

Science Education Review Letters

Chemistry Student Teachers' Scientific Inquiry Competencies

Jurik Stiller^{a*}, Andreas Nehring^a, and Rüdiger Tiemann^a

^a Chemistry Education, Department of Chemistry, Humboldt-Universität zu Berlin, Germany

* Corresponding author: jurik.stiller@hu-berlin.de

Received 21st March 2013, Accepted 12th November 2013

Abstract

We constructed paper-and-pencil test-items to assess chemistry student teachers' scientific inquiry competencies. We focused on the inquiry method experiment, divided into three *reasoning* steps (question & hypothesis; planning & performing; analysis & reflection). Test revision resulted in a test of 20 items, which was applied to a sample of undergraduate and graduate chemistry student teachers. The first explorative assessment ($N = 89$) based on that instrument (1) revealed acceptable item characteristics, such as an appropriate range of item difficulties ($.48 < P_i < .84$) and only slightly unsatisfactory item discrimination parameters ($.19 < r_{it} < .65$). The study (2) obtained preliminary information on the high prognostic potential of the planned degree.

Keywords

Scientific Inquiry, Science Teaching, Chemistry Education, Higher Education, Teacher Education

Introduction

Modeling and assessing competencies is currently promoted as the basis for policy control in Germany (e.g., Klieme, Leutner, & Kenk, 2010), particularly in the area of primary and secondary education. This is valid for students in different stages of their learning, as well as for teachers in different stages of their professional development. Most recently, a transfer of the term competency to the field of teacher education can be recognized. Von Aufschnaiter and Blömeke (2010) demand competence-oriented science teacher education research (see also Schecker & Parchmann, 2006). A fundamental base for students' participation in society can be provided through developing scientific literacy (American Association for the

Advancement of Science [AAAS], 1993). Of particular importance are the skills in terms of scientific inquiry (National Research Council [NRC], 2012). It can be concluded, that a teacher needs to be at least as capable as a student to successfully provide productive learning environments. Hence, looking at the teacher, the scientific inquiry competencies can be understood as a component of teachers' professional knowledge (see, e.g., Baumert & Kunter, 2006). These have mainly been investigated in the field of primary and secondary education. However, in higher education, so far only a few studies exist (see for example in physics Woitkowski, Riese, & Reinhold, 2011, or in chemistry Mamlök-Naaman, Taitelbaum, Carmeli, & Hofstein, 2006). For an evaluation of skills in

scientific inquiry in chemistry there is a lack of adequate measurement instruments. Thus, the aim of this study is the construction of a multiple-choice test instrument for assessing chemistry student teachers' scientific inquiry competencies, with sufficient item characteristics, average item-difficulties and acceptable item discrimination.

Theoretical Background

The competencies in the field of scientific inquiry as a part of Nature of Science (Lederman, 2007; McCormas & Olson, 1998) play a key role in the enactment of scientific literacy (AAAS, 1993; see also NRC, 2012). Several studies dealt with these competencies as well as their implementation in science classrooms (e.g. Martin, Mullis, Foy, & Stanco; 2012; Möller, Grube, & Mayer, 2006). Research in the field of scientific inquiry, however, lacks domain-specificity („science“ in the international community), or lacks a focus on comparing the achievement in different domains (Breslyn & McGinnis, 2011). Hence, based on a literature review and empirical findings, Nehring, Nowak, Tiemann, and Upmeier zu Belzen (2012) operationalized scientific inquiry competencies for chemistry and biology. The authors have the goal of identifying patterns of acquisition of knowledge (scientific inquiry) in biology and chemistry classes and the application of a competence structure model valid for both subjects. The model incorporates two dimensions, the *inquiry methods* on the one hand, and the *scientific reasoning* on the other. These two dimensions are valid as part of acts of knowledge acquisition (scientific inquiry) for both subjects (ib.; see Hacking, 1996). That operationalization was adapted, concentrating on a singular *inquiry method*, the experiment.

strument for assessing chemistry student teachers' scientific inquiry competencies, with sufficient item characteristics, average item difficulties and acceptable item discrimination.

