hodson, D. (2014). Learning science, learning about science, doing science: Different goals demand different learning methods. *International Journal of Science Education, 36*(15), 2534-2553.

Höffler, T., Lütjohann, F., & Parchmann, I. (2014). Welche Wirkungen erzielt ein naturwissenschaftlicher Anfangsunterricht? *Zeitschrift für Didaktik der Naturwissenschaft, 20*, 87-99.

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

Kidman, G., & Casinader, N. (2017). *Inquiry-based teaching and learning across disciplines*. London : Palgrave Macmillan UK.

Klahr, D. (2002). Exploring science: the cognition and development of discovery processes. Cambridge, Mass. [u.a.] : MIT Press

Klekovkin, T., Nonte, S., & Stubbe, T. C. (2015). Die Verbreitung von Schul- und Klassenprofilen an weiterführenden Schulen in Hamburg und Regierungsbezirk Braunschweig in Niedersachsen. (Working Paper No. 1 des Lehrstuhls Schulpädagogik / Empirische Schulforschung der Georg-August-Universität Göttingen). Retrieved from: http://www.publikationen.stubbe.info/Klekovkin_Nonte_Stubbe_2015_DVvSuK.pdf

Klos, S. (2008). Kompetenzförderung im naturwissenschaftlichen Anfangsunterricht – Der Einfluss eines integrierten Unterrichtskonzepts. Niedderer, H., Fischler, H., Sumfleth, E. (Eds.), *Studien zum Physik- und Chemielernen*, 89. Berlin: Logos.

Klos, S., Henke, C., Kieren, C., Walpuski, M., & Sumfleth, E. (2008). Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei verschiedene Kompetenzen. *Zeitschrift für Pädagogik*, *54*(3), 304-321.

Kluge, A. (2014). Combining laboratory experiments with digital tools to do scientific inquiry. *International Journal of Science Education, 36*(13), 2157-2179.

Koenen, J. (2014). Entwicklung und Evaluation von experimentunterstützten Lösungsbeispielen zur Förderung naturwissenschaftlich-experimenteller Arbeitsweisen. Niedderer, H., Fischler, H., Sumfleth, E. (Eds.), *Studien zum Physik- und Chemielernen*, 171. Berlin: Logos.

Kultusministerkonferenz (n.d.) Bildungsstandards der Kultusministerkonferenz. Retrieved from: https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsstandards.html#c5033

Kuo, C. Y., Wu, H. K., Jen, T. H., & Hsu, Y. S. (2015). Development and validation of a multimedia base assessment of scientific inquiry abilities. *International Journal of Science Education, 37*(14), 2326–2357.

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

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.

Mannel, S. (2011). Assessing scientific inquiry. Development and evaluation of a test for the low-performing stage. Berlin: Logos Verlag.

Martin, M. O., Mullis, I. V.S., Gonzales, E. J., & Chrostowski, S. J. (2004) TIMSS 2003 International science report. International association for the evaluation of educational achievement. *TIMSS & PIRLS International Study Center*, Lynch School of Education, Boston College.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-Based science instruction – What is it and does it matter? Results from a research synthesis years 1984 – 2002.

Journal of Research in Science Teaching, 47(4), 474-496.

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2020). Umgang mit dem Coronavirus an Schulen. Retrieved from: https://www.schulministerium.nrw.de/ministerium/schulverwaltung/schulmail-archiv/archiv-2020/13032020-umgang-mit-dem-corona-virus

National Research Council. (1996). National science education standards. National committee for science education standards and assessment. Washington, DC: National Academy Press.

Nold, G., Hartig, J., Hinz, S., & Rossa, H. (2008). Klassen mit bilinugalem Sachfachunterricht. Englisch als Arbeitssprache. Klieme, E. (Ed.). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie.* Weinheim u.a.:Beltz, 451-457.

Nonte, S. (2013). Entwicklungen und Auswirkungen der Schulprofilierung an allgemeinbildenden Schulen in ausgewählten europäischen Ländern und Implementationsperspektiven für Deutschland. *International Review of Education, 59*(2), 243-262.

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.

OECD (2000). Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy. Paris : OECD Publishing

OECD (2017). PISA 2015 assessment and analytical framework science, reading, mathematic, financial literacy and collaborative problem solving, revised edition, PISA, OECD Publishing, Paris

OECD (2019). PISA 2018 assessment and analytical framework, PISA, OECD Publishing, Paris

Osterhaus, C., Koerber, S., & Sodian, B (2020). The science-p reasoning inventory (SPR-I): measuring emerging scientific-reasoning skills in primary school. *International Journal of Science Education, 42*(7), 1087-1107.

Piekny, J., & Mahler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology, 31,* 153-179.

