KoKoHs

Modeling and Measuring Compentencies in Higher Education
- Validation and Methodological Innovations

KoKoHs Working Papers
No. 11

Hans Anand Pant, Olga Zlatkin-Troitschanskaia,
Corinna Lautenbach, Miriam Toepper & Dimitar Molerov (Eds.)

Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations (KoKoHs) Overview of the Research Projects

SPONSORED BY THE

Federal Ministry
of Education
and Research

Johannes Gutenberg University Mainz

Humboldt University of Berlin

KoKoHs Working Papers on the BMBF funding initiative "Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations"

The *KoKoHs Working Papers* series publishes papers from the BMBF funding initiative "Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations" (KoKoHs), including project information, conceptual papers, and preliminary results that are intended for rapid dissemination and public discussion. Publication of a text as a *KoKoHs Working Paper* does not preclude publication elsewhere. Responsibility for the content lies with the authors. The content does not necessarily reflect the views of the editors or the Scientific Transfer Project of KoKoHs.

Contact

corinna.lautenbach@hu-berlin.de
miriam.toepper@uni-mainz.de
molerov@hu-berlin.de

The program is funded by the Federal Ministry of Education and Research under grant numbers 01PK15100A and 01PK15100B.

The *KoKoHs Working Paper series* is also available for download:

http://www.kompetenzen-im-hochschulsektor.de

Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations (KoKoHs)

Overview of the Research Projects

Pant, H. A., Zlatkin-Troitschanskaia, O., Lautenbach, C., Toepper, M. & Molerov, D. (Eds.)

Contact:

corinna.lautenbach@hu-berlin.de
miriam.toepper@uni-mainz.de
molerov@hu-berlin.de

Bibliographical references:

Pant, H. A., Zlatkin-Troitschanskaia, O., Lautenbach, C., Toepper, M. & Molerov, D. (Eds.) (2016). Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations (KoKoHs) – Overview of the Research Projects (KoKoHs Working Papers, 11). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations (KoKoHs) – Overview of the Research Projects

Abstract

The 15 collaborative projects of the BMBF funding initiative "Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations (KoKoHs)" aim to make a significant contribution to advancing the field of modeling and valid measurement of academically acquired competencies. Based on findings, competency models, and assessment instruments from the first funding phase, the new projects assess competencies in three broad clusters, including (1) generic competencies, (2) competencies in teacher education, and (3) domain-specific competencies. Measurement approaches are refined, validated, and tested for use in higher education in Germany. Research and results of KoKoHs projects are systematically integrated to increase their visibility and compatibility with national and international research. Furthermore, results and findings are transferred on various levels into research, higher education practice, and educational policy.

Key Words

Competency models, Competency assessment, Test instruments, Higher education

Contents

Introduction

In recent years, higher education in Germany has undergone far-reaching educational reform (e.g., Bologna reform, new governance model). Comprehensive restructuring processes have been initiated to address inherent problems of higher education. In recent years, we have in parts seen an aggravation of problems and an increase in challenges in higher education, including considerable social preselection in the new Bachelor and Master degree courses (Zlatkin-Troitschanskaia et al. 2012; Maaz et al. 2014), lack of educational justice for students with immigrant family histories as well as gender effects in various disciplines, high failure rates as indicated, for instance, by dropout rates and long study durations (National Report on Education, Authoring Group 2014, 2015). Debates are increasingly centering on the quality, effectiveness, and efficiency as well as the individual and societal returns of higher education. Many key questions cannot be answered without empirical data on the influences, development, design, and effects of academic teaching-and-learning processes. To this end, intended outcomes of academic learning processes, such as competencies acquired in higher education need to be defined, conceptualized in a model, and assessed in a reliable and valid way.

Funded by the German Federal Ministry of Education and Research, the national research program „Modeling and Measuring Competencies in Higher Education – Validation and Methodological Challenges (KoKoHs)“ focuses on current practical and policy-related challenges of competency orientation in teaching and learning in higher education and contributes to addressing research deficits in current assessment practice (Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper & Lautenbach, 2016). In the previous KoKoHs initiative from 2011–2015, experts from various study disciplines, teaching methodologies, and measurement research collaborated in a multi-disciplinary mixed-methods research program. The program funded projects that modeled and assessed academically taught generic or domain-specific competencies in the study domains of engineering, business and economics, educational science, and teacher education – including teaching of STEM subjects. Altogether 24 collaborative projects developed and tested empirically competency models and instruments for valid assessment of academically acquired competencies in Germany (Zlatkin-Troitschanskaia et al. 2014). The results indicated that the developed competency models and instruments in various disciplines offer a sound basis for reliable and valid competency assessment as well as for prospective in-depth, longitudinal, field-experimental validation studies with multilevel analyses.

In the new KoKoHs initiative, running from 2016–2020[1], 15 collaborative projects systematically address central challenges of advanced competency research in higher education. The projects examine how, despite high conceptual and methodological requirements, we can ensure objective, reliable, and valid assessment of academic competencies. Thus, all collaborative projects aim to make significant contributions to advancing the still underresearched field of modeling and valid assessment of academically acquired competencies and to highlight effective approaches to addressing major challenges in current higher education practice (Zlatkin-Troitschanskaia et al., 2015). The collaborative projects focus on assessing competencies in three broad clusters:

1) Generic competencies,
2) Competencies in teacher education, and
3) Domain-specific competencies (medicine and business & economics).

As to the methods, the projects not only use multilevel analyses and longitudinal study designs in field-experimental validation studies, but moreover use and develop further internationally innovative

[1] For more information, see https://www.kompetenzen-im-hochschulsektor.de

methods of competency assessment, such as computer-based adaptive testing. In most projects, one major focus of the analyses is on establishing prognostic validity of the assessment. Two out of three KoKoHs projects have longitudinal study designs and examine competency development through change trajectories over the course of studies. By controlling for individual and study-related influence factors, the projects gather evidence on the effects of, for instance, certain learning opportunities on students' competency acquisition. These studies are relevant not only for current competency research, but also for improving higher education teaching and examination practice.

Substantial research progress is not possible without systematically integrating the research and results of individual projects so as to make them visible and compatible with national and international research. To this end, overarching scientific work and meta studies are necessary to aggregate the findings and results from individual projects and transfer them on various levels into higher education research, practice, and policy.

The necessary measures are taken by the Scientific Transfer Project of KoKoHs, directed by Prof. Hans Anand Pant at Humboldt University of Berlin and Prof. Olga Troitschanskaia at Johannes Gutenberg University Mainz. The activities of the Scientific Transfer Project serve to channel the overall research output of the KoKoHs program, to increase internal and external cooperation, optimize dissemination of results and positioning of KoKoHs research within the German and international research communities, and to support the implementation of research results and findings in higher education practice and policy.

KoKoHs Working Paper No. 11 offers an overview of the 15 collaborative projects of the KoKoHs funding initiative, outlining their theoretical approaches, central research questions, study designs, and expected outcomes. For each project, recommended literature, project details, and contact information are provided, as well, to encourage more in-depth exchange.

References

Authoring Group Educational Reporting (2014). Education in Germany 2014 - An indicator-based report including an analysis of the situation of people with special educational needs and disabilities. Bielefeld: W. Bertelsmann.

Maaz, K., Neumann, M. & Baumert, J. (Eds.) (2014). *Herkunft und Bildungserfolg von der frühen Kindheit bis ins Erwachsenenalter. [Origin and educational success from early childhood through to adulthood.]* Wiesbaden: Springer VS.

Zlatkin-Troitschanskaia, O., Pant, H. A., Kuhn, C., Toepper, M. & Lautenbach, C. (2016). *Messung akademischer Kompetenzen von Studierenden und Hochschulabsolventen – Ein Überblick zum nationalen und internationalen Forschungsstand*. Wiesbaden: Springer VS.

Zlatkin-Troitschanskaia, O., Shavelson, R. J. & Kuhn, C. (2015). The International State of Research on Measurement of Competency in Higher Education. *Studies in Higher Education*, 40(3), 393-411. DOI: 10.1080/03075079.2015.1004241

Zlatkin-Troitschanskaia, O., Kuhn, C. & Toepper, M. (2014). Modelling and Assessing Higher Education Learning Outcomes in Germany. In H. Coates (Ed.). *Higher Education Learning Outcomes Assessment – International Perspectives*. Frankfurt/Main: Peter Lang. 213-235.

Cluster 1: Generic Competencies

ASTRALITE

Assessment and Training of Scientific Literacy

KomPrü

Performance-based Assessment of Students' Communication Skills

MultiTex

Process-based Assessment of Multiple Documents Comprehension

PRO-SRL-EVA

Product- and Process-Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation

ValiDiS

Competence Modeling and Assessment: Validation of the Test Score Interpretations of a Scientific Reasoning Test

ASTRALITE – Assessment and Training of Scientific Literacy

Schmid, S., Richter, T. & Wild K.-P.

Abstract

In the ASTRALITE project, tests for the assessment of student competencies when dealing with scientific literature are validated by analyzing the declarative and procedural knowledge that is necessary to employ certain processing strategies. Systematic and receptive strategies are considered in connection to epistemic and heuristic strategies. The validation approach is based on training experiments, expert-novice studies, and ecologically valid achievement data.

Aims and Research Questions

The project aims to validate tests for the assessment of competencies university students need when dealing with scientific primary literature. Tests were constructed in the preceding KOSWO project (Students' Competencies when Dealing with Scientific Primary Literature; see von der Mühlen, Richter, Schmid, Schmidt & Berthold, in press; von der Mühlen, Richter, Schmid, Berthold & Schmidt, in press). A further aim of the project is to adapt the test system for application in additional academic disciplines.

Theoretical Framework

In almost all academic disciplines, university students are required to deal with primary academic literature. To do so successfully requires more than receptive reading comprehension. At the latest when students must write a bachelor or master thesis, they must sift through the relevant literature, identify appropriate texts, and adopt a critical stance to the text contents. Thus, in order to competently deal with scientific literature, students need to be equipped with a wide range of competencies, many of which are not formally taught at universities.

The project draws on contemporary conceptions of scientific literacy that have shifted in focus from scientific content knowledge to ways of dealing with scientific documents (see Britt, Richter, & Rouet, 2014). These conceptions consider receptive-systematic processes, epistemic processes (e.g., the evaluation of arguments; Larson, Britt, & Kurby, 2003), and heuristic processes (e.g, the use of source information [sourcing], Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012).

To systemize the set of potentially relevent subcompetencies necessary for dealing with scientific literature, we distinguish between two types of goals (receptive vs. epistemic goals) and two types of processing modes (systematic vs. heuristic strategies). Readers with a receptive goal strive to aquire knowledge, whereas readers with an epistemic goal strive to develop an own standpoint on the issues discussed in the text (Richter & Schmid, 2010). These goals can be pursued in an either systematic (deep-level processing of text contents) or heuristic (rule-based processing of surface features) fashion (see Petty & Wegener, 1999). By combining the dimensions epistemic-receptive and systematic-heuristic it is possible to construct a matrix with four categories of strategies:

(a) *Receptive-systematic strategies* serve to enrich or structure information in order to facilitate later recall (e.g., rehearsal, organization, and elaboration). These categories are at the core of research on

reading and learning strategies (e.g. Pintrich, 2004; Wild, 2000). (b) *Receptive-heuristic strategies* serve to gain a first impression of the text content (e.g., *skimming*, cursory reading to extract the gist, and *scanning*, cursory reading to find specific information). These strategies are particularly relevant in the phase of literature research and profit from generic knowledge of canonic text structures (Dillon, 1991). (c) *Epistemic-systematic strategies* serve to validate the argumentation of a text. In some taxonomies of learning strategies, they are covered by the construct *critical evaluation* (e.g., Pintrich, 2004). At a more specific level, it is possible to distinguish between *knowledge-based validation* and *consistency checking* (Richter & Schmid, 2010). Consistency checking requires the reader to identify the functional components of arguments such as *claim* and *ground*, to evaluate the acceptability of the ground, and to evaluate the relevance of the ground for the claim (Larson, Britt, & Kurby, 2009; Toulmin, 1958). (e) Finally, *epistemic-heuristic strategies* serve to gain a quick preliminary evaluation of the credibility of the text (e.g., by using source information). Such strategies are particularly important when readers lack relevant domain-specific content knowledge or the cognitive or motivational resources necessary for deep processing (Richter, Schroeder & Wöhrmann, 2009; Schroeder, Richter & Hoever, 200).

We find it likely that readers need to possess knowledge of strategies in all four categories for competently dealing with scientific literature. However, they also need to know when and for what purpose they can use a particular strategy (conditional knowledge, Lorch, Lorch, & Klusewitz, 1993). Skimming and scanning, for example, are particularly useful in the phase of literature research to find out whether a particular publication is relevant for the question at hand. Only if this condition is met, the reader should turn to systematic strategies.

Study Design

The project is based on a cognitive psychology conception of competencies as "systems of knowledge and belief" that enable successful actions in particular performance situations (Klieme & Hartig, 2007). In line with this, the focus of the project is on the analysis of declarative and procedural knowledge that allows students to competently deal with scientific literature. By this, we seek to validate tests constructed in the precursor project. Our validation approach includes training experiments, an expert-novice study using eye tracking, the assessment of ecologically valid achievement data, and a study on interdisciplinary differences.

For the purpose of experimental validation, we conduct computer-based training experiments. Students in different conditions participate in different training modules. Each module focuses on the declarative and procedural knowledge we assume to be necessary for employing strategies in one of the four categories (e.g., information on argument components, argumentation errors, heuristic indicators of scientific quality, and canonic text structures). We expect selective training effects on tests for different strategies. For the purpose of process-oriented validation, we conduct an expert-novice study with scientists as experts and students as novices. In this study, correlates of successful task performance are assessed using eye tracking. Eye tracking data are disambiguated by additional offline indicators of strategy use (e.g., questionnaires, short interviews, and reaction times). For the purpose of predictive validation, students who participated in the longitudinal study of KOSWO are contacted once more to obtain self-reports on the grades they received for their bachelor theses. We expect that considering heuristic and systematic strategies in connection to epistemic strategies and receptive competencies will enhance the assessment's ability to predict academic achievement. Finally, we plan

to adapt the test system for application in additional academic disciplines (education, teacher service, history, and biology), to explore the generalizability of the competency model.

Project data

Project management	Dr. Sebastian Schmid (University of Regensburg) Prof. Dr. Tobias Richter (University of Kassel) Prof. Dr. Klaus-Peter Wild (University of Regensburg)
Contact	Universität Regensburg Fakultät für Psychologie und Sportwissenschaft Universitätsstr. 1, 93053 Regensburg
Homepage	http://www.uni-regensburg.de/psychologie-paedagogik-sport/paedagogik-1/projekte/astralite/index.html
Time frame	01 Jan 2016 – 31 Dec 2018

References

Britt, M. A., Richter, T. & Rouet, J.-F. (2014). Scientific Literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist, 49*, 104-122.

Dillon, A. (1991). Reader's models of text structures: the case of academic articles. *International Journal of Man-Machine Studies, 35*, 913-925.

Goldman, S., Braasch, J. L., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly*, 47, 356-381.

Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. *Zeitschrift für Erziehungswissenschaft, 10*, 11-34.

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education, 77*, 339-366.

Lorch, R. F. Jr., Lorch, E. P. & Klusewitz, M. A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology, 85*, 239–252.

Petty, R. E. & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 41-72). New York: Guilford Press.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*, 385-407.

Richter, T. & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning, 5*, 47-65.

Richter, T., Schroeder, S. & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96, 538-558.

Schroeder, S., Richter, T. & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language, 59*, 237-259.

Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

von der Mühlen, S., Richter, T., Schmid, S., Berthold, K. & Schmidt, L.M. (in press). Use of source-related strategies in evaluating multiple psychology texts: A student-scientist comparison. *Reading and Writing*.

von der Mühlen, S., Richter, T., Schmid, S., Schmidt, L. M. & Berthold, K. (in press). Judging the plausibility of arguments in scientific texts: A student-scientist comparison. *Thinking & Reasoning.*

Wild, K.-P. (2000). *Lernstrategien im Studium. Strukturen und Bedingungen.* Münster: Waxmann.

KomPrü – Performance-based Assessment of Students' Communication Skills

Braun, E., Athanassiou, G., Gockel, S. & Pollerhof, K.

Abstract

The research group focuses on the construction of a test setting for the performance-based assessment of communication skills, more specifically of students of economics and teacher-training. Using Habermas' (1984) theory of communicative action and the distinction between strategic and communicative action, the conceptual elements corresponding with the respective communication types, i.e. strategic and understanding-oriented communication, were extracted. These elements were incorporated in simulated communication situations (role play), which correspond with either a strategic or with an understanding-oriented communication type; and in the construction of a standardized instrument for the observation-based assessment of communication skills. The complex situations are tightly coupled to the respective demands of the professional field, as obtained from a German nationwide survey of graduates. A total of 32 role plays were constructed: 16 for each group, 8 with a strategic and 8 with an understanding-oriented layout. The role play scenarios are being currently piloted in two courses at the University of Kassel. Video-recordings thereof will serve as training material for confederates (i.e. standardized interlocutors) and raters assigned with the task of the observation-based assessment. Role plays are to be tested and standardized to the respective populations (500 students) in autumn 2016 in 11 randomly selected Higher Education institutions throughout Germany.

Aims and Research Questions

Measuring competencies in higher education has largely been neglected in international empirical education research, despite the increasing importance of tertiary education to society" (KoKoHs, 2016). Additionally, performance-based competence assessments are gaining increasing importance for skill and competence measurement. This is the case for generic, cross-domain competences, such as communication skills that are explicitly referred to as an important outcome of higher education in national and international qualification frameworks. The current research project attends to the need for performance-based assessment of communication skills in Higher Education and aims at the construction of a theoretically founded psychometrically sound test setting that that can be efficiently employed in the HE context for this purpose.