In addition, the instrument shall be used to reveal findings concerning the correlation of scientific inquiry competencies with control variables such as domain knowledge, fluid intelligence (Sternberg, 2010), planned degree or self-ratings of students in terms of their performance. Finally, the influence of cohort (belonging) on scientific inquiry competencies shall be investigated.

Design and Methodology

To assess domain knowledge as covariate a multiple-choice (single-select) item test containing 16 items was developed. The items deal with various contexts of chemistry, such as carbohydrates, thin layer chromatography and solubility of gases in water. Each of the contexts of the items in the domain knowledge test is related to a context of the items in the competence test.

Moreover, to assess scientific inquiry competencies of chemistry student teachers a multiple-choice (single-select) item pool was generated, which contained at least 9 items for each cell of the underlying competence structure model (E1: 13 items; E2: 9; E3: 10). Those 32 items were included in individual test booklets. After a test revision and, hence, selection of items, it was used to assess chemistry student teachers' ($N = 89$; age: $M = 23.5$ years, $SD = 3.5$; undergraduate and graduate studies) scientific inquiry competencies according to the competence structure model shown in figure 1.

E1	formulating questions & hypotheses for experiments	E2	planning & performing of experiments	E3	analysis of exptl. data & reflection of experiments
-----------	--	-----------	--------------------------------------	-----------	---

fig. 1 Competence structure model of scientific inquiry through experiments (adapted; see Nehring et al., 2012; Mayer, 2007; NRC, 2012)

Scientific inquiry is assumed to be independent from domain knowledge (Nehring et al., 2012).

Aim

The aim of this explorative small-scale study is the construction of a multiple-choice test in-

The construction was based on a specific pattern: each item consisted of a stem, including information about the context and all required information to choose the correct option. According to Nentwig, Roennebeck, Schoeps, Rummann and Carstensen (2009) this is called high level of contextualization (hi-con). Thereby, the

solution of the items was assumed to be independent from prior domain knowledge. For each item a set of 4 options was given. The distractors were exclusively self-generated. The students were supposed to choose the right or, if they didn't think any option was right, at least the most appropriate option.

The items dealt with experiments in various fields of chemistry, such as thin layer chromatography, distillation, and substitution reactions. These fields are assumed to represent both the basic contents of chemistry (i.e. inorganic and organic chemistry) and fields, which the tested undergraduate students already are familiar with.

sample or not normally distributed trait (p. 93). In this case all possible reasons may have occurred – the item construction was theory-driven but susceptible to errors, and the sample was indeed very heterogeneous (including students of undergraduate and graduate studies). It's hard to judge the normal-distribution so far; other research projects, however, rely on the assumption of normal distribution of scientific inquiry-skills (e.g. Möller, Grube, & Mayer, 2006).

The test for assessing domain knowledge revealed an internal consistency of $\alpha = .588$, which is only marginally below the limit proposed by some authors for studies in empirical social sciences (Diekmann, 2007).

<p>E1</p> <p>4 thin layer chromatography 2 distillation 1 reactions of nitric acid 1 reactions of zinc 5 selectivity of subst. reactions $\Sigma = 13$</p>	<p>E2</p> <p>3 thin layer chromatography 2 solubility of gases 2 reactions of nitric acid 1 reactions of zinc 1 selectivity of subst. reactions $\Sigma = 9$</p>	<p>E3</p> <p>6 thin layer chromatography 2 distillation 1 reactions of nitric acid 1 selectivity of subst. reactions $\Sigma = 10$</p>
---	---	---

fig. 2 Contexts of the developed items and number of items per competence facet

Afterwards, data were coded dichotomously and analyzed by means of classical test theory. With the aim of testing for normal distribution, a Kolmogorov-Smirnov-test (K-S-test) was calculated (Janssen & Laatz, 2007). K-S-tests are applicable for small samples. A level of significance of $\alpha = .05$ was applied (Bortz, 2005). Cronbach's alpha was calculated as an indication of internal consistency according to Cronbach (1951), as well as item difficulties and item discrimination (Moosbrugger & Kelava, 2012). A relevant assumption is that all items measure the same construct. This is given for the test for scientific inquiry – although the subscales may constitute competence facets. In order to prevent an under- or overestimation of internal consistency, according to Schmitt's (1996) suggestion, Pearson subscale-inter-correlations were calculated as well. With the purpose of examining the relationships with covariates, a linear regression analysis was conducted.

