Differentiation and development of five levels in scientific inquiry skills: a longitudinal assessment of Biology students in grade 5 to 10

Andrea Möller\textsuperscript{1}, Stefan Hartmann\textsuperscript{2} and Jürgen Mayer\textsuperscript{2}

\textsuperscript{1} Institute of Science and Mathematics Education, University Vechta, Germany
\textsuperscript{2} Institute of Biology Education, University Kassel, Germany

This research was supported by a German Federal Ministry of Education & Research Grant awarded to Jürgen Mayer and by a Justus-Liebig University Giessen Young Researchers Award to Andrea Möller.

Correspondence concerning this article should be addressed to Andrea Möller, Institute of Science and Mathematics Education, University Vechta, Driverstrasse 22, 49377 Vechta, Germany. E-mail: Andrea.Moeller@uni-vechta.de
Abstract

The detailed developmental processes of scientific inquiry competence within the school environment have so far been studied only unsystematically. In the here presented study, we predict and measure five qualitative levels of scientific inquiry competence in the four central skills based on a differentiation through levels of complexity and a qualitative grading according to problem-solving processes. On the basis of 24 open test items a multi-matrix design was used to perform a longitudinal test on the inquiry competence of 1129 German students (age 10-16). We found all five predicted qualitative competence levels for each of the four inquiry skills present in the students’ test answers. However, the achievement of the highest two levels is low and inconsistent in distribution. The expected increase of inquiry competence within one school year is due to a significant qualitative increase within the skill levels. Detailed analyses reveal that performance differs significantly within the four skills. The results support a qualitative grading according to scientific problem-solving processes and provide more accurate information about the increase within the skills. The described threshold values for each competence level permit and facilitate a targeted choice of test items and thus provide a tool for gaining accurate insight in the individual student’s inquiry performance.

Keywords: scientific inquiry, school environment, qualitative levels, student’s performance, longitudinal test, Germany
Subject of the study

Providing students with competences of scientific inquiry represents a key goal of science education in general and of biology education in particular for many decades. However, results of international student assessment programs like TIMSS or PISA as well as several research studies in the field of science education show that German students in lower secondary biology education (grades 5-10) still lack competence especially in the field of scientific problem solving and science process skills (e.g. Mayer et al. 2003). It is speculated, that this results from the fact that planning and conducting experiments is either widely neglected or taught unsystematically in German schools (Mayer 2004). As one reaction to these findings, inquiry competence was being officially integrated in the 2004 German Federal Education Standards for Biology (KMK 2004). However, recent studies show that students of all ages still lack crucial competences in the process of scientific inquiry. They plan their experiments unsystematically, do not control the variables and hardly ever include experimental controls (Hammann et al. 2008) and over 50% do not meet the requirements in all four central skills “formulating questions”, “generating hypotheses”, “planning an investigation”, and “interpreting data” (e.g. Möller et al. 2008, 2009).

In order to address the students’ problems and to create a learning environment that encourages the individual to participate in the science process, it appears crucial to focus on finding out how inquiry competences can be described and assessed in detail.

Scientific inquiry is mainly described as a problem-solving process that involves several relevant skills (e.g. NRC 1996, Klahr 2000, Gott & Duggan 2001, Abd-El-Khalik et al. 2004, Mayer 2007). Möller, Grube & Mayer (2008) were able to empirically support the assumption that scientific inquiry can be differentiated in four central skills: “formulating questions”,
“generating hypotheses”, “investigation planning”, and “interpreting data”. However, the exact developmental processes of scientific inquiry competence within the school environment remain subject to further study. Particularly, the analysis of qualitative levels within the mentioned four central skills awaits detailed study. In order to establish a grading of competence levels within inquiry skills, several criterions are suggested. For example, Bybee (2002) proposes to use a gradual shift of students’ concepts of everyday life to scientific concepts. Others focus on the students’ use of relevant variables (Schauble et al. 1992) and the more or less systematic approach during investigation (Hammann 2004). In the here presented study, we predict and measure five qualitative levels of inquiry competence which are based on a differentiation through levels of complexity and a qualitative grading of each skill according to problem-solving processes (Mayer 2007, Möller & Mayer 2008). Longitudinal studies over all grades in lower secondary biology education show an overall significant increase of inquiry competence (Möller et al. 2009). However, it is not clear to which extend this is due to a qualitative increase within the skill levels.

Therefore our main research questions are: 1) Can all five predicted qualitative competence levels for each of the four inquiry skills be found in the students’ test answers? 2) Is the expected increase of inquiry competence within one school year due to a qualitative increase within the skill levels? 3) How are the competence levels distributed within the four subskills?
Design of the study