Theoretical Framework

Table 1 illustrates the conceptual elements of a communicative action that were extracted through an extended review of the respective body of literature. The elements were identified and selected on the basis of their relevance for a differentiated application-orientated approach to strategic and understanding-oriented communication (Braun, Pollerhof & Athanassiou, under review).

Study Design

1. Study: Development of a compositional theoretical framework

The conceptual elements of communicative action were extracted from existing theoretical models of communication in correspondence with the project objectives.

Context: Situation and reason for the interlocution		
Social role (Argyle et al., 1981; Goffman, 2008): higher/equal/lower (power) position (compared to counterpart in the interlocution)		
Intention on the	Understanding-oriented (Habermas, 1984)	Strategic (Habermas, 1984)
Factual level (Watzlawick, 1969; Watzlawick et al., 2011; Hinsch & Pfingsten, 2007)	Compliance with the cooperative principle: maxims of quantity, quality, relevance and manner (Grice, 1975)	Purpose-supporting use of the maxims of quantity, quality, relevance and manner (Grice, 1975)
Relational level Watzlawick, 1969; Watzlawick et al., 2011; Hinsch & Pfingsten, 2007)	Authentic self-disclosure (Hargie, 2011)	Purpose-supporting self-disclosure (Hargie, 2011)

Table 1: Conceptual framework for the simulated situations: Extracted theoretical components communicative action (marked with light color: provided as instruction in the role plays; marked with dark color: related to performance in the role plays and operationalized in the observation instrument)

2. Study: Development of role plays and construction of observation instrument

A series of role play scenarios was constructed. The construction was fitted to the extracted elements from the literature review. In addition, situations displayed in the scenarios were based on qualitative data derived from a nationwide, representative survey (10.000 graduates) of typical interaction situations of cross-domain Higher Education graduates. A total of 32 role plays were constructed: 16 for each undergraduate group of economics/teacher-education, (paired scenarios with a shared thematic core whereas the content was adjusted to the specific characteristics and demands of each professional field), of which 8 display a strategic and 8 an understanding-oriented layout. The corresponding observation instruments are tailored to the role play scenarios and identical for the paired scenarios. Table 2 illustrates the assignment of the constructed role play scenarios to the distinct communication types and the predetermined hierarchical position reflecting the power constellation (i.e. the social role of the test takers) in the simulated situation.

		Weaker power position	**Equal power position**	**Stronger power position**
Communication type	**Strategic**	**4**	**1**	**3**
	Understanding-oriented	**2**	**4**	**2**

Table 2: Assignment of the constructed role play scenarios to the distinct communication types and distribution of the test takers' power position among the simulated situations

3. Study: pilot phase of role plays (currently on going)

Role plays were validated through an expert workshop with the participation of practitioners from both relevant professional fields as well of experts in competence research. Subsequently, role play scenarios are currently piloted with undergraduate students of both study courses at the University of Kassel. The pilot process has so far demonstrated that the applicability of the role plays in the context

of Higher Education is very good. Participating students report high levels of motivation with the simulated situations and the activating character of the role play method. Students also report enhanced communication skill performance as a result of participating in the course. So far, there is no dropout at all.

4. Study: Training confederates and raters (summer 2016)

Role plays carried out in the teaching course are video recorded. Video recordings of the sessions will be used for training the confederates (simulated interlocutors) and the designated users of the observation instrument (raters). In addition, inter-rater agreement will be tested using the video material.

5. Study: Testing in the field (autumn 2016)

Role plays are to be tested and standardized to the respective populations (500 students) in autumn 2016 in 11 randomly selected Higher Education institutions (Universities, Universities of applied sciences, private Universities) in Germany. Item Response Theory will be employed for the analysis of the collected data.

Expected results

The research group is aiming at providing Higher Education with a performance-based competence test setting for the assessment of communication skills of undergraduate students, as an end product of the research. The respective training material for confederate and rater training will also be presented to the scientific community. In addition to an employment towards skill assessment, existing role play situations can also be used for the purpose of training communication skills.

Project data

Project Management	PD Dr. Edith Braun
Project team members	Dr. Georgios Athanassiou (Dipl. Psych.) Kathleen Pollerhof (M.A. Sociology) Svea Gockel (B.A. Political Science/Scandinavian Studies)
Contact	INCHER (International Centre for Higher Education Research) Mönchebergstr. 17 D - 34109 Kassel
Homepage	http://www.uni-kassel.de/einrichtungen/incher/forschung/studierende-und-absolventinnen/kompetenzorientierte-pruefung-kommunikativer-faehigkeiten-von-studierenden.html#c19315
Time frame	June 2014 – May 2017

References

Argyle, M., Furnham, A. & Graham, J. A. (1981). *Social situations*, Cambridge: Cambridge University Press.

Braun, E., Pollerhof, K. & Athanassiou, G. (under review). *Konzeption einer kompetenzorientierten Prüfung kommunikativer Fähigkeiten*.

Goffman, E. (2008). *Behavior in public places*, New York: Simon and Schuster.

Grice, H. P. (1975). *Logic and conversation*. In P. Coyle & J. Morgan (Eds.), Syntax and Semantics. 3rd ed. New York: Academic Press. 41–58.

Habermas, J. (1981). *Theorie des kommunikativen Handelns*. (Vol. 1: Handlungsrationalität und gesellschaftliche Rationalisierung, Vol. 2: Zur Kritik der funktionalistischen Vernunft). Frankfurt am Main: Suhrkamp Verlag.

Hinsch, R. & Pfingsten, U. (2007). *Gruppentraining sozialer Kompetenzen. GSK: Grundlagen, Durchführung, Anwendungsbeispiele.* 5th ed. Weinheim: Beltz PVU.

KoKoHs (2016). Forschungsprogramm der KoKoHs Förderlinie 2015-2019. http://www.kompetenzen-im-hochschulsektor.de/403_ENG_HTML.php, accessed on 26 Jan 2016.

Watzlawick, P. (1969). *Menschliche Kommunikation*. Bern: Huber.

Watzlawick, P., Bavelas, J.H. & Jackson, D.D. (2011). *Pragmatics of human communication. A study of interactional patterns pathologies and paradoxes*, New York: Norton.

MultiTex – Process-based Assessment of Multiple Documents Comprehension

Schoor, C., Artelt, C., Kröhne, U., Goldhammer, F., Lockl, K. & Roßbach, H.-G.

Abstract

In order to cope with the demands of their studies and their later profession, university students of all subjects have to be able to familiarize in a self-directed manner with different topics and filter out and integrate task relevant information. Especially in their academic life at universities, they are usually confronted with multiple sources (e.g., as a result of searching a database or the internet), which can be contradicting and which they have to integrate. These demands exceed the comprehension of (single) texts in such as that information from different sources have to be related to each other and, if applicable, the trustworthiness of the information has to be evaluated.
In the project MultiTex, a computer-based instrument to assess the competence of text comprehension with multiple documents (multiple documents literacy, MDL) of university students will be developed and validated. A special methodological focus lies in the usage of additional process-based diagnostics. Accordingly, in this project process-based measures will be used, first, as additional diagnostic information; second, in order to identify strategies of processing multiple documents; and third, to test theories of multiple text comprehension.

Aims and Research Questions

The aims of the present project can be characterized as follows:

1. Development of a theoretically sound computer-based test of multiple documents literacy (MDL)
2. Analysis of processes of multiple text comprehension for advanced diagnostics
3. Construct validation of the MDL text with two cohorts of university students of two subject groups at two locations
4. Empirical test of assumptions of theories of multiple documents comprehension

Theoretical Framework

The central task of multiple text comprehension is the integration of information from different documents that can be redundant, complementary or consistent (e.g., Bråten, Anmarkrud, Brandmo, & Strømsø, 2014) into a coherent mental model. While according to Kintsch (e.g., 1998) during reading of a text a surface representation, a propositional representation and a situational model are built, in the case of multiple documents the documents model is built in addition (e.g., Britt & Rouet, 2012). The documents model contains both a mental model of the topic that integrates information from all the documents and the intertext model. The intertext model contains a node per document in which meta-information such as author, genre, publication date, its intention, its intended audience etc. is represented. These document nodes are related to each other by predicates, like „are in accordance“, „contradict each other“ or „contains evidence for the claim“. By means of the intertext model, also contradicting information can be integrated into a coherent mental model.
Since the analysis of multiple text comprehension is a comparably young field of research (see Maier & Richter, 2014), there are so far hardly any approaches for a standardized assessment of multiple

documents literacy based on psychometric methods, especially for the population of university students. In addition, it is unclear how reading literacy (usually operationalized as comprehension of single texts) as a trait is related to multiple documents literacy (comprehension of multiple documents).
In prior research, multiple text comprehension has primarily been assessed by means of essays that had to be written by the participants on the topic of the multiple documents (e.g., Rouet, Favart, Britt, & Perfetti, 1997). This method is – at least in its current version where the essays have to be rated by human raters – very time-consuming. Moreover, not only the intended receptive competences of MDL are assessed this way but also productive skills (writing skills), which may result in a bias in the assessment of MDL. Promising approaches that will be pursued in the current project are, first, verification tests in order to differentiate mental representations (e.g., Schmalhofer & Glavanov, 1986; or Strømsø, Bråten, & Britt, 2010, who suggested this in their inference verification task), and, second, the highlighting of information, where information has to be highlighted that fulfills certain requirements (like information that is contradicting across documents). Hereby it will be possible to assess subcomponents of MDL like conflict detection (see Stadtler & Bromme, 2014).
In previous studies the concrete processing of multiple documents has been examined only rarely. Computer-based assessment, however, provides the opportunity to assess traces of comprehension processes in a way that is both relatively easy and not influencing the assessment. These data can be used for advanced diagnostics. Computer-based log files can inform – on top of mere time data (see Goldhammer et al., 2014) – about the concrete behavior of the tested person. In the current project, we will concentrate on computer-based highlights and notes that the test persons make during reading (see Hagen, Braasch, & Bråten, 2014). These can indicate which information was judged important and was elaborated more deeply.

Study Design

For the planned MDL test, three to five units with two to three texts per unit will be used. Each unit will contain 5 – 15 items. The MDL test aims at covering the theoretical profile of requirements/demands needed for multiple documents comprehension. Furthermore, the test generates process data. For the computerization of the test, the CBA ItemBuilder software (Rölke, 2012) will be used.
The MDL test, which will be submitted to a pilot study for the purpose of item selection, will afterwards undergo a validation study at the universities of Bamberg and Frankfurt. In this study, 400 university students of educational science and psychology on the one hand, and history and German language studies on the other hand are supposed to participate. In a cross-sectional design, both first-year bachelor students as well as last-year master students will be tested in order to obtain first results on the change sensitivity of the test. In addition, the relationship of conventional reading competence (assessed with a computer-based test by the National Educational Panel Study) and MDL will be investigated. Further variables, like motivation, study success, and epistemological beliefs, will also be measured in order to obtain indicators for convergent and discriminant validity. Moreover, it might be possible to analyze subsamples with regard to suboptimal MDL or suboptimal MDL strategies.

Project data

Project management	Prof. Dr. Cordula Artelt, Universität Bamberg Dr. Cornelia Schoor, Universität Bamberg Dr. Kathrin Lockl, Leibniz-Institut für Bildungsverläufe e.V. Bamberg Prof. Dr. Hans-Günther Roßbach, Leibniz-Institut für Bildungsverläufe e.V. Bamberg Dr. Ulf Kröhne, DIPF Frankfurt Prof. Dr. Frank Goldhammer, DIPF Frankfurt
Contact	cornelia.schoor@uni-bamberg.de
Homepage	http://multitex.bildungsforschung.uni-bamberg.de/
Time frame	01 Jan 2016 – 30 June 2019

References

Bråten, I., Anmarkrud, Ø., Brandmo, C., & Strømsø, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction, 30*, 9-24. doi: 10.1016/j.learninstruc.2013.11.002

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276-314). New York: Cambridge University Press.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*(3), 608-626. doi: 10.1037/a0034716

Hagen, Å. M., Braasch, J. L. G., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading, 37*(S1), S141-S157. doi: 10.1111/j.1467-9817.2012.01536.x

Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.

Maier, J., & Richter, T. (2014). Verstehen multipler Texte zu kontroversen wissenschaftlichen Themen: Die Rolle der epistemischen Validierung [Comprehension of multiple texts about controversial scientific topics: The role of epistemic validation]. *Unterrichtswissenschaft, 42*(1), 24-38.

Rölke, H. (2012). The ItemBuilder: A graphical authoring system for complex item development. In T. Bastiaens & G. Marks (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012* (Vol. 2012, pp. 344-353). Chesapeake, VA: AACE.

Rouet, J.-F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15*(1), 85-106. doi: 10.2307/3233756

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25*(3), 279-294. doi: 10.1016/0749-596X(86)90002-1

Stadtler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp & J. L. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379-402). Cambridge, MA: MIT Press.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20*(3), 192-204. doi: 10.1016/j.learninstruc.2009.02.001

PRO-SRL-EVA – Product- and Process-Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation

Steuer, G., Eckerlein, N., Först, N., Lang, J., Spiel, C., Schober, B., Schmitz, B., & Dresel, M.

Research Questions

Self-regulation competencies are an explicit goal of higher education (Standing Conference of Minister of Education and Cultural Affairs of the Länder 2005). At the same time, they are already of great importance during tertiary education, as it is expected of students to acquire an extensive body of knowledge, to monitor their own learning processes and to overcome motivational problems while studying. Thus, self-regulation competencies are both a requirement and a result of successful university education, and moreover, the foundation for lifelong learning (see Artelt, Baumert, Julius-McElvany, & Peschar, 2003).

Resting on extensive work on theoretical concepts as well as explaining, capturing and promoting competencies of self-regulated learning, a model of self-regulation competencies in higher education and several innovative measuring instruments (Situational Judgement Test, Standardized Diaries, Competence-Performance-Assessment) have been developed in the joint research project PRO-SRL.

The joint-research project of the University of Augsburg, the University of Vienna and the Technical University Darmstadt now concentrates on the further validation of the competence model and the developed measuring instruments. The focus lies on validity evidences from different sources, mainly convergent, discriminant and criterial evidence (see Joint Committee on the Standards for Educational and Psychological Testing, 2014). Therefore, a series of correlative and experimental studies will be conducted. In this context, the development of trainings is planned, which can be implemented in the university setting on a low threshold level and can be adapted to the needs of students and the institutional requirements of higher education.

Objectives

The main objective of the research project is the further validation and extended examination of the competence model (see “Theoretical Framework”) and the derived measuring instruments (Situational Judgement Test, Standardized Diaries, Competence-Performance-Assessment). To achieve this objective, a series of quantitative empirical studies is intended, which will be established in three fields of studies to ensure generalizability (see “Design and Method”).

Theoretical Framework

The developed competence model connects the well-established theoretical perspectives of research on self-regulation competencies. It is created by combining three dimensions with three differentiations (see Dresel et al., 2015; Schober et al., 2015; Steuer al., 2015). The “strategic dimension” reflects the substantial distinction between cognitive strategies, metacognitive strategies and resource management strategies, which can be found in component-oriented models of self-regulation competencies. The “process dimension” reflects the temporal distinction between self-regulation processes of process-oriented models and differentiates between pre-action phase, action phase and post-action phase. Furthermore, the “knowledge dimension” distinguishes between different types of strategic knowledge (declarative, procedural and conditional knowledge) which are represented cognitively as

self-regulation competencies. The competence model was designed as a framework and working model and claims to integrate - comprehensively as well as in detail – all relevant competencies that are necessary for a successful self-regulation of learning (see Neumann, 2013; Schütte, Wirth, & Leutner, 2010).[2]

There were convergent results from two conducted studies to verify the competence model which showed that the questioned experts rated strategies in all possible combinations of strategy and process dimensions as relevant and differentiated between them in dependence of the learning situation. Therefore, the findings are in accordance with the proposed model. The competence model, its theoretical conceptualization and indications to its validity have already been published (Dresel et al., 2015; Schober et al., 2015; Steuer al., 2015). Nevertheless, an in-depth evaluation of the highly differentiated model structure and a further analysis with factor analytical methods are still pending.

Design and Method

The joint-research project intends a series of eleven quantitative-empirical studies, which are established in three fields of studies (Mathematics, Psychology and Economics) to ensure generalizability.

In the first phase, the product- and process-oriented measuring instruments will be systematically integrated, their internal structure will be analyzed and they will be correlated with other variables like academic performance, global self-reports and general cognitive ability (Studies 1a (N=70), 1b (N=70) and 1c (N=70)).

In the second phase, eye-tracking will be used as a method to acquire objective behavioral data of self-regulation competencies in order to obtain criterial evidence for actual observable self-regulation processes (Study 2 (N=70)).

In the third phase, the results from the previous phases will be integrated and expanded by experimental manipulation. To achieve this, self-regulation competencies of students will be systematically trained and promoted in six (quasi-)experimental studies – two for the Situational Judgement Test (Studies 3a (N=120) and 3b (N=120)), two for the Standardized Diaries (Studies 4a (N=120) and 4b (N=120)) and two for the Competence-Performance-Assessment (Studies 5a (N=120) and 5b (N=120)).

The fourth and last phase concentrates on representative comparisons between students with differences in self-regulation competencies. The aim of the last phase is to verify, if the assumed differences in self-regulation competencies (e.g. study year and study program) can be captured with the developed measuring instruments (Study 6 (N=900)).

