Olga Zlatkin-Troitschanskaia
Hans Anand Pant · Miriam Toepper
Corinna Lautenbach *Editors*

Student Learning in German Higher Education

Innovative Measurement Approaches and Research Results

Springer VS

Student Learning in German Higher Education

Olga Zlatkin-Troitschanskaia · Hans Anand Pant · Miriam Toepper · Corinna Lautenbach
Editors

Student Learning in German Higher Education

Innovative Measurement Approaches and Research Results

Springer VS

Editors
Olga Zlatkin-Troitschanskaia
Johannes Gutenberg University Mainz
Mainz, Germany

Hans Anand Pant
Humboldt University of Berlin
Berlin, Germany

Miriam Toepper
Johannes Gutenberg University Mainz
Mainz, Germany

Corinna Lautenbach
Humboldt University of Berlin
Berlin, Germany

ISBN 978-3-658-27885-4 ISBN 978-3-658-27886-1 (eBook)
https://doi.org/10.1007/978-3-658-27886-1

This Springer VS imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH part of Springer Nature.
The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Contents

Modeling and Measuring Competencies in Higher Education

1

The KoKoHs Program

Zlatkin-Troitschanskaia, O., Pant, H. A., Toepper, M., and Lautenbach, C.

Over the past decade, tertiary education has increasingly been gaining importance in society. Developments such as the continuously growing number of students in higher education and increasing student mobility have raised questions of efficiency and effectiveness in tertiary education, calling for valid assessments of competencies and student learning outcomes. Assessments of the output of higher education can yield important evidence regarding the effectiveness of this highly important educational sector and thus provide a basis for improvement measures at the individual and institutional levels (Coates and Zlatkin-Troitschanskaia 2019).

Modern higher education focuses on the acquisition of domain-specific knowledge and on the development and promotion of generic (interdisciplinary) skills (e.g. critical thinking), which, according to current surveys amongst employers, are increasingly gaining significance in the 21st century (Association of German Chambers of Industry and Commerce (DIHK) 2015). Such a competence portfolio, acquired over the course of academic studies, is crucial for all professionals and globally engaged citizens and allows for lifelong learning, which is necessary in today's continuously changing age of information.

Despite this sociopolitical consensus and the growing competence orientation (in the context of the Bologna reform), there have been only few evidence-based in-

O. Zlatkin-Troitschanskaia et al. (Hrsg.), *Student Learning in German Higher Education* https://doi.org/10.1007/978-3-658-27886-1_1

sights into this field up until the last decade, particularly regarding the competencies of higher education students. Therefore, the German Federal Ministry of Education and Research established the Germany-wide research initiative "Modeling and Measuring Competencies in Higher Education (KoKoHs)"[1] in 2011 and – after a positive external evaluation in 2015 – decided to continue to fund this research in the context of the German program "Modeling and Measuring Competencies in Higher Education (KoKoHs) – Validation and Methodological Innovations" until 2020.

In the first research program, KoKoHs (2011–2015), more than 220 researchers from various fields such as subject-specific didactics, learning psychology, and psychometrics developed first modeling approaches and the corresponding measuring instruments for the valid assessment of student competencies in the context of 24 collaborative research projects at over 70 universities and research institutes, focused on central study domains such as business and economics, engineering, and teacher education (for a detailed description of the first KoKoHs research program (2011–2015) and the individual projects and results, see Zlatkin-Troitschanskaia et al. 2017). These models and tools developed in KoKoHs were one of the key results of this first working phase, which ran until 2015. Another equally important outcome of this research phase were the findings on students' competence levels in different study phases, which revealed many deficits. At the same time, the generalizability of these results was questionable, as some of the newly developed KoKoHs instruments had not yet been comprehensively validated in accordance with a number of validation criteria as recommended in the Standards for Psychological and Educational Testing by AERA et al. (2014). Another shortcoming had been the fact that most of the newly developed test instruments were paper-pencil-based and altogether only few innovative assessments had been developed in the first phase. Based on the results and recommendations from an international audit at the end of the first phase, the second phase of the KoKoHs program was launched nationwide in 2015 with a focus on validation and methodological evaluations.