Robert-Koch-Institute (2020). Risikobewertung zu COVID-19. Retrieved from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Risikobewertung.html

Roberts, D. A. (2007). Scientific literacy/science literacy. Abell, S.K. & Lederman, N.G. (Eds.), Handbook of research in science education, Mahwah, NJ: Erlbaum. 729-779.

Roberts, D. A. & Bybee, R. W. (2014). Scientific literacy, science literacy, and science education. Abell, S. K. & Lederman, N. G. (Eds.), *Handbook of research on science education*, *2,* 545-558.

Roth, A., Moll, C., Seidel, I. & Bös, K. (2017). Nachwuchsleistungssport an den NRW-Sportschulen – Talentsichtung unter Berücksichtigung sportartübergreifender, sportartspezifischer und sportpsychologischer Testverfahren. *Leipziger Sportwissenschaftliche Beiträge, 58*(1), 132-157.

Schiefer, J., Stark, L., Gaspard, H., Wille, E., Trautwein, U., & Golle, J. (2020). Scaling up an extracurricular science intervention for elementary school students: It works, and girls benefit more form it than boys. *Journal of Educational Psychology,* Advance online publication.

Schiepe-Tiska, A., Rönnebeck, S., Schöps, K., Neumann, K., Schmitdner, S. Parchmann, I. & Prenzel, M. (2016). Naturwissenschaftliche Kompetenz in PISA 2015 – Ergebnisse des internationalen Vergleichs mit einem modifizierten Testansatz. In: P**ISA 2015** : eine Studie zwischen Kontinuität und Innovation.

Schreiner, C. & Sjøberg, S. (2004). Sowing the seeds of ROSE. Background, rationale, questionnaire development and data collection for ROSE (The Relevance of Science Education) – a comparative study of students' views of science and science education (pdf) (Acta Didactica 4/2004). Oslo: Dept. of Teacher Education and School Development, University of Oslo

Schäfers, M-S., Schmiedebach, M., & Wegner, C. (2020). Virtuelle Labore im Biologieunterricht: Auswirkungen von Labster auf die Selbsteinschätzung von Schülerinnen und Schülern. *MedienPädagogik Zeitschrift für Theorie und Praxis der Medienbildung,* 140-167.

Schulte, T. (2017). *Desirable science education: Findings from a curricular delphi study on scientific literacy in germany*. Wiesbaden : Springer Fachmedien Wiesbaden

Schulte, A. & Wegner, C. (2020). Science-Klassen – Wie gelangt naturwissenschaftliche Förderung in die Schulen? *MNU – Journal, 73*(4), 309-312.

Schulte, A. & Wegner, C. (in prep.) Wie bewerten Lehrpersonen das Distanzlernen in den Science-Klassen?

Sjöström, J., Frerichs, N., Zuin, V. G., & Eilks, I. (2017). Use of the concept of *Bildung* in the international science education literature, its potential, and implications for teaching and learning. *Studies in Science Education, 53*(2), 165 – 192

Spörlein, E. (2003). *„Das mit dem Chemischen finde ich nicht so wichtig…": Chemielernen in der Sekundarstufe I aus der Perspektive der Bildungsdidaktik.* Wiesbaden : Springer Fachmedien

Walpuski, M. (2006): Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback. Eine empirische Studie. *Studien zum Physik- und Chemielernen.*Berlin: Logos Verlag.

Walpuski, M. & Schulz, A. (2011). Erkenntnisgewinnung durch Experimente – Stärken und Schwächen deutscher Schülerinnen und Schüler im Fach Chemie. *Chimica et ceterae artes rerum naturae didacticae. 37*(104), 6 – 27

Wellnitz, N., Fischer, H. E., Kauertz, A., Mayer, J., Neumann, I., Pant, H. A., Sumdleth, E. & Walpuski, M. (2012). Evaluation der Bildungsstandards – eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, *18*, 261 – 291.

Zachos, P., Hick, T. L., Doane, W. E., & Sargent, C. (2000). Setting theoretical and empirical foundations for assessing scientific inquiry and discovery in educational programs. *Journal of Research in Science Teaching, 37*(9), 938–962.

About the Authors

Alena Schulte is currently a PhD student at the Osthushenrich center for giftedness at the Faculty of Biology (OZHB) at Bielefeld University. In her project "Science-Klassen", concepts are developed to support scientifically gifted students in schools. The concepts are implemented in the schools and scientifically accompanied by her. In this context, she investigates the development of the students in the science classes and the design of the lessons by the teachers.

Dr. Claas Wegner is a professor at Bielefeld University, Germany, in the Department for Didactics of Biology. His research focuses on gifted students in science education and he is head of the just recently founded Osthushenrich center for giftedness (OZHB) at Bielefeld University.

---

[i] The technical terms in italics were translated from German into English.

[ii] One of the three types of secondary schools students may choose after primary school. This school type allows students to take an entrance exam to later study at university.