After each study the assessments will be further developed and the acquired data will be specifically managed. At the end of the project time frame, an interdisciplinary symposium will be organized that promotes the transfer of the project results to the practice of higher education. The results will open several scientific and practical possibilities of usage, so that a sustainably transferable and implementable contribution to the research and training of a key competence in higher education is ensured.

Project Data

Project management	Prof. Dr. Markus Dresel (University of Augsburg)

[2] The combination of all aspects results in 27 cells, which represent potential part competencies. However, cognitive strategies are reserved for the actional phase. Thus, the combination results in 21 theoretically meaningful cells, which represent potential part competencies of self-regulated learning.

Prof. Dr. Christiane Spiel (University of Vienna)
Prof. Dr. Barbara Schober (University of Vienna)
Prof. Dr. Bernhard Schmitz (Technical University Darmstadt)

Contact Dr. Gabriele Steuer
gabriele.steuer@phil.uni-augsburg.de

Homepage https://www.philso.uni-augsburg.de/lehrstuehle/ psychologie/psycho1/forschung_int/projekte/pro_srl_eva

Time frame 01 Oct 2015 – 30 Sept 2018

References

Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning. Results from PISA 2000*. Paris: OECD.

Dresel, M., Schmitz, B., Schober, B., Spiel, S., Ziegler, A., Engelschalk, T. et. al. (2015). Competencies for successful self-regulated learning in higher education: Structural model and indications drawn from expert interviews. *Studies in Higher Education, 40*. 454–470.

Joint Committee on the Standards for Educational and Psychological Testing of the American Education Research Association, the American Psychological Association and the National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* Washington, DC: American Education Research Association.

Neumann, P. D. K. (2013). Mit welchem Auflösungsgrad können Kompetenzen modelliert werden? In welcher Beziehung stehen Modelle zueinander, die Kompetenz in einer Domäne mit unterschiedlichem Auflösungsgrad beschreiben? [Up to which level of detail can a competence be described? Which relation do competence models describing a competence at different levels of detail have to each other?] *Zeitschrift für Erziehungswissenschaft*, *16* (1), 35–39.

Schober, B., Klug, J., Jöstl, G., Spiel, C., Dresel, M., Steuer, G., Schmitz, B., & Ziegler, A. (2015). Gaining substantial new insights into university students' SRL-competences: How can we succeed? *Journal auf Psychology 223*, 64–65.

Schütte, M., Wirth, J., & Leutner, D. (2010). Selbstregulationskompetenz beim Lernen aus Sachtexten - Entwicklung und Evaluation eines Kompetenzstrukturmodells [Self-regulated learning from factual texts – Development and evaluation of a competence structure model]. *Zeitschrift für Pädagogik Beiheft, 56*, 249–257.

The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (2005). *Qualifikationsrahmen für Deutsche Hochschulabschlüsse* [Qualifications Framework for German Higher Education Qualifications].
Retrieved January 11, 2001 from
http://www.dgps.de/studium/rahmen/BS_050421_Qualifikationsrahmen_AS_Ka.pdf

Steuer, G., Engelschalk, T., Jöstl, G., Roth, A., Wimmer, B., Schmitz, B., Schober, B., Spiel, C., Ziegler, A., & Dresel, M. (2015). Kompetenzen zum selbstregulierten Lernen im Studium: Ergebnisse der Befragung von Expert(inn)en aus vier Studienbereichen [Competencies of self-regulated learning in higher education: Results from an interview study with experts from four areas of studies]. *Zeitschrift für Pädagogik, Beiheft 61*, 203-225.

ValiDiS – Competence Modeling and Assessment: Validation of the Test Score Interpretations of a Scientific Reasoning Test

Krüger, D., Upmeier zu Belzen, A. & Hartmann, S.

Abstract

The goal of project ValiDiS is to gather validity evidence for the test score interpretations of a scientific reasoning test for preservice science teachers, which has been developed during KoKoHs I in project Ko-WADiS. In ValiDiS, we will continue the longitudinal panel study which has been started in Ko-WADiS, in order to evaluate the instrument's sensitivity to known-group differences and the impact of covariates on competence development. Furthermore, we will investigate the test's sensitivity to interventions in which scientific reasoning competencies are trained. We will also investigate the predictive validity of our test score interpretations, evaluating the test scores' prognostic qualities on the participants' performance in real-life scientific problem solving situations. In addition to validity, we examine the generalizability towards other empirical science disciplines, testing to what degree the competencies which are assessed with the Ko-WADiS test can be used to solve scientific problems in disciplines such as psychology.

Aims and Research Questions

Our long-term goal is to provide a test instrument which can be used to evaluate academic programs in science teacher education. Only valid test score interpretations enable policymakers to base their decisions on the test results. In ValiDiS, we gather validity evidence from different sources, aiming at the following research questions: (1) To what extent can the performance measures of preservice teachers in the Ko-WADiS test be interpreted as valid indicators for scientific reasoning competencies? (2) To what degree can the competencies which are assessed with the test be generalized to other empirical science disciplines?

Theoretical Framework

In ValiDiS we will continue the longitudinal assessment of preservice science teachers' scientific reasoning competencies, which has been initiated in project Ko-WADiS (Hartmann, Upmeier zu Belzen, Krüger, & Pant, 2015). To evaluate the validity of the test score interpretations (see Rupp & Pant, 2006), we will examine several sources of validity evidence: test content, response processes, internal structure, relationships to other variables, and consequences of testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014). In ValiDiS, we will especially focus on the test's sensitivity to known-group differences in competence development (see AERA, APA, & NCME, 2014, p. 16), its sensitivity to interventions (see AERA, APA, & NCME, p. 17), and its relationships to other variables such as preservice teachers' professional knowledge and general cognitive ability. In addition, the transfer of the competence structure of scientific reasoning in biology, chemistry, and physics to other fields of empirical science will be investigated.

Study Design

ValiDiS is designed as a longitudinal panel study, assessing competencies of preservice teachers in academic training at Freie Universität Berlin (FU) and Humboldt-Universität zu Berlin (HU). We will continue assessing cohorts of preservice teachers from project Ko-WADiS. As control groups, cohorts in Austria will be tested in cooperation with our partners at the universities of Innsbruck and Vienna (Prof. Dr. S. Kapelari) and Salzburg (Prof. Dr. U. Unterbruner). In addition, we will conduct interventional studies at FU and HU as well as in cooperation with other universities such as Ludwig-Maximilians-Universität München (project "Reason", Prof. Dr. F. Fischer; see Fischer et al., 2014).
In order to assess the test's sensitivity to curricular measures, we will conduct pre-post observations in selected seminars and classes at universities in Berlin (FU and HU) and Cologne (Prof. Dr. K. Schlüter). To assess prognostic validity, test instruments from two projects, ExMo (Hasse, Joachim, Bögeholz, & Hammann, 2014) and KeiLa (Großschedl, Harms, Kleickmann, & Glowinski, 2015), will be used together with the Ko-WADiS test and a cognitive ability test (Liepmann, Beauducel, Brocke, & Nettelnstroth, 2012). Study design and data analysis are conceptually supported by Prof. Dr. H. A. Pant (HU). To initially examine the generalizability of the test score interpretations, psychology students at HU will be tested in cooperation with Prof. Dr. M. Ziegler.

Project data

Project management	Prof. Dr. Dirk Krüger; Prof. Dr. Annette Upmeier zu Belzen
Contact	Dr. Stefan Hartmann, stefan.hartmann@hu-berlin.de
Homepage	www.bcp.fu-berlin.de/biologie/arbeitsgruppen/didaktik/ www2.hu-berlin.de/biologie/biodidaktik
Time frame	01 Jan 2016 – 31 Dec 2019

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME] (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Hartmann, S., Upmeier zu Belzen, A., Krüger, D., & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of pre-service science teachers' competencies. *Zeitschrift für Psychologie, 223*, 47-59.

Hasse, S., Joachim, C., Bögeholz, S., & Hammann, M. (2014). Assessing teaching and assessment competences of biology teacher trainees: Lessons from item development. *International Journal of Education in Mathematics, Science and Technology, 2*, 191-205.

Großschedl, J., Harms, U., Kleickmann, T., & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: Structure and learning opportunities. *Journal of Science Teacher Education, 26*, 291-318.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, … Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 5*, 28-45.

Liepmann, D., Beauducel, A., Brocke, B., & Nettelnstroth, W. (2012). *IST-Screening - Intelligenz-Struktur-Test - Screening (IST-SC). PSYNDEX Tests Info.* Göttingen: Hogrefe.

Rupp, A. A., & Pant, H. A. (2006). Validity theory. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 1032-1035). Thousand Oaks, CA: SAGE Publications.

Cluster 2: Competencies in Teacher Education

BilWiss-UV

Returns and Development of Educational Knowledge—Validation of a Competence Test for Teacher Students

ELMaWi

Assessing Subject-Specific Competencies in Teacher Education in Mathematics and Economics – A Quasi-Experimental Validation Study with a Focus on Domain-Specificity

PlanvoLL-D

The Role of Professional Knowledge of Pre-Service German Teachers in their Lesson Planning: Validation and Methodological Innovation

ProfiLe-P+

Professional Competence in Academic Physics Teacher Education

Pro-KomMa

Professionalization of Early Childhood Teacher Education: Convergent, Discriminant, and Prognostic Validation of the KomMa Models and Tests

TEDS-Validierung

Validation of the Instruments from the International Comparative TEDS-M Study and its Follow-Up TEDS-FU

BilWiss-UV – Returns and Development of Educational Knowledge — Validation of a Competence Test for Teacher Students

Kunter, M., Leutner, D., Seidel, T., Kunina-Habenicht, O., Dicke, T. & Linninger, C.

Abstract

The joint project of the Goethe University Frankfurt, the University of Duisburg-Essen and the Technische Universität München aims to optimize and validate a test that has been developed during the previous research projects "BilWiss" and "Bilwiss-Beruf" for the assessment of teacher students' educational knowledge. The optimized and validated instrument should allow for a formative assessment during university studies on the one hand, and for the investigation of the impact of educational knowledge on teachers' professional practices on the other hand. We plan to investigate several strands of validity evidence. First, we investigate how the test content matches the curricula of courses on educational foundations at university. Second, we analyze response processes of testees while working on the test. Third, we analyze the internal structure of the test and the relationship between test scores and other variables, such as the reflection on experiences at school and professional behavior. Forth, we investigate how learning opportunities at university can contribute to the development of educational knowledge. Therefore, we plan to conduct a multi-cohort longitudinal study with teacher students to assess changes in educational knowledge throughout university studies.

Aims and Research Questions

The aim is to provide an instrument for the assessment of teacher students' and teacher candidates' educational knowledge in two applicable test versions which both have specific advantages. The long version of the test should provide a high curricular validity in order to address questions on the population level, such as evaluation studies and formative assessment. The short version of the test should allow for an efficient assessment of educational knowledge in order to address questions on the individual level, such as studies that assess changes in knowledge or focus on other aspects of professional competence and need to control for educational knowledge.

Theoretical Framework

A sound base of evidence on the impact of educational university studies on teachers' professional practice is lacking (Brouwer, 2010; Zeichner, 2005). This is mostly due to the fact that, unfortunately, studies which directly assess teacher students' educational knowledge and the long-term relation to the prospective teachers' professional practices are rare. Our understanding of educational knowledge is based on the model of Baumert and Kunter (2006). They draw on Weinert (2001) and conceptualize teachers' professional competence as an individual prerequisite necessary for an effective professional practice (Kunter et al., 2011). Several aspects of professional competence, such as professional knowledge, professional beliefs, motivational orientations or self-regulative skills, are supposed to enable teachers to act appropriately across different work-related situations (Blömeke, Gustafsson, & Shavelson, 2015; Kunter et al., 2011). Thereby, the model implies a distinction between competence (in terms of learnable and malleable personal prerequisites) and acting (in terms of performance). In our research project we aim to assess those aspects of professional competence that are particularly addressed within university studies on educational foundations.

References

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*. doi: 10.1007/s11618-006-0165-2

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3-13. doi: 10.1027/2151-2604/a000194

Brouwer, C. N. (2010). Determining Long Term Effects of Teacher Education. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education*. Oxford: Elsevier. 503-510.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2011). *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.

Weinert, F. E. (2001). A concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies*. Seattle: Hogrefe & Huber. 45–65.

Zeichner, K. M. (2005). A research agenda for teacher education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The AERA Panel on Research and Teacher Education*. Mahwah, N. J.: Lawrence Erlbaum. 737–760.

Study Design

Our study design includes several approaches (see Figure 1):

1) During the phase of optimizing the test we will analyze the statistical parameters of all items and gather expert ratings of how the test content matches curricula in teacher education. Furthermore, we will conduct cognitive labs with teacher students in order to get information on construct-irrelevant variance and response processes for each item. The revised items will be tested in pilot studies and a field test with teacher candidates at the beginning of their induction phase. The data will be collected at teacher training institutes in the federal state of North Rhine-Westphalia.
2) The revised test will be administered along with other instruments measuring aspects of professional development within a multi-cohort longitudinal sample of teacher students at the universities in Frankfurt, Essen, and Munich (Longitudinal study II, see Table 1).
3) The longitudinal sample of beginning teachers of the previous research projects *BilWiss* and *Bilwiss-Beruf* (Longitudinal study I) will be continued to gain further evidence on the long-term effects of educational knowledge.
4) Furthermore, we will conduct two experiments to assess the predictive validity and instructional sensitivity of the test scores. The data will be collected within specific educational courses on educational foundations by the project partners in Frankfurt and Munich.

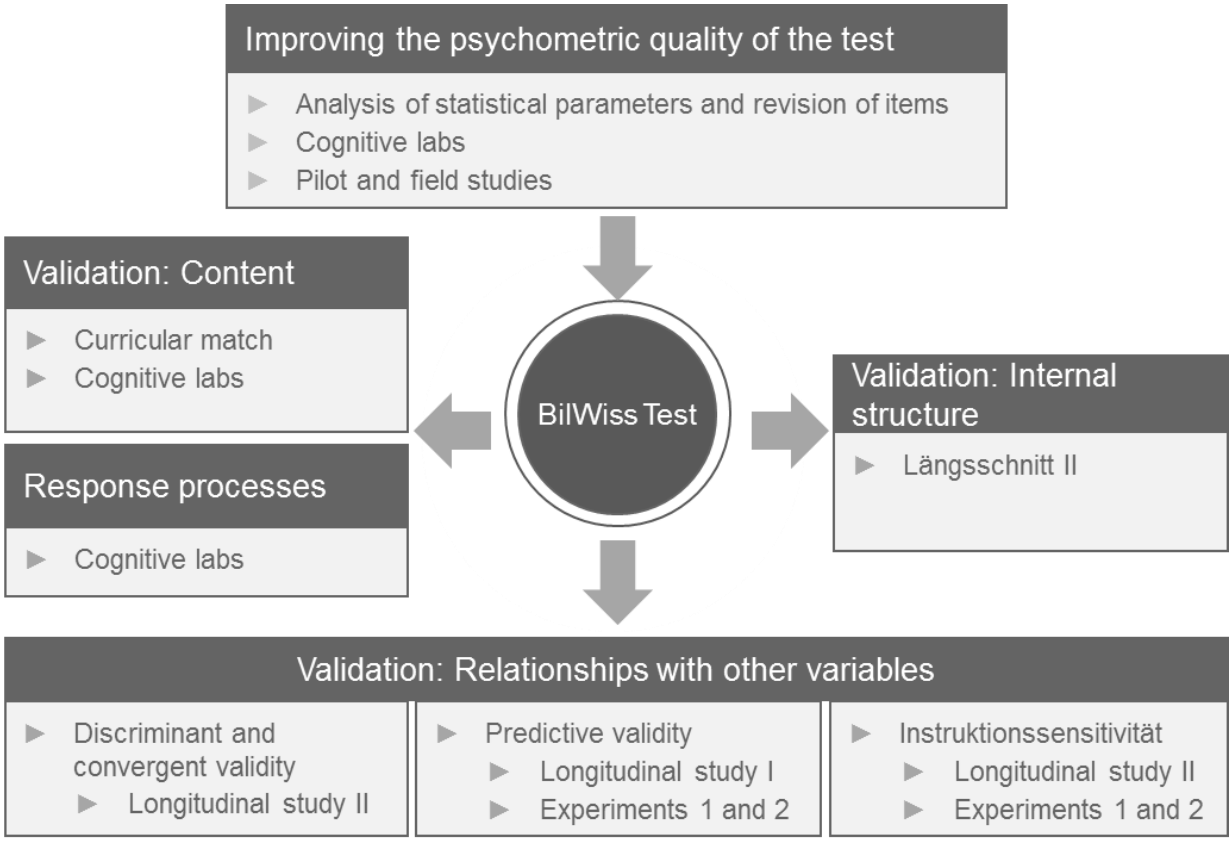


Figure 1. Research program of the BilWiss-UV project.

Table 1

Design of Longitudinal study II

Participants		Measurement points (MP) of the longitudinal study			
Cohort	**Number**	**MP 1: WT 16/17**	**MP 2: ST 17**	**MP 3: WT 17/18**	**MP 4: ST 18**
Cohort 1	150	Semester 1	Semester 2	Semester 3	Semester 4
Cohort 2	150	Semester 3	Semester 4	Semester 5	Semester 6
Cohort 3	240	Semester 5	Semester 6	Semester 7	Semester 8
Cohort 4	240	Semester 7	Semester 8	Semester 9	Semester 10
Sum:	780				

Note. WT = winter term; ST = summer term.