In this follow-up research program, KoKoHs (2016–2020), more than 100 researchers comprehensively validated KoKoHs assessments and developed new innovative modeling approaches and the corresponding measuring instruments for the valid assessment of student competencies in the context of 16 collaborative research projects at over 40 universities and research institutes, again focused on central study domains such as business and economics as well as teacher education. In this program, one new study domain was included: medicine. Moreover, some of

1 For further information on KoKoHs, see https://www.blogs.uni-mainz.de/fb03-kokohs-eng/

the projects focused on transferring and adapting modeling approaches and assessments from one domain to another (e.g. from mathematics to economics). Overall, this program consists of three large clusters: four projects focusing on domain-specific competencies in economics and medicine, five projects with a focus on domain-independent competencies such as scientific reasoning and self-regulation skills (for domain-specific and generic competencies, see Zlatkin-Troitschanskaia, Pant, and Greiff 2019), and the largest cluster with seven projects and a focus on teacher education in different domains such as mathematics, physics, or economics (for teachers' competencies, see Cortina, Pant, and Zlatkin-Troitschanskaia 2019).

A common focus of all projects was the in-depth validation of KoKoHs assessments following the validation criteria of AERA et al. (2014). Most projects were also characterized by their focus on the development and validation of complex technology-based assessments, which are mostly performance-oriented (for performance assessment, see Zlatkin-Troitschanskaia and Shavelson 2019). Innovative technology-based test formats such as computer-based learning diaries or mobile apps were also developed and implemented. In this research phase, some of the projects have had a longitudinal design, which has allowed for valid statements about the development of competencies over the course of academic studies. In addition, several instruments developed and validated in KoKoHs have now also been tested and used in many other countries such as Japan, the US, and China, and comparative analyses have already been carried out (for cross-national studies, see Zlatkin-Troitschanskaia et al. 2018).

Overall, in the 40 collaborative KoKoHs projects (which, in turn, comprised about 100 individual projects), theoretical-conceptual competence models and corresponding measurement instruments were developed and successfully validated for selected large study domains (e.g. economics, teacher education, STEM). These models differentiate, reliably describe, and assess the competences of students in different phases of higher education – entry, undergraduate and postgraduate studies. Over 100 newly developed innovative video-, computer-, and simulation-based test instruments were validated across Germany at more than 350 universities with over 75,000 undergraduate and master's students. The assessments focused on both discipline-specific competencies and generic skills, which students and graduates should acquire over the course of their studies and which employers and other stakeholders expect according to the professional and social requirements of the 21st century.

Building on best practices from the first funding phase of the KoKoHs program (2011–2015), the subsequent funding phase ran from 2015 to 2020 and brought together experts from various fields and with different methodological backgrounds in cross-university project alliances within a joint international and interdiscipli-

nary research network. Based on the models and instruments for the reliable and valid assessment of competencies acquired in various study domains in higher education that were developed and empirically tested in the first funding phase, this follow-up research phase of KoKoHs aimed to increase the explanatory power and broaden the scope of use of the KoKoHs test instruments through in-depth validation and to drive methodological innovation in higher education competency assessment.

KoKoHs is the only existing nationwide program in which students' learning outcomes in higher education are systematically, validly, and objectively assessed and analyzed. The KoKoHs program provides unique findings on the acquisition and development of students' competencies in German higher education, which form a significant basis for the optimization of learning and teaching practice.

This book is based on the research and development work conducted in KoKoHs over the past decade and offers a comprehensive overview of current innovative tools and approaches to assessing domain-specific and generic student learning processes and learning outcomes in higher education. It presents the work of all KoKoHs projects, thus offering an insight into the most significant research program focused on student learning outcomes in higher education to date. In this volume, innovative modeling and measuring approaches as well as the newly developed objective, valid, and reliable assessment tools for student learning in higher education are presented and critically discussed, with a particular focus on using the developed models and assessments in both further research and higher education practice.

In addition to presenting key conceptual and methodological findings from work within the KoKoHs program, the 88 authors in this book also present key research results and lessons learned from their research to provide new insights into how student learning in higher education can be assessed in various contexts and to show what we can learn from the assessment results. Most contributions also provide an outlook on possible approaches to implementing the instruments into teaching and learning practice and transfer studies. The authors also give a few examples of how higher education practitioners in particular can effectively support teaching and learning at their universities by using the KoKoHs assessments and tools.