Findings

The test scores are normally distributed ($KS-Z = 1.139$; $p = .149$), although with a skewness of -1.191 ($SE = .327$) and an excess of 1.364 ($SE = .644$). According to Moosbrugger and Kelava (2012) three plausible reasons can cause that: deficient item construction, a heterogeneous

The items constructed for assessing scientific inquiry competencies showed ambivalent item parameters. While 10 items could not be selected due to item difficulty indices of $P_i > .84$, two more items showed an unacceptable item discrimination of $r_{it} < .10$. These limits were chosen concordant with Moosbrugger and Kelava (2012), whereby the broadly accepted upper limit for the item difficulty of $P_i = .80$ was slightly raised in order to examine the trait of scientific inquiry even between students in the extreme groups (p. 87).

Partly notwithstanding the previous assumptions, the remaining items covered an acceptable range of item difficulty ($.48 < P_i < .84$). However, the test was too easy. Detailed information about the item characteristics with regard to the competence facets are provided in table 1 (see next page).

The resulting test for assessing scientific inquiry competencies with 20 items showed a test difficulty index of $.75$ ($M = 14.98$; $SD = 3.94$). It moreover revealed a good internal consistency of $\alpha = .812$. The index for internal consistency is dependent on the number of items and could, thus, be overestimated (Urbina, 2004).

The expectable subscales E1, E2 and E3 could

table 1 Item Characteristics (P_i : item difficulty; Var : item variance; r_{it} : item discrimination)

Competence Facet (Subscales) ¹	P_i	Var	r_{it}
E1 (6 items)	.64 - .84	.102 - .222	.243 - .507
E2 (7 items)	.52 - .84	.102 - .250	.192 - .557
E3 (7 items)	.48 - .84	.102 - .254	.230 - .648

¹ after test-revision: 12 items were excluded (E1: 7 items; E2: 2 item; E3: 3 item)

more or less sufficiently be reproduced: $\alpha_{E1} = .682$, $\alpha_{E2} = .529$, $\alpha_{E3} = .724$. Also, the subscale-inter-correlation revealed significant interrelations between the three subscales, though on a smaller level than expected. That may indicate a three-dimensional structure.

table 2 Subscale-Inter-Correlation

	E1	E2	E3
E1			
E2	.429**		
E3	.365**	.613**	

** = $p < .01$

To assure the expected dimensionality and due to the relatively small inter-correlations, a factor analysis was calculated. The items load on mainly four factors (elicited by exploratory factor analysis), which are not immediately recognizable in terms of the theoretical model. The items load, independent from context as well as almost independent from the competence facet they have been constructed for, on the four factors (determined by confirmatory factor analysis).

In order to examine the relationships with covariates and the assumed relationship with cohort belonging, a linear regression analysis was conducted. Self-ratings of students in terms of their performance, age, fluid intelligence, domain knowledge and gender revealed no prognostic potential.

A significant influence could only be found for one predictor of the model ($F(6,28) = 3.449$,

$p < .016$, $R^2 = .496$, $R^2_{corrected} = .352$), the planned degree. Results are shown in table 3 (see end of page).

Discussion

The constructed instrument for assessing domain knowledge shows a slightly unsatisfactory internal consistency. This could be caused by the multidimensionality of the underlying construct (Schermelleh-Engel & Werner 2012) and is, hence, acceptable for domain knowledge tests.

Regarding the test for assessing scientific inquiry there of course is a limitation concerning the validity: In this case due to the measurement of procedural knowledge/skills by applying a multiple-choice test, which may appear inappropriate. However, the application of multiple-choice test instruments tends to result in higher test interpretation objectivity and test economy. Hence, this format was preferred in this study.