Project Data

Project management	Prof. Dr. Mareike Kunter (Goethe University Frankfurt), Prof. Dr. Detlev Leutner (University Duisburg-Essen), Prof. Dr. Tina Seidel (Technische Universität München)
Contact	Christina Linninger Goethe University Frankfurt Theodor-W.-Adorno-Platz 6 60323 Frankfurt am Main Phone +49 (0)69 798 35367 Email: linninger@psych.uni-frankfurt.de
Homepage	www.bilwiss.uni-frankfurt.de
Time frame	Feb 2016 – Jan 2019

References

Kunter, M., Kunina-Habenicht, O., Baumert, J., Dicke, T., Holzberger, D., Lohse-Bossenz, H., Leutner, D., Schulze-Stocker, F. & Terhart, E. (in press). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung – Ergebnisse des Projekts BilWiss. In C. Gräsel & K. Trempler (Eds.), *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven.* Wiesbaden: Springer-Online.

Dicke, T., Parker, P.D., Holzberger, D., Kunter, M., & Leutner, D.(2015). Investigating longitudinal changes in beginning teachers' efficacy and emotional exhaustion: Can they be predicted by professional knowledge? *Contemporary Educational Psychology, 41,*62-72.

Kunina-Habenicht, O., Lohse-Bossenz, H., Kunter, M., Dicke, T., Förster, D., Gößling, J., Schulze-Stocker, F., Schmeck, A., Baumert, J., Leutner D. & Terhart, E. (2012). Welche bildungswissenschaftlichen Inhalte sind wichtig in der Lehrerbildung? Ergebnisse einer Delphi-Studie. *Zeitschrift für Erziehungswissenschaft, 15*(4), 649–682. doi: 10.1007/s11618-012-0324-6

Kunina-Habenicht, O., Schulze-Stocker, F., Kunter, M., Baumert, J., Leutner, D., Förster, D., Lohse-Bossenz, H. & Terhart, E. (2013). Die Bedeutung der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens. *Zeitschrift für Pädagogik*, 59(1), 1-23.

Linninger, C., Kunina-Habenicht, O., Emmenlauer, S., Dicke, T., Schulze-Stocker, F., Leutner, D., Seidel, T., Terhart, E. & Kunter, M. (2015). Assessing teachers' educational knowledge: Construct specification and validation using mixed methods (Messung des Bildungswissenschaftlichen Wissens von Lehrkräften: Konstruktspezifikation und Validierungsansätze). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*. *47(2),* 62-74, doi: 10.1026/0049-8637/a000126.

Lohse-Bossenz, H., Kunina-Habenicht, O., Dicke, T., Leutner, D., & Kunter, M. (2015). Teachers' knowledge about psychology: Development and validation of a test measuring theoretical foundations for teaching and its relation to instructional behavior. *Studies in Educational Evaluation, 44*, 36–49.

Schulze-Stocker, F., Holzberger, D., Kunina-Habenicht, O. & Terhart, E. (2014). Bildungswissenschaften in der Lehrerbildung: Ergebnisse des Studiums und Entwicklungen im Referendariat. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttke (Eds.), *Kompetenzen von Lehramtsstudierenden und angehenden ErzieherInnen*. Bad Heilbrunn: Klinkhardt.

Schulze-Stocker, F., Holzberger, D., Kunina-Habenicht, O., Terhart, E. & Kunter, M. (accepted). Spielen Studienschwerpunkte wirklich eine Rolle? Zum Zusammenhang von bildungswissenschaftlichen Studienschwerpunkten, selbst eingeschätzten Kenntnissen und gemessenem Wissen am Ende eines Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft.*

ELMaWi – Assessing Subject-Specific Competencies in Teacher Education in Mathematics and Economics – A Quasi-Experimental Validation Study with a Focus on Domain-Specificity

Heinze, A., Kuhn, C., Lindmeier, A. & Zlatkin-Troitschanskaia, O.

Abstract

In a quasi-experimental validation study with pre-service and in-service teachers of mathematics and economics, the ELMaWi project team pursues two key objectives: i) to examine by means of innovative assessment designs the validity of the assessment of subject-specific teacher competency, that is, their knowledge as well as their ability to respond appropriately in challenging situations in the classroom, and ii) to examine the validity of the modeling of domain-specific and generic components of teacher competency.
Quasi-experimental variation is achieved through the sampling (N=600), varying with regard to training in the two domains (mathematics, economics) and expertise in teaching (university students, teacher trainees, teachers). For the three status groups, the sampling comprises 150 participants from each subject and 300 who have had academic training in both subjects. Interdependencies between the domains are examined based on the subsample of university students, teacher trainees, and teachers who are educated in both mathematics and economics.
The project contributes to a domain-specific understanding of the structure of subject-specific teacher competency in mathematics and economics by providing the validated competency models and the corresponding tests. It also enriches general understanding of subject-specific teacher competency by examining it in two domains and over three phases of teacher education.

Aims

In a quasi-experimental validation study with pre-service and in-service teachers of mathematics and economics, the ELMaWi project team pursues two key objectives: i) to examine by means of innovative assessment designs the validity of the assessment of subject-specific teacher competency, that is, their knowledge as well as their ability to respond appropriately in challenging situations in the classroom, and ii) to examine the validity of the modeling of domain-specific and generic components of teacher competency.

Theoretical Framework

This project is based on a competency structure model by Lindmeier (2011) and Kuhn (2014) which focuses on the situational requirements of teaching and differentiates between two components of domain-specific teacher competency: reflective competency (RC) and action-related competency (AC). RC is needed to prepare and evaluate specific subject-related situations during pre- and post-instructional phases, and AC is needed to handle specific subject-related situations during instruction and under time pressure. While RC requires conscious (reflective) cognitive processes that draw on domain-specific propositional knowledge (content knowledge (CK) and pedagogical content knowledge (PCK)), AC also requires knowledge but usually involves intuitively controlled or automated (less reflective) cognitive processes. In addition to central domain-specific constructs (RC, AC,

CK, PCK), both domain models include basic cognitive competency (facets of intelligence), generic competency (complex problem-solving and situational awareness) as well as the degree of expertise in teaching (university students, teacher trainees, and teachers) as non-domain-specific variables. By investigating teacher competency in the two different but related subjects of mathematics and economics, we are able to examine the validity of the assessment across domains.

Study Design

In the study's quasi-experimental design, constructs are examined for convergent, discriminant, incremental, and predictive validity by means of a multitrait-multimethod matrix and two contrast groups. Quasi-experimental variation is achieved through the sampling (N=600), varying with regard to training in the two domains (mathematics and economics) and expertise in teaching (university students, teacher trainees, and teachers). For the three status groups, the sampling comprises 150 participants from each subject and 300 who have had academic training in both subjects. Interdependencies between the two domains are examined based on the subsample of university students, teacher trainees, and teachers who are educated in both mathematics and economics. Domain-specific competencies, non-domain-specific competencies, and the influence of expertise are analyzed. This framework covers the in-depth validation of the competency structure model, which differentiates between domain-specific and non-domain-specific competencies. It also covers the associated instruments for competency assessment of pre-service and in-service teachers in both domains; this includes testing with computer-based test formats to facilitate the valid assessment of action-based teacher competency. Particular focus is on explaining the test results in terms of their domain-specificity (mathematics or economics) as well as their non-domain-specific proportions. By means of different expertise groups, we examine whether a change in competency can be mapped over the various stages of teacher education.

Project Data

Project management	IPN – Leibniz Institute for Science and Mathematics Education Department of Mathematics Education Prof. Dr. Anke Lindmeier, Prof. Dr. Aiso Heinze Johannes Gutenberg University Mainz Department of Business and Economics Education Dr. Christiane Kuhn, Prof. Dr. Olga Zlatkin-Troitschanskaia
Contact	info@elmawi.de
Homepage	www.elmawi.de
Time frame	Mar 2016 – Feb 2019

References

Knievel, I., Lindmeier, A.M. & Heinze, A. (2015, online first). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education*, 1–21. doi:10.1007/s10763-014-9608-z

Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung [Pedagogical Content Knowledge of (Future) Teachers in Business and Economics: Theoretical Modeling, Test Development, and Validation].* Landau: Verlag Empirische Pädagogik.

Kuhn, C., Happ, R., Zlatkin-Troitschanskaia, O., Beck, K., Förster, M. & Preuße, D. (2014). Kompetenzentwicklung angehender Lehrkräfte im kaufmännisch-verwaltenden Bereich – Erfassung und Zusammenhänge von Fachwissen und fachdidaktischem Wissen [Competence Development among Future Teachers in the Field of Business Administration – Assessment and Relationship of Content Knowledge and Pedagogical Content Knowledge]. In E. Winther & M. Prenzel (Eds.), *Perspektiven der empirischen Berufsbildungsforschung: Kompetenz und Professionalisierung*. *Zeitschrift für Erziehungswissenschaft*, *17*(1), Special Issue 22. Wiesbaden: Springer. 149–167. doi:10.1007/s11618-013-0456-3

Lindmeier, A. (2011*). Modeling and measuring knowledge and competencies of teachers: A threefold domain-specific structure model for mathematics.* Münster: Waxmann.

Lindmeier, A.M., Heinze, A. & Reiss, K. (2013). Eine Machbarkeitsstudie zur Operationalisierung aktionsbezogener Kompetenz von Mathematiklehrkräften mit videobasierten Maßen. *Journal für Mathematik-Didaktik*, 34(1), 99–119. doi:10.1007/s13138-012-0046-6

PlanvoLL-D – The Role of Professional Knowledge of Pre-Service German Teachers in their Lesson Planning: Validation and Methodological Innovation

Bremerich-Vos, A., Buchholtz, C. & König, J.

Abstract

PlanvoLL-D aims at investigating the relation of professional knowledge pre-service German teachers acquire in their university education (first phase of teacher preparation) and a core requirement of their practical training (second phase): lesson planning.
PlanvoLL-D relies on two previous studies. In the "Teacher Education and Development Study: Learning to Teach" (TEDS-LT; funded by the BMBF), reliable tests were designed which measure content knowledge (literary and linguistic knowledge), pedagogical content knowledge (literary and linguistic pedagogical knowledge), and general pedagogical knowledge (Bremerich-Vos & Dämmer, 2013; König et al., 2013). It remains an open question whether the competences the future teachers gain in their university education determine their professional development in the second phase of training and therefore possess predictive validity for their further competence development. Lesson planning of teachers as a research field has received little attention so far in terms of modelling and measuring relevant competences. In a second previous study we used an innovative approach to measure planning competence: In "Planungskompetenz von Lehrerinnen und Lehrern" (PlanvoLL; König et al., 2015), we developed a standardized method for analyzing written lesson plans which highlights the demand of pedagogical adaptivity. Since this measures the pedagogical adaptivity of teacher trainees as a crucial aspect of their planning competence irrespective of their subject, in PlanvoLL-D, we refine the approach by applying it to German as a specific subject.

Aims

The project PlanvoLL-D aims at investigating the relation of professional knowledge pre-service German teachers acquire in their university education (first phase of teacher preparation) and a core requirement of their practical training (second phase): lesson planning. Professional knowledge comprises content knowledge (CK), pedagogical content knowledge (PCK), and general pedagogical knowledge (GPK), i.e., three major cognitive components of professional competence of teachers (Baumert & Kunter, 2006).
We hypothesize that professional knowledge of pre-service German teachers has predictive validity for their general and subject-specific lesson planning skills (planning competence) in the second phase of teacher preparation. While knowledge is defined as “cognitive disposition” and captured via standardized paper-pencil tests, lesson planning will be considered as “situation-specific skill” (see Blömeke, Gustafsson, & Shavelson, 2015) and measured according to an innovative approach of analyzing written plans of demonstration lessons (“Lehrproben”), i.e., typical work samples (Audit-Report on Funding Initiative, p. 14).
To our knowledge, this research question has not been investigated by empirical competence research yet. Following reform debates on teacher education in Germany – on the coherent structure of first and second phase, on the orientation towards teaching practice, on the support of cumulative learning of student teachers – PlanvoLL-D is considered to provide significant findings on the development of teacher preparation at university. Subsequently, our utilization concept aims at concrete measures to

transfer project findings to current challenges on designing and evaluating extended forms of teaching practice at university.

Theoretical Framework and Research Questions

Lesson planning is a major challenge for teachers. Planning a lesson can be understood as a problem solving process, for which specific knowledge in the area of the subject, subject-specific pedagogy, and general pedagogy should be relevant (König, Buchholtz, & Dohmen, 2015). Paper-pencil tests that capture knowledge in a de-contextualized way are of limited value when measuring and modelling planning competence, though: As research on expertise of teachers has shown, typical problems occurring during the process of lesson planning are inextricably linked to the specific situation, in particular to the characteristics of the learning group as perceived by the planning teacher (summarized in König et al., 2015).

There is very little empirical evidence on how pre-service teachers plan their lessons. Research on measuring and modelling lesson planning competence does not exist at all (e.g., Aufschnaiter, & Blömeke, 2010). A first approach has been provided by the project *„Planungskompetenz von Lehrerinnen und Lehrern“* (PlanvoLL; König, Buchholtz & Dohmen, 2015; Buchholtz & König, 2015), with analyzing written plans of demonstration lessons (“Lehrproben”) in a standardized way. Applying this methodological approach to the planning process implies that the pre-service teachers are required to relate their knowledge to a concrete, real learning group. It thus differs from standardized knowledge tests that require pre-service teachers to respond to test items capturing aspects of lesson planning in a de-contextualized way. However, with the innovative approach developed in the PlanvoLL project, pedagogical adaptivity was captured irrespective of the subject only. Since the purpose of lesson planning is always related to a specific subject, the domain-specific modelling and measuring of planning competence is still a research gap.

PlanvoLL-D just takes up this point. We apply the innovative approach of analyzing written plans of demonstration lessons as developed in PlanvoLL to German as one of the main school subjects (secondary level). Besides, we use paper-pencil tests developed the „Teacher Education and Development Study: Learning to Teach“ (TEDS-LT, funded by the BMBF) that measure content knowledge (literary and linguistic knowledge), pedagogical content knowledge (literary and linguistic pedagogical knowledge), and general pedagogical knowledge (Bremerich-Vos & Dämmer, 2013; König et al., 2013). The aim of our study is to examine the role of pre-service teachers’ professional knowledge in their lesson planning. Two overarching research questions are focused on:

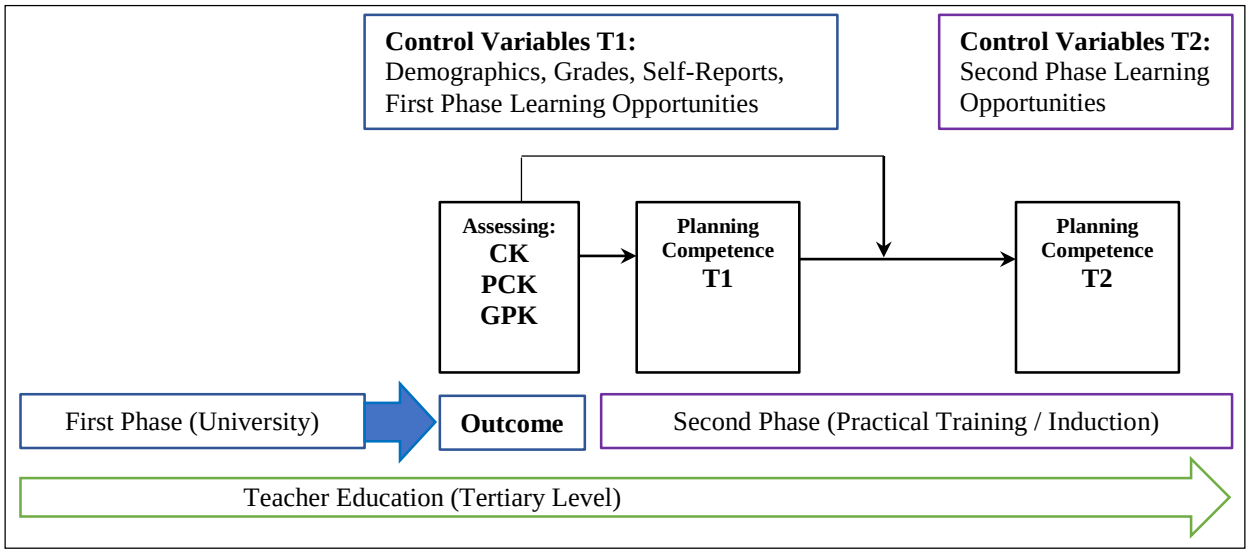
1) Is it possible to concretize the generic planning competence (pedagogical adaptivity) with regards to German-specific pedagogy (subject-specific pedagogical adaptivity)?

2) Does pre-service teachers’ knowledge as acquired at university predict planning competence as well as the development of planning competence? And does, as a consequence, teachers’ knowledge have predictive validity for their planning competence?

Study Design

Target group of our study PlanvoLL-D are pre-service middle school teachers of German in the federal states of Berlin and North Rhine-Westphalia. In our research model (see Figure 1), professional knowledge of pre-service teachers is defined as an outcome of the first phase of teacher preparation at university. That knowledge is tested at the beginning of the second phase, using the tests developed in TEDS-LT. In contrast, pre-service teachers’ planning competence will be measured at two time points: We will collect and analyze the written plans of the lessons they demonstrate at the very start

and the very end of their practical training. This enables us (a) to relate their knowledge on their planning competence at the start of practical training and (b) to relate their knowledge on the change of planning competence (gain) during practical training.



Abbreviations: CK – German Content Knowledge, PCK – Pedagogical Content Knowledge, GPK – General Pedagogical Knowledge; T1 – Time Point 1, T2 – Time Point 2)

Figure 1. Research Model.