In this study a longitudinal test design with two test points over one school year was used in order to test the levels and development of students’ inquiry competence. A paper-and-pencil test was conducted with 1129 German students from grade 5-10 (age 10-16) at the start and end of the school term, respectively. Both tests were conducted in a multi matrix design with 24 open test items, equally representing the four inquiry skills “formulating questions,” “generating hypotheses,” “planning an investigation,” and “interpreting data,” which were already confirmed in previous studies. Each individual test included six items. Data were analyzed with the help of specific coding schemes for each of the four skills (intercoder reliability: $Cohens Cappa = 0.88 - 1.0$ for all proposed skills). Up to 14 individual codes (ranked in degree of difficulty) were combined to five assumed levels of competence according to increased level of difficulty and problem-solving processes. Data from both tests were entered and analyzed using the probabilistic Rasch model (Rasch 1960). The model is used for so called “person-item maps” providing equal linear (interval) measures in logit units of student performance (here in scientific inquiry) and item difficulty. Because these scales are invariant they allow researchers to confidently compare results over time (see Boone & Scantlebury 2006 for details). In this study, the program ConQuest (Wu, Adams & Wilson 1997) was used for item-response (Rasch) modelling. In order to achieve threshold values for each defined skill level (level 1 - 5) we defined the threshold at 50%, meaning that a student must be able to solve at least half of the items within one skill in order to achieve the next level. This method is used in accordance to the definition of threshold values in international large-scale student assessment projects, such like TIMSS and PISA. Further analyses were conducted using the Statistical Software Program Package (SPSS) 15.0.
Findings and Analysis

45.1% of all students across all grades achieve the first competence level, while 11.8% are below level 1, meaning that they do not reach any level of scientific inquiry competence aimed for at the beginning of the school term. 37.7% reach the second, and 5.4 the third level of competence, the latter representing overall the requirements of the national science education standards. No student reaches level 4 and 5. At the end of the term, the percentage of students that are below level 1 decreases significantly to 8.1 (p < .001, Wilcoxon). Also, the number of students that reach level 1 decrease significantly from 45.1 to 31.2% (p < .001, Wilcoxon). At the same time the percentage of students that reach level 2 and 3 increase significantly over all four inquiry skills (8 and 8.8%, respectively; both p < .001, Wilcoxon) (see fig. 1). At the second test 0.8% of the students reach level 4. Although some students reached level 5, numbers were low and highly inconsistent in distribution among grades. So far, it cannot be included in the Rasch-Model. In all four inquiry skills and throughout all age groups, less than 50% of all students achieved level 3.

A more detailed analysis of the inquiry competence at the end of the school term reveals that achieved levels differ significantly between the four subskills “formulating questions,” “generating hypotheses,” “planning an investigation” and “interpreting data” and thus further support that they represent four independent domains of one underlying factor which is assumed to be scientific inquiry (Grube et al. 2007) (p < .001, Wilcoxon, see fig. 2). Students performed best when they were planning an investigation, where 18.3% reached level 3 and 3% level 4. However, there are still 24.7% that did not even achieve level 1. 69.3% reached level 1, 15.5% level 2 and 2.5% level 3 when generating hypotheses. The amount of students that reached level 3 was second highest in interpreting data, 35.2% reached level 2 and almost 50% level 1. Over
all age groups, formulating scientific questions posed the most difficulties to the students. 35 % did not even reach level 1, 41% level 1 and 24% level 2. Note that some students did show performance on level 3, 4 and 5 in that inquiry skill but numbers were so low that again they could not be included in the analyses.

Conclusions and Implications/Contribution to the teaching and learning of science

First, the five predicted qualitative competence levels for each of the four inquiry skills were all present in the students’ test answers. Therefore, the results of this study support the predicted qualitative grading of each of the four skills according to scientific problem-solving processes (Möller/Mayer 2009). The open test items used in this test can be applied on all accounts to evaluate inquiry competence up to level 3 that represents the requirements of national education standards. Competence level 4 was only achieved in few cases and a too small amount of students performed on level 5 to be included in the model. We believe that those two highest competence levels cannot be evaluated by paper pencil tests alone but probably need other assessment methods such like portfolios or video analysis. The results also support the assumption that inquiry competence is not affected by age alone but can be acquired in the classroom.

Second, the expected increase of inquiry competence within one school year is due to a significant qualitative increase within the skill levels (p <.001 over all included levels).

Third, the detailed analyses show that there are significant differences in performance between the four subskills of scientific inquiry. Formulating scientific questions seems to be
especially challenging to the students. Why students perform so poorly in this particular skill is not entirely clear yet. Information on classroom setting and learning environment that will be collected in addition to the inquiry tests during a third testing time will presumably provide a more detailed picture. In interviews, teachers who participated in this study report that most of the time they start a new investigation by providing the question in order to save time. According to them this happens especially often within the last two years (grade 9 and 10) when the biology curriculum becomes more challenging. While the results in formulating scientific questions were especially alarming, students’ performance in all four skills strongly suggest more biology lessons (and possibly other science lessons) that integrate the process of scientific inquiry.

Overall, our results show that in addition to a prediction about the overall competence increase of each student ($b$-value), our model allows more precise information about the qualitative increase within the skills. For science teachers, an accurate assessment of students’ scientific inquiry skills is crucial in order to gain insight in the progress of each individuals’ performance. It facilitates lesson planning and enables the teacher to provide accordant feedback, as well as distribute adequate exercise units. With a greater knowledge of students’ difficulties and needs, novel teaching concepts can be developed that are possibly adjusted to different performance levels of scientific inquiry and still meet the actual classroom setting. In addition, the defined threshold values for each competence level permit and facilitate a more targeted choice of test items in assessments.
References


Figure 1. Reached levels of students’ scientific inquiry competence over all inquiry skills and grades 5-10 (N = 1129) at the beginning (T1) and end of the schoolterm (T2).
Figure 2. Detailed distribution of levels in the four skills of scientific inquiry at the end of the school term (N = 1129). The reached levels differ significantly over all skills (p < .001, Wilcoxon).