The items in the test constructed to assess scientific inquiry competencies are generally too easy, while item discrimination seems to cover an acceptable range. It is striking that almost all items with item difficulty indices of $P_i > .90$ are allocated in the competence facet E1 (formulating questions & hypotheses for experiments). This could be caused by hierarchical complexity (see Commons et al., 2007), which has not been a-priori-controlled completely. A post hoc analysis showed, that the five easiest items all were addressing less complex cognitive operations – such as the so called “one fact” (see Neumann, Fischer, & Kauertz, 2010). Oppositely, the three items with item difficulty indi-

table 3 Results of linear regression analysis (significant predictor)

Predictor	B	β	SE	T	p
planned degree ²	1.162	.552	.552	2.105	$p < .05$

² 0 = no teaching degree planned

ces of $P_i < .30$ all addressed the competence facet E3 (analysis & reflection of experiments). This may be induced by the difficulty of interpreting tables and figures, a cognitive operation that the solution of these items required. However, the findings of explanatory factor analysis and confirmatory factor analysis are inconsistent, do not show the expected factor loadings (each competence facet representing a factor) and contribute to a need of further research.

The findings concerning the internal consistency of the competence-test are surprising, in this respect: high inter-correlation increase the α -estimate of internal consistency (Schermele-Engel & Werner, 2012). While the inter-correlation is quite small ($.458 < r < .584$), the α ($= .81$) is acceptably high. That, in line with the inconsistent results of the factor analysis, is a first indication of a one-dimensional competence structure, while the differential facets need to be investigated more precisely in follow-up studies.

The results of the linear regression analysis are surprising, especially concerning the non-predictive influence of fluid intelligence. Moreover, they contradict common findings in other projects (see for example Danner, 2011). That might be due to the relatively easy items in the scientific inquiry test and the relatively small internal consistency in the domain knowledge test. Thus, the (missing) effects might be caused by the instrument itself.

Summing up, the constructed test is an objective and reliable instrument for the assessment of scientific inquiry competencies. Further research might stress the a-priori control of item characteristics and a bigger sample. Plus, an expansion of the research to different *inquiry methods*, such as *usage of models*, and on the other science subjects biology and physics, could be useful. Also, as Kremer (2010) shows, an additional assessment of the views concerning the nature of science (i.e. epistemological beliefs concerning the processes of science) might add explanatory information.

Acknowledgment

We would like to thank Stefan Hartmann for his advice and two anonymous reviewers for their comments and suggestions that were essential for improving the quality of the paper.

References

- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York, Oxford University Press.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Professional competencies of teachers]. *Zeitschrift für Erziehungswissenschaft*, 9(4), pp. 469 - 520.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* [Statistics for human and social scientists]. Heidelberg: Springer.
- Breslyn, W., & McGinnis, J. R. (2012). A comparison of exemplary biology, chemistry, earth science, and physics teachers' conceptions and enactment of inquiry. *Science Education*, 96(1), 48 - 77.
- Commons, M. L., Goodheart, E. A., Pekker, A., Dawson, T. L., Draney, K., & Adams, K. M. (2007). Using Rasch scaled stage scores to validate orders of hierarchical complexity of balance beam task sequences. In E. V. Smith, Jr. & R. M. Smith (Eds.), *Rasch measurement: Advanced and specialized applications* (pp. 121 - 147). Maple Grove, MN: JAM Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297 -334.
- Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*, 39(5), 323-334.
- Diekmann, A. (2007). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendung* [Empirical research in the social sciences. Principles, methods, application]. Hamburg: Rowolth.
- Grube, C., Möller, A., & Mayer, J. (2007). Dimensionen eines Kompetenzstrukturmodells zum Experimentieren [Dimensions of a competence structure model for scientific inquiry]. In H. Bayrhuber, U. Harms, D. Krüger, A. Sandmann, A. Unterbrunner, A. Upmeyer zu Belzen, & H. Vogt (Eds.), *Ausbildung und Professionalisierung von Lehrkräften* [Education and professionalisation of teachers]. International conference of Fachgruppe Biologiedidaktik in Verband Biologie, Biowissenschaften & Biomedizin (VBio) (pp. 31 - 34). Kassel: Universitätsbibliothek Kassel.
- Hacking, I. (1996). *Einführung in die Philosophie der Naturwissenschaften* [Introduction to the philosophy of science]. Stuttgart: Reclam.
- Janssen, J., & Laatz, W. (2007). *Statistische Datenanalyse mit SPSS für Windows: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests* [Statistical Analysis with SPSS for Windows: An application-oriented introduction to the basic system and the module „Exact tests“]. Berlin, New York: Springer.
- Klieme, E., Leutner, D., & Kenk, M. (2010). Kompetenzmodellierung [Modeling of competence]. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung. Eine aktuelle Zwischenbilanz des*