Project data

Project management	Prof. Dr. Albert Bremerich-Vos, University Duisburg-Essen Dr. Christiane Buchholtz, Technical University of Berlin Prof. Dr. Johannes König, University of Cologne
Contact	Dr. Sandra Lammerding, University of Cologne sandra.lammerding@uni-koeln.de
Homepage	http://www.hf.uni-koeln.de/37700
Time frame	01 Apr 2016 – 31 Mar 2019

References

Bremerich-Vos, A. & Dämmer, J. (2013). Professionelles Wissen im Studienverlauf: Lehramt Deutsch. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Keßler & K. Schwippert (Eds.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerausbildung aus TEDS-LT*. Münster: Waxmann. 47-75.

Buchholtz, C. & König, J. (2015). Erfassung von Planungskompetenz im Praxissemester. *Journal für LehrerInnenbildung*, 15(1), 39-45.

König, J., Blömeke, S. & Schwippert, K. (2013). Pädagogisches Professionswissen im Studienverlauf. In S. Blömeke, A. Bremerich-Vos, G. Kaiser, G. Nold, H. Haudeck, J.-U. Keßler & K. Schwippert (Eds.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerausbildung aus TEDS-LT*. Münster: Waxman. 145-166.

König, J., Buchholtz, C. & Dohmen, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, *18* (2), 375-404. http://dx.doi.org/10.1007/s11618-015-0625-7

ProfiLe-P+ – Professional Competence in Academic Physics Teacher Education

Vogelsang, C., Borowski, A., Fischer, H., Kulgemeyer, C., Reinhold, P., Riese, J. & Schecker, H.

Abstract

The research project ProfiLe-P+ aims at an in-depth analysis of various aspects of validity concerning models and test instruments for physics teachers' professional knowledge developed in the prior project ProfiLe-P. We will examine how professional competencies achieved in academic physics teacher education (focusing on the aspects content knowledge (CK) and pedagogical content knowledge (PCK)) affect the performance of teacher trainees in standard situations of physics teaching. According to Miller (1990) we differentiate between (a) tests for competence (in this case: using knowledge in PCK and CK to master teaching-related problems in a written test), (b) tests for performance (mastering standardized teaching situations), and (c) tests for professional action (observations of classroom action). We will develop standardized performance tests concerning two standard situations of teaching physics (planning physics teaching and analyzing physics teaching) in addition to a performance test for explaining physics, which has already been developed in the prior project ProfiLe-P. One of our main research aims is to analyze how CK and PCK (measured by curricular valid test instruments) are used to master the standard situations of physics teaching and how these aspects of competence affect the quality of performance. In addition, we conduct a longitudinal survey to study on the development of (a) CK and PCK and (b) the performance in explaining physics during academic physics teacher education. The aim of this part of our project is to gain evidence for the development of competencies during academic teacher education and about the prognostic validity of our test instruments and models. We will conduct the longitudinal study with two different cohorts (basically Bachelor and Master students) including teacher trainees completing a long-term internship (one semester) at schools. We will especially focus on the development of the competencies and their relationship (e.g. the impact of CK on PCK, impact of PCK and CK on planning physics teaching) during this internship. We aim to include universities with very different concepts of teacher education and explore the impact of these concepts on competence development.

Aims and Research Questions

In order to evaluate academic teacher trainee programs, we need valid diagnostic instruments to assess competencies achieved in academic teacher education programs. During the first funding period (project ProfiLe-P, c.f. relevant literature section) we developed models and paper-and-pencil test instruments for CK and PCK. These tests measure reliably different aspects of CK and PCK. Using these instruments we examined the relationship between CK and PCK more in detail. Whether or not this relationship depends on the progress of teacher trainees in their academic teacher education is an open question – that includes the question whether or not the models for CK and PCK are stabile during the educational progress. Even more, the analysis of construct validity regarding CK and PCK is an open research question. Therefore, in ProfiLe-P+ we will use the models and instruments developed in Profile-P (c.f. Riese et al., 2015) to collect arguments for further aspects of validity, among them nomological networks and the stability of the interdependence of CK and PCK during the academic teacher education. We will focus on three research aims. (1) Firstly we will analyze construct validity by using a nomological network that includes CK and PCK as well as the performances in planning and analyzing physics teaching (c.f. Fig. 1). (2) Secondly we will analyze validity regarding prognostic validity and the

stability of the interdependence of CK and PCK. (3) Thirdly we will gain first evidence about the impact of different academic teacher trainee programs by comparing the development of competencies for extreme groups of these programs. These groups will be formed based on characteristics of the program (e.g. the existence of long-term internships). Currently each university in Germany has a different concept for their academic teacher trainee program and some of them differ a lot. Our research program might give some hints for successful elements of these programs.

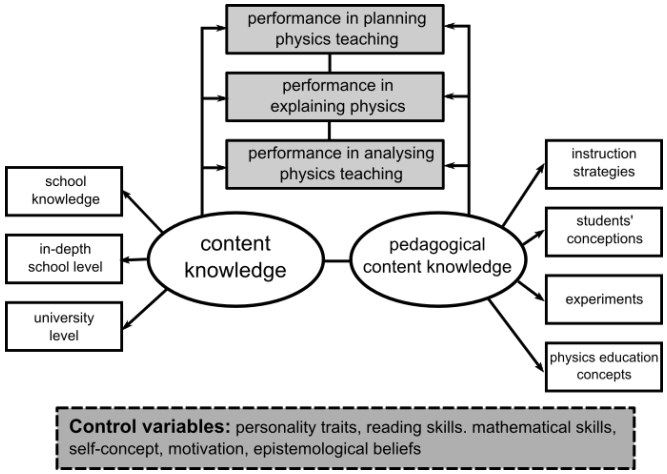


Figure 1. Nomological network: relationships between different parts of teachers' professional knowledge and performances in standard situations of physics teaching.

Theoretical Framework

Many research projects concerning teachers' professional competencies focus on professional knowledge based on Shulman's (1986) influential differentiation between content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical knowledge (PK). However, prior studies in physics reveal a very small correlation between the level of competence (or professional knowledge) and actual teaching quality – some evidence even implies that there is no correlation at all (see Cauet et al., 2015; Vogelsang, 2014). On the other hand the German academic teacher education relies on developing CK, PCK and PK. Based on current research it is an open question at which point the "functional chain" from these aspects of competence achieved during academic teacher education and teaching quality tears off. One reason for this could be that teaching quality is mostly measured by observing physics teaching – and actual physics lessons are a very complex setting influenced by a lot of variables beneath teacher competencies. That is why we develop performance tests instead of observing teachers' classroom action. For one particular standard situation of teaching physics – explaining physics – we could show an impact of CK and PCK on explaining performance (Kulgemeyer, Tomczyszyn & Schecker, 2015). Performance assessment for the standard situation of planning physics teaching has already been conducted (Stender, 2014). Nevertheless, a deeper analysis of the relationship between professional competencies and the performances in planning physics teaching or analyzing physics teaching has not been undertaken. This kind of analysis however is needed to use existing paper-and-pencil tests for an evaluation of the impact of academic teacher education programs. Even more, it is a crucial question for teacher education in general how (or even if) academically achieved competencies are of use for actual teaching.

Study Design

ProfiLe-P+ relies on four different sub-projects in order to reach its research aims. The *first sub-project* will develop a performance test for the standard situation of planning physics teaching. This development will be approached in two different ways. Firstly, we will develop a standardized test that challenges each test person with the same problem: planning a physics lesson with a given topic in a given context (e.g. students, prior teaching, etc.). Secondly we will compare the performance in this test with an analysis of documents the test persons prepare to plan lessons at different stages during their long term internship at a school. We will evaluate whether or not an analysis of these documents leads to the same results as the performance test. We will also research how we can use both the documents and the test to describe a possible progress during the internship. The *second sub-project* will develop a similar performance test for the standard situation of analyzing physics teaching. Also in this sub-project we will compare a standardized test (a standardized lesson to analyze) with the analysis of the lessons the students conduct during the internship. We will use the instruments developed in the first two sub-projects in order two examine the relationship between CK, PCK and the performance quality in the two standard situations of physics teaching we focus on. Additionally, we will test for control variables like cognitive abilities and include them into this analysis. The *third sub-project* will focus on the development of CK, PCK and performance quality in explaining physics using the instruments developed in the prior project ProfiLe-P. We will collect our data in a longitudinal study. Beneath these three tested areas we will also consider variables such as prior knowledge at the beginning of the longitudinal study. Especially we will keep record concerning the opportunities for learning during study. The fourth sub-project will on the one hand coordinate the collection and management of the data. On the other hand it will lead the exploration of the impact of elements of an academic teacher trainee program on competency development. We will focus on the impact of a long-term internship at a school but also consider the quantity and quality of courses for CK and PCK in the programs. All sub-projects will collect their data at least five universities based on a longitudinal design with two cohorts (first three years of academic teacher education, year four and five of academic teacher education). We will therefore include students from different stages of teacher education (Bachelor and Master). All in all the number of tested students will reach about 800 physics teacher trainees.

Project Data

Project management	Josef Riese, Christoph Kulgemeyer, Peter Reinhold, Andreas Borowski
Contact	Prof. Dr. Josef Riese, AG Didaktik der Physik und Technik, I. Physikalisches Institut, RWTH Aachen, Sommerfeldstr. 14, D-52074 Aachen, Tel. +49 241 80-27163, Email: riese@physik.rwth-aachen.de Prof. Dr. Andreas Borowski, Didaktik der Physik, Universität Potsdam, Karl-Liebknecht Str. 24/25, D-14476 Potsdam, Tel. +49 331 977 1025, Email: andreas.borowski@uni-potsdam.de Dr. Christoph Kulgemeyer, Institut für Didaktik der Naturwissenschaften (IDN), Universität Bremen, Bibliothekstraße, D-28359 Bremen, Tel. +49 421 218-62782,

Email: kulgemeyer@physik.uni-bremen.de
Dr. Christoph Vogelsang, AG Didaktik der Physik, Universität Paderborn, Warburger Str. 100, D-33098 Paderborn, Tel. +49 5251/60 2680, Email: cvogelsa@mail.uni-paderborn.de

Homepage	http://physik.uni-paderborn.de/reinhold/forschung/profilep+/
Time Frame	Apr 2016 – Sept 2019

References

Cauet, E.; Liepertz, S.; Borowski, A.; Fischer, H.E. (2015). Does it Matter What We Measure? - Domain-specific Professional knowledge of Physics Teachers. In *Schweizerische Zeitschrift für Bildungswissenschaften*, 37 (3), 463-480.

Kulgemeyer, C., Tomczyszyn, E. & Schecker, H. (2015). *Was beeinflusst die Performanz beim Erklären von Physik?*. Vortrag auf der Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik vom 14.-17. September 2015 in Berlin, 17 Sept 2015.

Miller, G. E. (1990): The assessment of clinical skills/competence/performance. In *Academic Medicine* 65 (9), 563–567.

Riese; J.; Kulgemeyer, C.; Zander, S.; Borowski, A.; Fischer, H.; Gramzow, Y.; Reinhold, P.; Schecker, H. & Tomczyszyn, E. (2015). *Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik*. In Blömeke, S. & Zlatkin-Troitschanskaia, O. (Eds.): *Kompetenzen von Studierenden: 61. Beiheft der Zeitschrift für Pädagogik*. Weinheim: Beltz. 55-79.

Shulman, L. (1986). Those who Understand: Knowledge Growth in Teaching. In *Educational Researcher*, 15, 4-15.

Stender, A. (2014). *Unterrichtsplanung: vom Wissen zum Handeln. Theoretische Entwicklung und empirische Überprüfung des Transformationsmodells der Unterrichtsplanung*. Berlin: Logos.

Vogelsang, C. (2014). *Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften. Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz* . Berlin: Logos.

Pro-KomMa – Professionalization of Early Childhood Teacher Education: Convergent, Discriminant, and Prognostic Validation of the KomMa Models and Tests

Jenßen, L., Jegodtka, A., Eilerts, K., Eid, M., Koinzer, T., Schmude, C., Rasche, J., Szczesny, M., & Blömeke, S.

Abstract

From a methodological point of view, Pro-KomMa will provide further steps in validating inferences to be drawn from the tests developed in the KomMa project. Aim of the tests is to represent preschool teachers' professional competencies in the field of mathematics at the end of their training. It was hypothesized that these competencies are acquired during preschool teacher education. Furthermore, it was hypothesized that they are relevant for everyday job situations in preschool, in particular for fostering children's development of mathematical literacy. To test convergent, discriminant, and prognostic validity, in a first study a broad range of preschool teachers' and children's characteristics will be assessed. In a second study, preservice preschool teachers' competence development during their training will be assessed in a longitudinal design. Results of these two studies will be related to results obtained with the mathematics content knowledge (MCK), mathematics pedagogical content knowledge (MPCK) and general pedagogical knowledge (GPK) tests of the forerunner project KomMa. From a substantive point of view, Pro-KomMa contributes to the examination of the effectiveness of early childhood teacher education. It will also be possible to examine prospective preschool teachers' competence development during teacher education. Additionally, test scores representing preschool teachers' professional competence will be related to test scores representing children's mathematical competence.

Aims and Research Questions

Main aim of the Pro-KomMa project is the validation of the tests developed in the forerunner project KomMa. These tests assess different constructs assumed to be acquired during teacher education and to be relevant for fostering children's mathematical literacy. The substudy of Pro-KomMa carried out at Humboldt-Universität zu Berlin and Freie Universität Berlin examines this hypothesized causal relationship. Consequently, Pro-KomMa partly examines the effectiveness of early childhood education studies with regard to contents. In the substudy carried out at Alice-Salomon-Hochschule (ASH), preschool teachers' performance will be assessed and related to competence profiles. Hence, knowledge about preschool teachers' practice and specifically about preschool teacher-child interaction will be gathered.

Based on modern validity theory (Kane, 2013), the following main and sub-aims will be followed:

1) Examination of the validity of the tests concerning preschool teachers' competence development during their studies
    a. Examination of the factorial structure of MCK, MPCK and GPK in a longitudinal design (beginning and end of academic studies)
    b. Examination of the factorial structure of cognitive dispositions and cognitive skills
    c. Examination of the sensitivity of the tests concerning competence development from the beginning of the studies to the end
    d. Examination of convergent and discriminant validity concerning predictors for competence development

   e. Examination of convergent, discriminant and incremental validity concerning individual characteristics which were deducted from a nomological net
2) Examination of the prognostic validity of the tests developed in the KomMa project
   a. Prognostic validation concerning preschool teachers' performance in daily-life situations in preschool
   b. Prognostic validity concerning the development of children's mathematical competencies

Theoretical Framework

The theoretical framework of Pro-KomMa is based on Weinert's (2001) definition of competence as a multidimensional and demand-related construct. Blömeke, Gustafsson and Shavelson (2015) have expanded this definition concerning situation-specific skills which mediate competence and performance. According to Shulman (1986), the cognitive facet of professional competence in the field of mathematics can be distinguished into MCK, MPCK and GPK. These knowledge facets represent the cognitive disposition of competence whereas math-related and pedagogical beliefs, emotions and motivations (Benz, 2012) represent the affective-motivational disposition.

Blömeke, Gustafsson and Shavelson (2015) pointed out that situation-specific skills are needed to identify math-related daily-life situations (Fröhlich-Gildhoff, Nentwig-Gesemann & Pietsch, 2011, 2014), to plan specific math-related activities (Bruns, 2014), and to observe math-related abilities of children (Dunekacke, Jenßen & Blömeke, 2015a, 2015b). One main characteristic of these situation-specific skills is their strong relation to practical contexts (Lindmeier, 2013). The mathematical activities of the children are observable and represent the outcome of competence (Bruns, 2014).

Teacher education is regarded crucial for competence development because competence can be seen as a learnable characteristic. Specifically, entry requirements for the studies and quantity and quality of opportunities to learn (OTL) may have an effect for this development. These assumptions are based on research on educational effectiveness and provide information on identification of individual and institutional predictors (Mujis et al., 2014; Scheerens, 2015).

For Pro-KomMa assumptions from competence research, educational effectiveness theories, research on early childhood education and developmental psychology were integrated to model causal relationships for academic studies of preschool teachers in the field of mathematics (Figure 1).

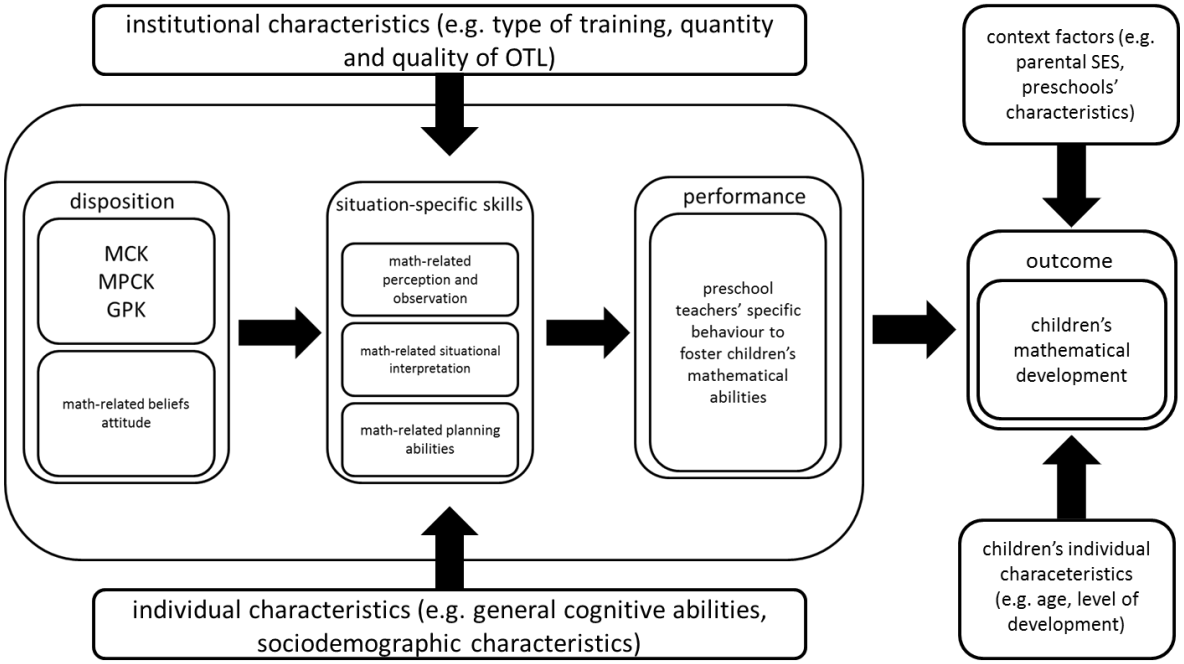


Figure 1: Model of causal relationships for preschool teachers' competence in the field of mathematics

Design of the quantitative Substudy

In the substudy at Humboldt-Universität zu Berlin and Freie Universität Berlin, relevant constructs (e.g. OTL, knowledge facets, situation-specific skills) were deducted from a nomological net to realize the aims of the Pro-KomMa project. Thus, the assessment battery contains the KomMa paper-pencil-based and video-based tests, OTL questionnaires, beliefs questionnaires, a diagnosing tool for children's mathematical performance (Walter-Laager et al., 2012) and a standardized observation procedure to assess children's mathematical performance.

To investigate the model developed for Pro-KomMa (Figure 1), two samples will be tested (Table 1). Sample 1 consists of pre- and in-service preschool teachers who were tested in the KomMa project and now once again in a longitudinal design (approx. *n*=500). Sample 2 consists of preschool teachers from Berlin and Brandenburg who will be tested to investigate the prognostic validity of the KomMa tests (approx. *n*>50). These preschool teachers will be observed in their daily-life activities in preschool and 10 children per preschool teacher will be chosen at random to test prognostic validity of the KomMa tests concerning their mathematical performance.

sample	demo-graphy	OTL	MCK, MPCK, GPK	video-based tests	beliefs	KiDiT®	obser-va-tions	children's math literacy
I	X	x	x	x	x			
II	X	x	x	x	x	x	x	x

Table 1. Overview of the samples and assessments

The KomMa tests will be analyzed by application of 2PL IRT-based models (Eid & Schmidt, 2014; Bond & Fox, 2013). All the other assessments will be analyzed by applying classical test theory. Competence structure will be examined by multidimensional factor models. To test our hypothetical assumptions regarding relationships of all variables, confirmatory factor analysis will be applied (Eid, Gollwitzer &

Schmitt, 2013). To identify developmental processes, the two-factorial model (person x item) of the KomMa project will be expanded by integrating a third factor “occasion” (Eid, 2000; Davier, Xu & Carstensen, 2011; Hartig & Kühnbach, 2005). Effects of preschool teachers’ competencies on children’s mathematical performance will be examined by multi-level models with cross-level interactions (Eid, Gollwitzer & Schmitt, 2013).

Design of the Qualitative Substudy

In the substudy at the ASH, preschool teachers’ performance in their math-related daily-life practice will be investigated by sequential videography (Bohnsack, 2009). The preschool teachers will be accompanied during two typical structured daily-life situations, the morning circle and lunch, as well as the most typical unstructured daily-life situation, the free play (Tietze et al., 1998). Because of considerations of feasibility (König, 2009) these three situations will be videotaped sequentially. The focus of the investigation is on interaction processes containing mathematics between the child and preschool teacher. The sample of *n*=12 is recruited from the total sample.
In addition, a video-based group discussion will be conducted: Theoretically derived daily-life situations relating to mathematical topics will be used to survey the competency of situational awareness and analysis, which is located in the transition area between disposition and performance of preschool teachers. The sample of *n*=approx. 25 will be recruited from the total sample, excluding the participants of the videography-study.
To identify performance types of preschool teachers as well as for the evaluation of models for competence of early childhood education staff (Fröhlich-Gildhoff et al., 2011), a typology will be generated according to the documentary method (Bohnsack, 2009). Moreover, the quantitative and qualitative data will be related to each other according to a mixed methods design (Kelle, 2008; Kuckartz, 2014).

Project Data

Project Management	Prof. Dr. Katja Eilerts and Prof. Dr. Thomas Koinzer (Humboldt-Universität zu Berlin) Prof. Dr. Michael Eid (Freie Universität Berlin) Prof. Dr. Corinna Schmude ("Alice-Salomon"-Hochschule für Sozialarbeit und Sozialpädagogik Berlin)
Contact	Prof. Dr. Katja Eilerts (director of the collaborative project) Kultur-, Sozial- und Bildungswissenschaftliche Fakultät Institut für Erziehungswissenschaften Grundschulpädagogik, Lernbereich Mathematik Unter den Linden 6, 10099 Berlin phone: 030/2093-4012 , Email: katja.eilerts@hu-berlin.de
Time frame	01 Jan 2016 – 31 Dec 2018

References

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015a). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223*(1), 3-14.

Blömeke, S., Dunekacke, S., Jenßen, L., Suhl, U., Grassmann, M., & Wedekind, H. (in press). Leistungstests zur Messung der professionellen Kompetenz frühpädagogischer Fachkräfte. In press in *Zeitschrift für Pädagogische Psychologie*.

Dunekacke, S., Jenßen, L., & Blömeke, S. (2015a). Effects of Mathematics Content Knowledge on Pre-school Teachers' Performance: a Video-Based Assessment of Perception and Planning Abilities in Informal Learning Situations. *International Journal of Science and Mathematics Education, 13*, 267-286.

Dunekacke, S., Jenßen, L., Eilerts, K., & Blömeke, S. (2015b). Epistemological beliefs of prospective pre-school teachers and their relation to knowledge, perception and planning abilities in the field of mathematics: A process-model. *ZDM The International Journal on Mathematics Education*. DOI 10.1007/s11858-015-0711-6.

Jenßen, L., Dunekacke, S., Gustafsson, J.-E., & Blömeke, S. (submitted). Intelligence and competence: Clarifying the hierarchy of preschool teachers' cognitive abilities. *Journal of Educational Psychology*.

Jenßen, L., Dunekacke, S., & Blömeke, S. (2015a). Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. *Zeitschrift für Pädagogik, 61. Beiheft*, 11-31.

Jenßen, L., Dunekacke, S., Eid, M., & Blömeke, S. (2015b). The Relationship of Mathematical Competence and Mathematics Anxiety -An Application of Latent State-Trait Theory. *Zeitschrift für Psychologie, 223*(1), 31-39.

Jenßen, L., Dunekacke, S., Baack, W., Tengler, M., Koinzer, T., Schmude, C., Wedekind, H., Grassmann, M., & Blömeke, S. (2015c). KomMa: Kompetenzmodellierung und Kompetenzmessung bei frühpädagogischen Fachkräften im Bereich Mathematik. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttke (Eds.), *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte*. Bad Heilbrunn: Klinkhardt. 59-79.

TEDS-Validation – Validation of the Instruments from the International Comparative TEDS-M Study and its Follow-Up TEDS-FU

Kaiser, G., König, J., Buchholtz, N., Busse, A. & Blömeke, S.

Abstract

The TEDS-Validation project aims at answering the question of whether research findings brought forward by measurement instruments to test professional competence of mathematics teachers have predictive validity for the quality of their instruction and the learning progress of their students. TEDS-Validation strongly relies on previous work on the quality-controlled development of instruments measuring professional competence of mathematics teachers that was conducted in the context of TEDS-M, TEDS-FU, and TEDS-Instruction. TEDS-Validation examines the last part related to predictive validity. It is a joint project of the University of Hamburg, the University of Cologne, and the CEMO (Centre for Educational Measurement of the University of Oslo). It will be conducted in cooperation with Thuringia Institute for Teacher Professional Development, Curriculum Development and Media (Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien; Thillm) and is supported by the project "*Kompetenztest.de*" of the Friedrich-Schiller-University Jena. TEDS-Validation will be conducted in Thuringia.

Aims

The project TEDS-Validation aims at answering the question of whether research findings brought forward by measurement instruments to test professional competence of mathematics teachers have predictive validity for the quality of their instruction and the learning progress of their students. The instruments, which will be used, have been developed by the comparative „Teacher Education and Development Study: Learning to Teach Mathematics“ (TEDS-M) (Blömeke, Kaiser & Lehmann, 2010a, b) and measure the professional competencies of future mathematics teachers. If predictive validity of teachers' competencies for quality-oriented instruction and students' achievements can be confirmed, then such assessments will fulfill being objective, reliable, and valid and enable examination of teacher effectiveness in future higher education research. The TEDS-M measurement instruments have already been validated. They capture the central cognitive elements of mathematics teachers' professional competence (Shulman, 1987; Baumert & Kunter, 2006): mathematical content knowledge (MCK), mathematical pedagogical content knowledge (MPCK), and general pedagogical knowledge (GPK). However, it remains an open question whether predictive validity can be assumed with regards to the mastering of professional tasks that are relevant during teaching as the core task of teachers.
Another objective of the project will be to answer the question which forms of teacher competencies – declarative and procedural– are acquired and where during teacher preparation and teachers' practice these forms are significant for teaching. The TEDS-M measurement instruments measure mainly – not exclusively - declarative parts of cognitive knowledge ("knowing that…"), which is acquired during teacher preparation and can be considered as part of a teacher's knowledge base needed for teaching. By contrast, procedural knowledge ("knowing how…") depends strongly on practical experience and is more related to situations and performance in class. That is why in the follow-up of TEDS-M, the so-called TEDS-FU study, innovative forms of situation-specific teacher competence measures were de-

veloped ("video-cued testing"). TEDS-FU differentiates these situation-specific abilities triply: perception of classroom situations, interpretation of these situations and decision-making concerning these situations (in the following called PID-model).
Based on expert reviews, these test instruments have been proven to be reliable and valid as well as being suitable for capturing situation-specific skills. If they also turn out to have predictive validity, then in-depth analyses on the acquisition and on the relation of declarative and procedural knowledge during teacher preparation as well as on their differential predictivity can be carried out in order to get evidence on designing teacher education at university, especially for the theory-practice-relationship of opportunities to learn. Such evidence will be related to mathematics teacher preparation, but still they contain high potential for generalization.

Theoretical Framework

The international state of research on the development of professional competence during teacher preparation has strong gaps. Although all studies assume a chain of effects such as teacher preparation – teacher competence – instructional quality – student progress, there is no empirical evidence whether teacher competencies that teacher had acquired during teacher preparation have an influence on the quality of their instruction and the learning progress of their students. Also the relationship of teacher categories content knowledge, pedagogical content knowledge, and general pedagogical knowledge on instructional quality and student learning has not yet been modelled simultaneously, neither have situation-specific cognitive skills been accounted for.
To make it concrete, in the project TEDS-Validation the following research questions will be investigated:

(1) Can we provide empirical evidence that TEDS-M and TEDS-FU measurement instruments have predictive validity for teaching mathematics at high quality? We expect that MCK, MPCK, and GPK as well as video-based measures significantly impact instructional quality and correlate with the learning gain of students.
(2) Do situation-specific abilities based on the PID model (as measured via video-based assessments) contribute to explain instructional quality and learning gain of students – in addition to the effects of professional knowledge of teachers (as measured via paper-pencil tests), i.e., do they have added value? We expect that video-based skills of perception, interpretation, and decision-making correlate higher with instructional quality and students' learning gain than knowledge-based tests (MCK, MPCK, GPK). Moreover, we expect that the relationship between video-based measures and learning gain will be mediated by instructional quality.

Study Design

TEDS-Validation will be conducted in Thuringia in 2016. About 150 teachers will participate. Evaluation of cognitive and situation-specific tests will take about three hours testing time and will be provided online. Moreover, instructional quality of a subgroup of teachers (n=25) will be captured through classroom observations over four lessons. A recently developed rating instrument will be used. Students' learning progress will be measured via central assessments (Lernstandserhebungen) that regularly take place during grade 6 and 8 in Thuringia. These assessments are aligned on the national educational standards for mathematics.
Data will be provided through the project "kompetenztest.de" by the University of Jena. After data collection and cleaning, descriptive statistics will be computed. Test data will be scored using Item-Response-Theory. Instructional quality scales will be computed according to the procedure applied in

other studies. We will examine our hypotheses based on covariance and factor analyses as well as by using structural equation modelling on the level of latent variables.
Access to the field will be supported through the cooperation with the Thuringia Institute for Teacher Professional Development, Curriculum Development and Media (Thüringer Institut für Lehrerfortbildung, Lehrplanentwicklung und Medien, Thillm). Data collection will be conducted online with the software package "UniPark". Incentives will be given to the teachers, e.g., by offering professional teacher development workshops. After publication of first project findings, instruments of the project and project data will be offered to the scientific community according to APA-standards.

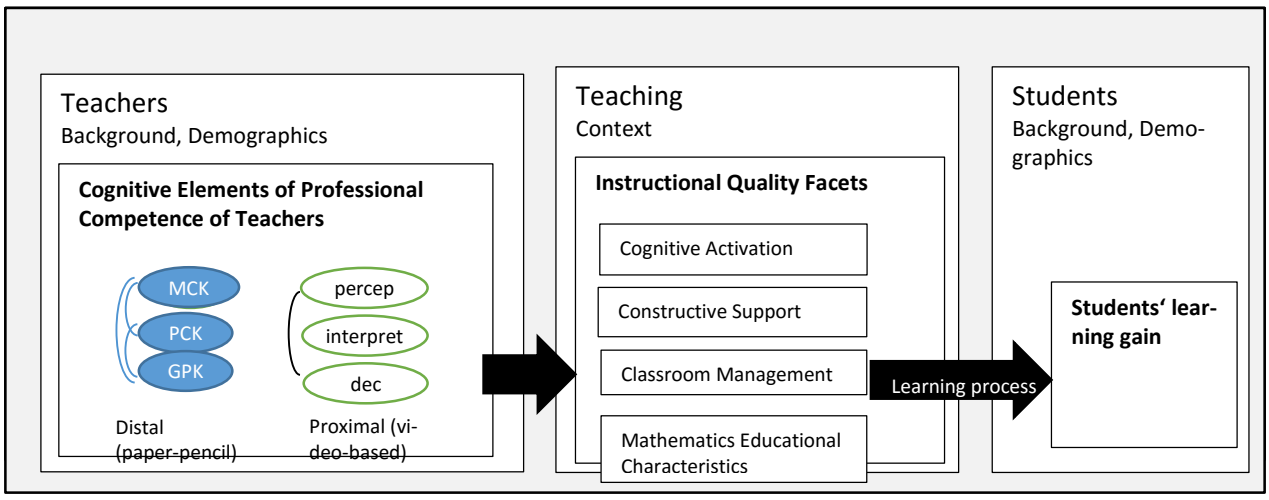


Figure 1. Research model
Key: mathematical content knowledge (MCK), mathematics pedagogical content knowledge (MPCK) and general pedagogical knowledge (GPK), Percept (Perception of classroom situations), Interpret (Interpretation), Dec (Decision-making)

Project data

Project management	Prof. Dr. Gabriele Kaiser (leader of research group) Universität Hamburg Prof. Dr. Johannes König, Universität zu Köln Prof. Dr. Sigrid Blömeke (Universität Oslo, Norwegen)
Contact	Dr. Nils Buchholtz, Universität Hamburg Nils.buchholtz@uni-hamburg.de Dr. Andreas Busse, Universität Hamburg Andreas.busse@uni-hamburg.de
Homepage	https://www.teds-validierung.uni-hamburg.de/
Time frame	01 Feb 2016 – 31 Mar 2019

References

Blömeke, S., Gustafsson, J.-E. & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie, 223,* 3–13.

Blömeke, S., Hoth, J., Döhrmann, M., Busse, A., Kaiser, G. & König, J. (2015). Teacher Change During Induction: Development of Beginning Primary Teachers' Knowledge, Beliefs and Performance. *International Journal of Science and Mathematics Education*, 13, 287-308.

Blömeke, S., König, J., Suhl, U., Hoth, J. & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit von videobasierten Performanztests. *Zeitschrift für Pädagogik, 61,* 310-327.

Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M. & Kaiser, G. (2014). Von der Lehrerausbildung in den Beruf - Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 17 (3), 509-542.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. Teaching and Teacher Education, 56(2016), 35-46.

Kaiser, G., Busse, A., Hoth, J., König, J. & Blömeke, S. (2015). About the complexities of video-based assessments: Theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *International Journal of Science and Mathematics Education, 13,* 369-387.

König, J., Blömeke, S. & Kaiser, G. (2015). Early career mathematics teachers' general pedagogical knowledge and skills: Do teacher education, teaching experience, and working conditions make a difference? *International Journal of Science and Mathematics Education, 13,* 331-350.

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education, 38,* 76-88.

Cluster 3: Domain-specific Competencies (Medicine / Business & Economics)

ÄKHOM

ÄKHOM – Medical Competences: Hamburg-Oldenburg-Munich

KoNa-Ma

Simulation-based Measurement and Validation of a Competence Model for Sustainability Management

WiWiKom II

Valid Assessment of Students' Development of Professional Business and Economic Competencies over the Course of their Studies - A Quasi-experimental Longitudinal Study

WiWiSET

Validation of an Entrance Examination in the Study Domain of Business and Economics – A National and International Comparative Study of Universities and Universities of Applied Sciences

ÄKHOM – Medical Competences Hamburg – Oldenburg – Munich

Harendza, S., Kadmon, M. & Berberat, P.