- DFG-Schwerpunktprogramms [Competence modeling. A Recent report of the DFG-programme] (pp. 9 - 11). Weinheim, Basel: Beltz.
- Kremer, K. H. (2010). *Die Natur der Naturwissenschaften verstehen - Untersuchungen zur Struktur und Entwicklung von Kompetenzen in der Sekundarstufe I* [Understanding the nature of science - investigations concerning structure and development of competencies in secondary school]. Doctoral Dissertation. Kassel: Universitätsbibliothek Kassel.
- Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831 - 879). Mahwah: Erlbaum.
- Mamluk-Naaman, R., Taitelbaum, D., Carmeli, M., & Hofstein, A. (2006). A model of professional development of chemistry teachers. In I. Eilks & B. Ralle (Eds.), *Towards research-based science teacher education*. Proceedings of the 18th Symposium on Chemical and Science Education held at the University of Bremen in June 2006 (pp. 113 - 124). Aachen: Shaker.
- Martin, M.O., Mullis, I.V.S., Foy, P., & Stanco, G.M. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mayer, J. (2007): Erkenntnisgewinnung als wissenschaftliches Problemlösen [Scientific inquiry as scientific problem-solving]. In D. Krüger & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* [Theories of research in biology education. A handbook for student teachers and doctoral students] (pp. 177 - 186). Berlin, New York: Springer.
- McComas, W.F., & Olson, J.K. (1998). The nature of science in international science education standard documents. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 41 - 52). Dordrecht: Kluwer Academic.
- Möller, A., Grube, C., & Mayer, J. (2006, September). *Inquiry Competence in German Biology Education: First Results of the National Research Project "biology in context"*. Paper Presented at the meeting of European Researchers in Didactics of Biology (ERIDOB), London.
- Moosbrugger, H., & Kelava, A. (Eds.). (2012). *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion* [Test theory and construction of surveys]. Berlin, Heidelberg: Springer Berlin Heidelberg.
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Nehring, A., Nowak, K. H., Upmeyer zu Belzen, A., & Tiemann, R. (2012). "VerE"-Studie – Aufgabenentwicklung für eine modellbasierte Erfassung von Schülerkompetenzen im Bereich der Erkenntnisgewinnung des Chemie- und Biologieunterrichts [VerE-study – construction and evaluation of test-items for model-based assessment of scientific inquiry competencies of students in biology and chemistry lessons]. In S. Bernholt (Ed.), *Konzepte fachdidaktischer Strukturierung für den Unterricht* [Concepts of didactically structuring lessons] (pp. 269 - 271). Münster: LIT.
- Nentwig, P., Rönnebeck, S., Schöps, K., Rumann, S., & Carstensen, C. (2009). Performance and levels of contextualization in a selection of OECD countries in PISA 2006. *Journal of Research in Science Teaching*, 8, 897 - 908.
- Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards - the impact of large scale assessments on science education research in Germany. *Journal of Mathematics and Science Education*, 8(3), 545 - 563.
- Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz [Modeling of scientific competence]. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45 - 66.
- Schermelleh-Engel, K., & Werner, C. (2012). Methoden der Reliabilitätsbestimmung [Methods of measuring reliability]. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* [Test theory and construction of surveys] (pp. 119 - 141). Berlin, Heidelberg: Springer.
- Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assessment*, 8(4), 350 -353.
- Sternberg, R.J. (2010). Intelligence. In P. Peterson, E. Baker, & B. McGaw (Hrsg.), *International encyclopedia of education (Third edition)*. Oxford: Elsevier.
- Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: Wiley.
- Veal, W. (2004). Beliefs and knowledge in chemistry teacher development. *International Journal of Science Education*, 26(3), 329 - 351.
- Von Aufschnaiter, C., & Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata [Assessing professional competencies of (pre-service) teachers – Desiderates]. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 361 - 367.
- Woitkowski, D., Riese, J., & Reinhold, P. (2011). Modellierung fachwissenschaftlicher Kompetenz angehender Physiklehrkräfte [Modeling of competence of prospective physics teachers in the field of content knowledge]. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 315 - 337.