Abstract

The Third Section of the Medical Exam, the oral-practical state examination after the final practice year of undergraduate medical education, is a non-structured, non-validated exam. Its design is almost entirely allocated to the medical faculties. Hence, it is hardly possible to compare the students' medical competences between the different medical faculties by the results from this assessment. The aim of this study is the further development of a newly designed competence-based assessment format for medical students in their final year at different German universities. Students' development of competences in different curricula at three medical faculties (Hamburg, Oldenburg, TU München) will be compared by means of this assessment. This enterprise is based on completed validations of an international project comparing medical students' competences at the end of their final year of undergraduate training. Within the scope of the planned Federal-State initiative "Master Plan Undergraduate Medical Education 2020" it could provide an essential component for the introduction of a national, standardized, and structured clinical-practical exam with oral and practical parts based on the National Competence-Based Catalog of Learning Objectives in Medicine (NKLM) which has been passed by the Medical Faculty Association (MFT).

Aims and Research Questions

The aim of this project is the further validation of a new competence-based assessment format for undergraduate medical students in the final year of their curriculum, which has already been tested in an international context in a cooperation between Hamburg and Utrecht medical school. Students from three German medical faculties with different undergraduate medical curricula (Hamburg, Oldenburg, TU München) will be tested with respect to the development of their competences using this assessment format. If appropriate, it could replace the current Third Section of the Medical Exam, a non-structured, non-validated oral-practical exam to provide better comparison of exam results on a federal level.

Theoretical Framework

Medical competence is the ability to act responsibly and adequately within medical contexts during the daily routine as well as in unexpected situations for the individual and the collective advantage based on justified benefit and the reflection of communication, complex knowledge, technical skills, emotions, values, and attitudes (compare the concepts of Epstein and Hundert 2002**,** Van der Blij et al. 2002 and Erpenbeck and Sauter 2013). A Delphi survey amongst clinical professors resulted in 25 facets of competence, which were considered to be particularly relevant for medical activities (Wijnen-Meijer et al. 2013a). They can all be subsumed under the seven medical roles of the Canadian CanMEDS framework (Medical Expert, Scholar, Communicator, Collaborator, Health Advocate, Leader, Professional), which were adopted for the definitions of the NKLM roles. In a ranking study in the Netherlands and Germany, the same ten facets of competence were found on the first places (Wijnen-Meijer et al. 2013b). Those were the basis for this project (Figure 1).

With the introduction of competence-based postgraduate medical education in England, Canada, and the USA, a quest for suitable assessment methods began which resulted in so called “entrustable professional activities” (EPAs) being particularly suitable to achieve a “constructive alignment” between theoretical competence-based postgraduate education and clinical practice (Ten Cate and Scheele 2007). In 1997, the “Patient Assessment and Management Examination” (PAME) was developed as assessment format for physicians in postgraduate education in Canada, where six stations with standardized patients have to be passed and the medical management of these patients is assessed (MacRae et al. 1997). This is a valid and reliable assessment format. A feasibility study for EPAs in internal medicine residency has also been successfully conducted (Hauer et al. 2013). It is also hypothesized that EPAs could be successfully introduced into undergraduate medical education (Chen et al. 2015). A comprehensive assessment for undergraduate medical students based on a competence framework has been established neither nationally nor internationally.

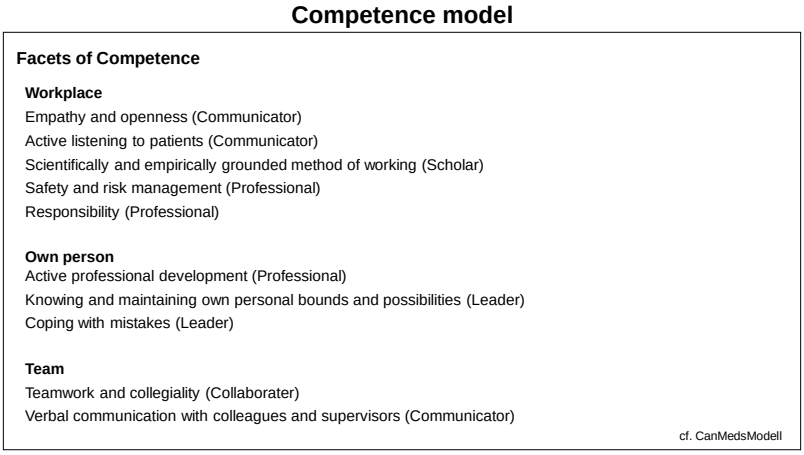


Figure 1: Competence model with facets and roles of competence

Study Design

The new competence-based assessment, which will be further validated with additional instruments and executed in the project described here, simulates the first day of work in a physician’s postgraduate education in a hospital with a consultation-hour, subsequent patient management and a concluding report to a supervisor. The participating students will complete this new assessment twice, the first time at the start of their final practice year (after finishing the Second Section of the Medical Exam) and the second time at the end of their practice year (before taking the Third Section of the Medical Exam). Assessment of the competences is carried out by means of a 360°-feedback (simulated patients, supervisors, peers, nurses) and self-assessment. Different instruments will be applied for the measurement of the different facets of competence. Some of them are already validated and others are still within the process of validation. The collected data will be evaluated quantitatively and qualitatively.

To assess the content validity of the newly designed assessment further, a Delphi-study to select the facets of competence to be studied similar to the one described above will be conducted at the three project sites. For assessment of the discriminant validity, all participating students will take a knowledge test with multiple-choice questions (Raupach et al. 2013). Participants’ exam results of the Second Section of the Medical Exam will serve as benchmark. Furthermore, they can be put in relation to the results of the whole year’s cohort of the respective university and to other universities in Ger-

many, respectively. As a test for convergent validity, the participants will complete a part of the competence assessment, which is used as an aptitude test for aerospace personnel testing the facets of competence teamwork, communication, leadership, and the ability to work under pressure for comparison in addition the newly designed competence-based assessment (Oubaid 2015).

Project Data and Contact

Project Management	Prof. Dr. med. Sigrid Harendza, MME (Coordinator) Prof. Dr. med. Marina Kadmon, MME Prof. Dr. med. Pascal Berberat, MME
Contact (Coordination)	University Hospital Hamburg-Eppendorf Center for Internal Medicine III. Medical Clinic Martinistr. 52 20246 Hamburg, Germany Tel: +4940-7410-54167 Fax: +4940-7410-40218 Email: harendza@uke.de
Time Frame	01 Mar 2016 – 28 Feb 2019

References

Chen, H.C., van den Broek, W.E. & Ten Cate, O. (2015). The case for use of entrustable professional activities in undergraduate medical education. *Acad Med,* 90(4), 431-436.

Berberat, P., Harendza, S. & Kadmon, M. (2013). Entrustable professional activities – visualization of competencies in postgraduate training. Position paper of the committee on postgraduate medical training of the German society for medical education (GMA). *GMS Z Med Ausbild,* 30(4), Doc47.

Epstein, R.M. & Hundert, E.M. (2002). Defining and assessing professional competence. *JAMA*, 287, 226-235.

Erpenbeck, J. & Sauter, W. (2013). *So werden wir lernen!* Berlin, Heidelberg: Springer-Verlag.

Hauer, K.E., Soni, K., Cornett, P., Kohlwes, J., Hollander, H., Ranji, S.R., Ten Cate, O., Widera, E., Calton, B. & O'Sullivan, P.S. (2013). Developing entrustable professional activities as the basis for assessment of competence in an internal medicine residency: a feasibility study. *J Gen Intern Med,* 28(8), 1110-1114.

Harendza, S., Alofs, L., Huiskes, J. & Wijnen-Meijer, M. (2013). Ordering patterns for laboratory and radiology tests by students from different undergraduate medical curricula. *BMC Med. Educ,* 13, 109.

Kadmon, M., Bender, MJ., Adili, F., Arbab, D., Heinemann, MK., Hofmann, HS., König, S., Küper, MA., Obertacke, U., Rennekampff, HO., Rolle, U., Rücker, M., Sader, R., Tingart, M., Tolksdorf, MM., Tronnier, V., Will, B., & Walcher, F.; für die Chirurgische Arbeitsgemeinschaft Lehre der Deutschen Gesellschaft für Chirurgie (CAL). (2013). Competency-based medical education: National Catalogue of Learning Objectives in surgery. *Chirurg,* 4, 277-285.

MacRae, H.M., Cohen, R., Regehr, G., Reznick, R & Burnstein, M. (1997). A new assessment tool: The patient assessment and management examination. *Surgery*, 122(2), 335–344.

Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM). www.nklm.de [Accessed on 04 Jan 2016]

Oubaid, V. Wie und Warum? Psychologische Eignungsuntersuchungen in Luft- und Raumfahrt. web.ev-akademie-tutzing.de/cms/get_it.php?ID=1576 [Accessed on 22 Mar 2015].

Raupach, T., Vogel, D., Schiekirka, S., Keijsers, C., Ten Cate, O. & Harendza, S. (2013). Increase in medical knowledge during the final year of undergraduate medical education in Germany. *GMS Z Med Ausbild*, 30(3), Doc33.

Ten Cate, O. & Scheele, F. (2007). Competency-based postgraduate training: can we bridge the gap be-tween theory and clinical practice? *Acad Med,* 82(6), 542-547.

Van der Blij, M., Boon, J., Van Lieshout, H., Schafer, H., & Schrijen H. Competentieprofielen: over schillen en knoppen. [e-Comope-tence profiles], Utrecht: Digitale Universiteit, 2002.

Wijnen-Meijer, M., van der Schaaf, M., Nillesen, K., Harendza, S. & Ten Cate, O. (2013a). Essential facets of competence that enable trust in graduates: A Delphi study among physician educators in the Netherlands. *J Grad Med* Educ, 5, 46-53.

Wijnen-Meijer, M., van der Schaaf, M., Nillesen, K., Harendza, S. & Ten Cate, O. (2013b). Essential facets of competence that enable trust in medical graduates: A ranking study among physician educators in two countries. *Perspect Med Educ*, 2, 290-297.

Wijnen-Meijer, M., Ten Cate, O., van der Schaaf, M. & Harendza, S. (2013). Graduates form vertically integrated curricula. *Clin Teach,* 10, 155-159.

Wijnen-Meijer, M., van der Schaaf, M., Booij, E., Harendza, S., Boscardin, C., van Wijngaarden, J. & Ten Cate, Th. J. (2013). An argument-based approach to the validation of UHTRUST: Can we measure how recent graduates can be trusted with unfamiliar tasks? *Adv. Health Sci Educ Theory Pract,* 18, 301-307.

Ko-NaMa – Simulation-based Measurement and Validation of a Competence Model for Sustainability Management

Seeber, S., Hartig, J., Dierkes, S. & Schumann, M.

Abstract

The co-operative project Ko-NaMa aims at the validation of a multidimensional competency model for sustainability management. A total of three test components is to be implemented: (1) a test aiming at the measurement of declarative knowledge in the domain of sustainability management, (2) a test assessing declarative and procedural knowledge in business administration, aiming primarily at economic concepts and corresponding efficiency-geared decision-making, and (3) a complex simulation-based test, in which authentic business situations related to problems of sustainability are represented. In addition, several validation studies are projected which are intended to provide empirical evidence substantiating the underlying competency model. These studies comprise analyses examining the curricular validity of assessments in the tertiary sector, the significance of the measured competencies for making decisions, as well as the corresponding construct validity. In the context of an intervention study involving the application of a specific training component for sustainability management in three locations, the effects of specific learning opportunities are to be investigated. Control groups are to be recruited among students from locations, where learning opportunities concerning sustainability management are provided either based on a general societal perspective or by way of integrating them into existing modules in the field of business administration.

Aims and Research Questions

Within the co-operative project Ko-NaMa, three aims are pursued: 1. validation of a competency model for sustainability management, 2. implementation and evaluation of a pertinent intervention, 3. analysis of variables influencing the emergence and development of respective competencies.

Theoretical Framework

The competency model for sustainability management is underpinned by a business model linking a systemic- with a process-based perspective. In this model, enterprises are complex systems embedded in specific changing environments and interact with different stakeholder groups. From a process perspective, the value chain consists of business, management and support processes (see Ruegg-Stürm, 2005), which allows to identify relevant management requirements in enterprises and to analyze the domain of acting. Additionally, we take into account the decision theory which tries to identify the optimal alternative for different preferences in a given situation. Thus, both concepts are suitable to model requirements concerning sustainability management in business contexts in an appropriate way. With regard to sustainability management, a differentiation of sustainability into the three dimensions economy, ecology and social issues is prevailing (Elkington, 1999). Taking this approach as a basis, problems of sustainability in the entire value-added chain can be determined. The implementation of sustainability in companies is connected with decisions which are accompanied by normative ideas, problem formulation, target definition processes and with the design of models which facilitate the identification and assessment of action alternatives. On the basis of a decision-orientated approach, it is at the same time possible to over-

come the contrast between the measuring of "dispositions" (Competence) and the observation of "behavior" (Performance), by including situation analysis, problem definition, target formulation and decision as constitutive factors each (see Blömeke, Gustafsson, & Shavelson, 2015, 7). For this purpose, test methods are applied which are both authentic, i.e. approximated to corporate reality, as well as computer- and simulation-based.

Competence in sustainability management are defined as a complex ability to act adequately in business contexts, in particular to be able to take into account the medium- and long-term economical, ecological and social – intra-company as well as external – consequences of management decisions. Illustration 1 shows the (preliminary) competency model for sustainability management. Preliminary work on competency modelling in commercial and business administration professions showed that such knowledge is of substantial importance – contrary to some assumptions and explanatory models in the area of environmental education (see critical opinion by Lehmann, 1999; Roczen, 2011). These preliminary studies have shown, however, that decisions on sustainability do not exclusively depend on the knowledge about sustainability, but that affectional-motivational dispositions regarding sustainability are important determinants of the decision on how to act, as well (see Seeber & Michaelis, 2014). Since the situations concerning sustainability management are differing in complexity, competency modelling takes into account differently demanding modelling capabilities in terms of cognition as well as different levels of information networking.

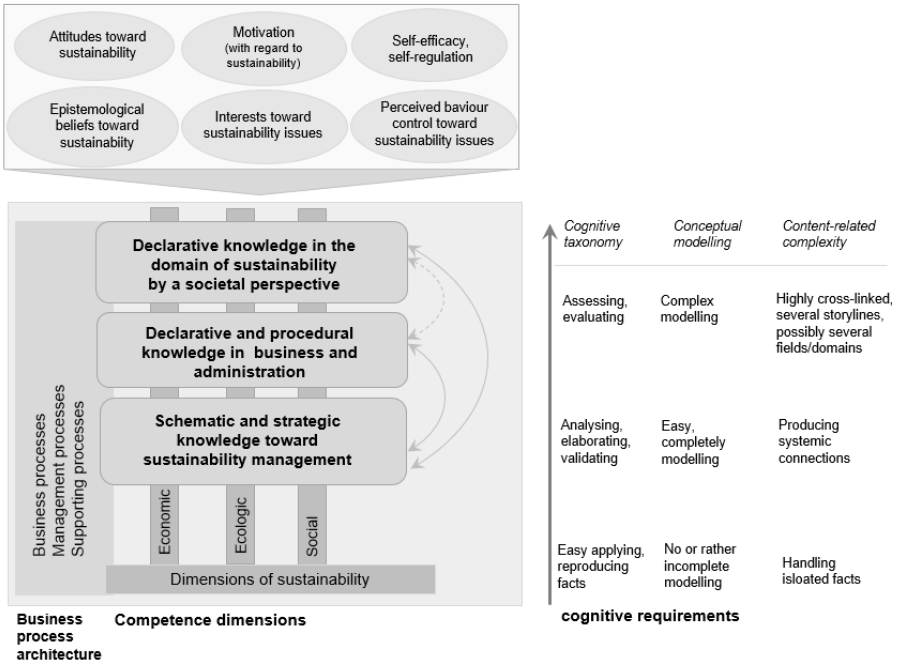


Fig. 1: competency structure model "Operational sustainability management (see Seeber, & Michaelis, 2014; Winther, 2010; kinds of knowledge see Shavelson, Ruiz-Primo, & Wiley, 2005, Rüegg-Stürm, 2005).

Study Design

The central aim of the project is the validation of the competency model of sustainability management. The competency model and the planned assessment takes into account different kinds of knowledge. Thus, the measurement requires also different test accesses (see Shavelson, Ruiz-Primo, & Wiley, 2005, p. 415-416). The declarative knowledge about generally prospects for sustainability as well as the declarative knowledge in business administration will be assessed by closed and brief open answer test for-

mats. For measuring procedural, schematic and strategic knowledge closely related to action and business decision making in complex situations (including perception and interpretation of a specific management situation, problem formulation, target definition, perspective taking, decision making a.s.o.) a performance assessment is needed and may mediate between disposition and performance. For this purpose, a computer-based authentic assessment with simulations of real business problems under the perspective of sustainability in companies will be implemented.

First curricular analyses on the integration of sustainability issues into business administration studies refer to three central guiding principles of anchoring: (a) Basic modules on sustainability are offered from a societal perspective as interdisciplinary approach and thus usually without domain-specific integration (b) Sustainability issues are included into different basic and specialized business administration modules on an integrative level. (c) Specific modules on sustainability management are offered along selected operational function areas or on a cross-functional level. In the context of the planned intervention study involving the application of a specific training component for sustainability management at three universities, the effects of specific learning opportunities are to be investigated. Control groups are to be recruited among students from universities, where learning opportunities for sustainability management are provided either based on a general societal perspective or by way of integrating them into exiting business administration modules. Furthermore, experts' judgments of the representativeness and quality of tasks sampled are used to examine validity claims.

Test items have been developed which are taken to be scalable according to Latent-Trait Models (Rasch and Partial-Credit) and/or Latent-Class Models, if applicable. The structure assumed in the competency model will be subjected to confirmatory procedures. With regard to the intervention and control groups, the invariance of measurement between the groups will be tested by appropriate methods (e. g., multigroup-IRT analyses). In this context, "differential item functioning" (DIF) should not only be considered as a source of error variance and as indicative of a lack of criterion validity. Instead, it could be considered as a result of different learning opportunities leading to different group profiles with regard to the same construct (Scheuneman, & Gerritz, 1990). In order to measure changes in the proportions of students meeting proficiency in sustainable management over two occasions within the intervention study, growth models are used in which the performance at the different times of testing is scaled in terms of two latent dimensions. The analysis of changes over time is facilitated by the application of anchor items and the use of a multi-matrix design.

Project data

Project management:	Prof. Dr. Susan Seeber (Georg-August-University of Göttingen) Prof. Dr. Johannes Hartig (German Institute for International Educational Research (DIPF), Frankfurt) Prof. Dr. Stefan Dierkes (Georg-August-University of Göttingen) Prof. Dr. Matthias Schumann (Georg-August-University of Göttingen)
Contact	Prof. Dr. Susan Seeber Georg-August-University of Göttingen, Department of Business Education and Human Resource Development Platz der Göttinger Sieben 5; 37073 Göttingen Phone: +49 (0) 551 39 4421; Email: susan.seeber@wiwi.uni-goettingen.de
Homepage	http://www.uni-goettingen.de/en/531774.html

Time frame 01 Jan 2016 – 31 Dec 2018

References

Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223(1), 3-13.

Elkington, J. (1999). Triple bottom-line reporting: Looking for balance. *AUSTRALIAN CPA*, 69, 18-21.

Lehmann, J. (1999). *Befunde empirischer Forschung zu Umweltbildung und Umweltbewusstsein*. Opladen: Leske+ Budrich.

Roczen, N. (2011). *Environmental competence - The interplay between connection with nature and environmental knowledge in promoting ecological behaviour*. Eindhoven: University of technology Library.

Rüegg-Stürm, J. (2005). *Das neue St. Galler Management-Modell*. Bern: Haupt Verlag.

Scheuneman, J. D., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, 27(2), 109-131.

Seeber, S. & Michaelis, C. (2014). *Development of a Model of Competencies Required for Sustainable Economic Performance among Apprentices in Business Education*. http://www.aera.net/Publications/OnlinePaperRepository/tabid/10250/Default.aspx

Shavelson, R. J. (2012). Assessing business-planning competence using the Collegiate Learning Assessment as a prototype. *Empirical Research in Vocational Education and Training*, 4(1), 2012, 77–90.

Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (2005). Windows into the mind. *Higher education*, 49(4), 413-430.

Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. Bielefeld: W. Bertelsmann Verlag.

WiWiKom II - Valid Assessment of Students' Development of Professional Business and Economic Competencies over the Course of their Studies - A Quasi-experimental Longitudinal Study

Zlatkin-Troitschanskaia, O., Pant, H. A., Förster, M., Brückner, S. & Fox, J.-P.

The WiWiKom II project is based on the competency model developed and tested and the test instrument used in WiWiKom I (Zlatkin-Troitschanskaia et al., 2014). The overarching aim of WiWiKom II is objective, reliable, and valid assessment of students' *development* of professional business and economic competencies that allows for valid interpretations of test results. Building on the validation approach from WiWiKom I (based on AERA, AEA & NCME, 2014; Kane 2013), WiWiKom II addresses more in-depth validation questions and expands the results gained so far by including individual *change* measurement.

In WiWiKom II, we implement a longitudinal design over a period of four years and four measuring dates. The project follows a quasi-experimental design, examining constructs according to the multitrait-multimethod matrix approach and two group comparisons for convergent, discriminant, incremental, and predictive validation. Quasi-experimental variation is achieved through sampling by including not only two study domains (business & economics and social studies) but also different stages of training (orientation, advanced, and specialization study phases in the Bachelor degree course). In order to conduct discriminant validation, both the WiWiKom test and an abridged version of the *Berliner Test zur Erfassung fluider und kristalliner Intelligenz* [Berlin test for the assessment of fluid and crystalline intelligence] (BEFKI) (Schipolowski, Schroeders & Wilhelm, 2014) are implemented. The BEFKI is a general, subject-independent knowledge test that assesses general cognitive abilities (GCA) in terms of fluid and crystalline intelligence.

In WiWiKom II, professional business and economic competencies and their development over time are modeled as central dependent variables. In a longitudinal study, we examine the extent to which these dependent variables can be explained by variables related to business and economics degree courses (e.g., learning opportunities attended over the course of Bachelor studies). Further validation questions concern the differentiation in measurement between the dependent variables and theoretically related criteria (such as GCA) as well as correlations between the dependent variables and construct-relevant external criteria (such as Bachelor degree final grade). In a quasi-experimental comparative analysis, we examine further aspects of discriminant validity by analyzing the manifestation of the dependent variables among the target group and a control group in cross-sectional and longitudinal studies. Altogether, WiWiKom II focuses on four validation aspects:

(1) *Cross-sectional and longitudinal convergent validation*: In a longitudinal study, the development of professional business and economics competencies is assessed and analyzed in relation to learning opportunities attended during and prior to university studies. One research focus is on the influence of learning opportunities within a business and economics degree course on the dependent variables.

(2) *Cross-sectional and longitudinal discriminant validation*: We analyze the differentiation, that is, the empirical separability, between GCA and professional business and economic competencies. In a longitudinal study, we examine the extent to which GCA can explain the increase in the dependent variables and whether this influence remains constant over the course of studies.

(3) *Incremental and predictive validation*: In longitudinal studies, we gather evidence on the ability of the test to predict study success (grade and completion of a Bachelor degree course) or successful transition into a Master degree course or into professional practice. Here, we focus on two external criteria (e.g., grade and transition) that are central to study success. The WiWiKom test should be able to explain significant proportions of variance in the external criteria and should have greater explanatory power than domain-independent predictive factors, such as university entrance qualifications and intelligence tests. Furthermore, we assume that students' degree of expertise will increase over the course of three years of studies, which, according to the competency model, should become evident in an increase in the level of competency in business and economics.

(4) *Discriminant validation by means of the known-group approach*: The specificity of the test to the study domain of business and economics is analyzed by means of a comparison between two groups of students with different opportunities for acquiring professional business and economic competencies. We assume that the target group of business and economics students acquire business and economic knowledge by attending courses in this field of studies. In contrast, exploratory findings from WiWiKom I suggest that the control group of students in social studies primarily relies on competencies not specific to this field and should therefore rather refer to everyday experiences or knowledge from prior education when responding to business and economic test items. By means of the group comparison (Förster et al., 2015), we also examine how sensitive to changes the WiWiKom test is by comparing the progression of competency development while controlling for other study-related influence factors.

To test our hypotheses, we conduct comprehensive analyses. They offer an insight into the students' developmental progress in acquiring competency in business and economics as well as GCA. Competencies are modeled on the basis of Item-Response Theory and Structural Equation Modeling (Wilson, 2005). Comparability of groups is tested by means of measurement invariance analyses in cross-sectional and longitudinal studies. Due to expected missing values in the longitudinal data, multi-level models as well as growth curve models are estimated according to frequentist and Bayesian approaches (Schmidt et al., in review). Furthermore, propensity score matching is implemented in order to enable valid comparisons of groups with different sizes and distributions of covariates (e.g., different higher education institutions, different gender ratios etc.) in cross-sectional and longitudinal studies.

Project data

Project management	Prof. Dr. Olga Zlatkin-Troitschanskaia (Johannes Gutenberg University Mainz, JGU) Prof. Dr. Hans Anand Pant (Humboldt-University of Berlin, HU) JP Dr. Manuel Förster (JGU) Prof. Dr. Jean-Paul Fox (University of Twente)
Contact	Dr. Sebastian Brückner, JGU Mainz Jakob-Welder-Weg 4, 55128 Mainz Tel: +49 6131-39-22009
Homepage	www.wiwi-kompetenz.de
Time Frame	01 Nov 2015 – 31 Oct 2019

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2004, 2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Förster, M., Zlatkin-Troitschanskaia, O., Brückner, S., Happ, R., Hambleton, R. K., Walstad, W. B., . . . Yamaoka, M. (2015). Validating Test Score Interpretations by Cross-National Comparison: Comparing the Results of Students From Japan and Germany on an American Test of Economic Knowledge in Higher Education. *Zeitschrift für Psychologie*, *223*(1), 14-23.

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. Journal *of Educational Measurement*, 50(1), 1–73.

Schipolowski, S., Schroeders, U., & Wilhelm, O. (2014). Pitfalls and Challenges in Constructing Short Forms of Cognitive Ability Measures. *Journal of Individual Differences, 35*(4), 190-200.

Schmidt, S., Fox, J.-P., & Zlatkin-Troitschanskaia, O.. A Multilevel IRT Model for Hierarchical and Longitudinal Data for Valid Assessment of Competence Growth among Students in Higher Education. In *Journal of Educational Measurement*, Volume 4/2016.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ.: Erlbaum.

Zlatkin-Troitschanskaia, O., Förster, M., Brückner, S. & Happ, R. (2014). Insights from a German assessment of business and economics competence. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment – International Perspectives* (pp. 175-197). Frankfurt am Main: Lang.

WiWiSET – Validation of an Entrance Examination in the Study Domain of Business and Economics – A National and International Comparative Study of Universities and Universities of Applied Sciences

Pant, H. A., Zlatkin-Troitschanskaia, O., Schipolowski, S. & Förster, M.

The WiWiSET project employs a German edition (so called TEL-D) of the standardized Test of Economic Literacy IV (TEL IV), developed by the U.S. Council for Economic Education (CEE) (Walstad, Rebeck & Butters, 2013). The test has been translated according to the TRAPD process model (including translation, review, adjudication, pretesting, and documentation) (Harkness, 2008), and has been adapted according to the International Test Commission's Guidelines on Test Adaptation (ITC, 2010) for initial diagnostic assessment in higher education in the domain of business and economics. We analyze economics-related competencies at the beginning of studies in an objective, reliable, and valid way. Furthermore, we examine correlations with additional, theoretically significant personal variables, such as prior knowledge and motivation for enrolling in business and economics studies, as well as institutional variables, such as types of higher education institution and degree course. By means of various statistic models, such as structural equation modeling and multilevel analysis, we examine the effects that can be explained, according to the preliminary study, by students' individual preconditions, by differences between degree courses, such as business administration and economics, and by differences between higher education institutions, that is, universities compared to universities of applied sciences. Among others, we examine whether the test enables measurement-invariant assessment of economics-related competencies at the beginning of studies.

Building on validation work from the preliminary study, conducted according to the international Standards for Educational and Psychological Testing (AERA et al., 2014), the WiWiSET project pursues discriminant, predictive, and incremental validation perspectives. We examine whether subject-related competencies at the beginning of studies assessed by means of the TEL-D can be empirically differentiated from general cognitive abilities (GCA) measured by means of intelligence tests and university entrance qualifications (discriminant validation). Moreover, we use the TEL-D to predict subject-related academic performance or study success in economics (predictive validation). We expect that the TEL-D should be able to explain significant proportions of variance in subject-related academic performance and should have greater explanatory power than subject-independent predictors such as intelligence test scores and university entrance qualifications (incremental validation). In addition, predictive validation can include prediction of dropouts; we examine the extent to which business and economics students' results on the TEL-D at the beginning of their studies can predict discontinuation of studies within the first academic year, which is the most critical one in this study domain. To be judged a suitable predictor of study success at the beginning of studies and to serve as a useful information tool for intervention planning in higher education teaching, the TEL-D must increase the reliability of predictions of study success when used alongside university entrance qualifications.

The population for the WiWiSET study includes students at the beginning of a Bachelor's degree program in the domain of business and economics (the largest study domain in Germany) at a university or university of applied sciences in Germany. As an efficient selection process, we will use a cluster sample with clustering according to types of higher education institutions (universities or universities of applied sciences). In the sampled clusters, we will assess first-year students in the domain of business and economics. This method of sampling implies that the data has a nested structure, as students are clustered in higher education institutions. This nesting must be considered in the data analysis in order to distinguish intra-institutional differences among students at one institution from inter-institutional differences between students from different institutions. This aspect of the multilevel approach defines the required sample size in terms of the amount of clusters (level 2 units) and the number of students (level 1 units). To estimate ideally undistorted and generalizable fixed and random effects, we must assess at least 30 higher education institutions (Hox, 2010). The German Federal Statistical Office has provided a list (04/2015) of all higher education institutions included in the population, from which we will choose 40 institutions at random. For each geographical area, we have determined two substitute institutions with similar characteristics (such as similar degree courses and department sizes).

To test the research hypotheses of the WiWiSET project, we need two assessment rounds. In the first round of assessment, which is scheduled to take place before the beginning of the winter term 2016 (September-October 2016), first-year students will be assessed at randomly chosen higher education institutions during introductory courses in economic studies. In the second round of assessment to take place at the beginning of the winter term 2017, the same students will be assessed a second time after having completed one year of studies. This way, we aim to assess their study progress during the orientation phase of the Bachelor's degree program so as to examine the predictive validity of the TEL-D in terms of academic courses completed during the first year of studies.

Project data

Project management	Prof. Dr. Hans Anand Pant (Humboldt University of Berlin)
Homepage	www.wiwi-kompetenz.de

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2004, 2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Harkness, J. A. (2008). Comparative survey research: Goals and challenges. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology*. New York: L. Erlbaum Associates. 56–77.

Hox, J. J. (2010). *Multilevel Analysis – Techniques and Applications* (2nd ed.). New York: Routledge.

International Test Guidelines (ITC). (2010). *Guidelines for Translating and Adapting Tests: Version 2010*. Retrieved from http://www.intestcom.org/upload/sitefiles/40.pdf

Walstad, W. B., Rebeck, K., & Butters, R. B. (2013). *Test of economic literacy: Examiner's manual.* New York: Council for Economic Education.

Previously published:

KoKoHs Working Papers, 1

Blömeke, S. & Zlatkin-Troitschanskaia, O. (2013). Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor: Ziele, theoretischer Rahmen, Design und Herausforderungen des BMBF-Forschungsprogramms KoKoHs [Modeling and Measuring Competencies in Higher Education: Aims, theoretical framework, design, and challenges of the BMBF-funded research program KoKoHs] (KoKoHs Working Papers, 1). Berlin & Mainz: Humboldt University & Johannes Gutenberg-University.

KoKoHs Working Papers, 2

Blömeke, S. (2013). Validierung als Aufgabe im Forschungsprogramm "Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor" [The task of validation in the research program „Modeling and Measuring Competencies in Higher Education"] (KoKoHs Working Papers, 2). Berlin & Mainz: Humboldt University & Johannes Gutenberg-University.

KoKoHs Working Papers, 3

Blömeke, S. & Zlatkin-Troitschanskaia, O. (Eds.) (2013). The German funding initiative "Modeling and Measuring Competencies in Higher Education": 23 research projects on engineering, economics and social sciences, education and generic skills of higher education students (KoKoHs Working Papers, 3). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

KoKoHs Working Papers, 4

Berger, S., Hammer, S., Hartmann, S., Joachim, C. & Lösch, T. (2013). Causal Inference in Educational Research. Approaches, Assumptions and Limitations. (KoKoHs Working Papers, 4). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

KoKoHs Working Papers, 5

Toepper, M., Zlatkin-Troitschanskaia, O., Kuhn, C., Schmidt, S. & Brückner, S. (2014). Advancement of Young Researchers in the Field of Academic Competency Assessment – Report from the International Colloquium for Young Researchers from November 14-16, 2013 in Mainz (KoKoHs Working Papers, 5). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

KoKoHs Working Paper, 6

Kuhn, C., Toepper, M., & Zlatkin-Troitschanskaia, O. (2014). Current International State and Future Perspectives on Competence Assessment in Higher Education – Report from the KoKoHs Affiliated Group Meeting at the AERA Conference on April 4, 2014 in Philadelphia (USA) (KoKoHs Working Papers, 6). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

KoKoHs Working Paper, 7

Brückner, S., Dunekacke, S., & Happ, R. (2014). Report from the "AERA Institute on Statistical Analysis for Education Policy" from 6th till 9th May 2014 in Washington, DC. (KoKoHs Working Papers, 7). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

KoKoHs Working Paper, 8

Lautenbach, C., & Schulz, K. (Eds.) (2015). Developing International Research Projects in the Field of Academic Competency Assessment – Report from the "KoKoHs-Autumn Academy" from October 6-10, 2014 in Berlin. (KoKoHs Working Papers, 8). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

KoKoHs Working Paper, 9

Toepper, M., & Zlatkin-Troitschanskaia, O. (2015). Modelling and Measuring Competencies in Higher Education - Report from the International Conference at Johannes Gutenberg University Mainz from 28-29 November 2014. (KoKoHs Working Papers, 9). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

KoKoHs Working Paper, 10

Zlatkin-Troitschanskaia, O., Pant, H.A., Lautenbach, C. & Toepper, M. (2016). Kompetenzmodelle und Instrumente der Kompetenzerfassung im Hochschulsektor – Validierungen und methodische Innovationen (KoKoHs): Übersicht der Forschungsprojekte (KoKoHs Working Papers, 10). Berlin & Mainz: Humboldt University & Johannes Gutenberg University.