With its very broad spectrum of contributions focused on both innovative research and the practical application of assessments in higher education, this volume offers valuable insights for scientists in higher education research as well as related disciplines such as psychology, educational sciences, lecturers in university practice, university evaluation, accreditation agencies, higher education pol-

icy-makers, students, companies and all other stakeholders interested in higher education student learning outcomes.

We would like to thank everyone who contributed to this book. This includes, of course, the 88 authors from the KoKoHs projects and all of the researchers and student assistants who contributed to the work conducted in the KoKoHs program and documented in this volume. We would like to thank all national and international critical advisors of this program, especially *Daniel Koretz, Fritz Oser, James Pellegrino*, and *Richard Shavelson*, who have significantly supported the work conducted in this program over the past decade. Our sincere thanks also go to all of our colleagues who provided external reviews of the contributions and thus contributed significantly to the quality of the articles in this volume. Special thanks go to the sponsor of the KoKoHs program, the German Federal Ministry for Education and Research, which, thanks to its long-term support, has enabled us to carry out sustainable research and development in this field for almost a decade now, thus also contributing to the emergence of a new field of research and to establishing empirical research in higher education in a sustainable manner. In this context, we would like to thank *Martina Diegelmann* in particular, who has critically supervised the program over the past decade and has decisively contributed to its structural and conceptual development. We would also like to thank the DLR project management agency for providing administrative support to all KoKoHs projects.

Many others were involved in the preparation of this book, including our student assistants in KoKoHs and *Mirco Kunz* in particular, who was responsible for the technical preparation of the manuscript, as well as our staff members from the field of translation studies, *Katja Kirmizakis* and *Annika Weibell*, who proofread the contributions in this volume as well as this article.

References

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

Association of German Chambers of Industry and Commerce (DIHK) (2015). *Licht und Schatten (DIHK-Onlineumfrage zur Berufsschulsituation in den IHK-Regionen)*. Verfügbar unter http://www.dihk.de/themenfelder/aus-und-weiterbildung/schule-hochschule/schule/umfragen-und-prognosen/dihk-berufsschulumfrage

Coates, H., & Zlatkin-Troitschanskaia, O. (2019). The Governance, Policy and Strategy of Learning Outcomes Assessment. *Higher Education Policy, 32*(19), 1–6.

Cortina, K., Pant, H. A., & Zlatkin-Troitschanskaia, O. (2019). Kompetenzerwerb zukünftiger LehrerInnen in der universitären Ausbildung. [Themenheft]. *Zeitschrift für Pädagogik (4/2019)*.

Zlatkin-Troitschanskaia, O., Pant, H. A., & Greiff, S. (2019). Assessing Academic Competencies in Higher Education. [Special Issue]. *Zeitschrift für Pädagogische Psychologie, 33*(2).

Zlatkin-Troitschanskaia, O., & Shavelson, R. J. (2019). Performance Assessment of Student Learning in Higher Education [Special issue]. *British Journal of Educational Psychology, 89*(3).

Zlatkin-Troitschanskaia, O., Pant, H. A., Lautenbach, C., Molerov, D., Toepper, M., & Brückner, S. (2017). *Modeling and Measuring Competencies in Higher Education. Approaches to Challenges in Higher Education Policy and Practice*. Wiesbaden: Springer.

Zlatkin-Troitschanskaia, O., Toepper, M., Pant, H. A., Lautenbach, C., & Kuhn, C. (Eds.) (2018). *Assessment of Learning Outcomes in Higher Education – Cross-national Comparisons and Perspectives*. Wiesbaden: Springer.

Teachers' Competence Development 2

2.1
Putting Educational Knowledge of Prospective Teachers to the Test

Further Development and Validation of the BilWiss Test

Kunter, M., Kunina-Habenicht, O.[1], Holzberger, D., Leutner, D., Maurer, C., Seidel, T., and Wolf, K.

Abstract

Teachers' generic educational knowledge theoretically constitutes an aspect of their professional competence. However, empirical evidence for its importance for teachers' daily practice is scarce. In this chapter, we describe findings from the BilWiss research program, which aimed to investigate the development and relevance of the type of generic educational knowledge typically addressed in university teacher education. We developed a standardized test that assesses generic knowledge in the following six domains: learning and development, instruction, assessment, educational theory (and history), school system and educational policy, and the teaching profession. We present findings from a series of studies that (a) provide evidence for the validity of the test score interpretations and (b) prove the predictive value of this test for diverse professional activities. These results are discussed regarding their theoretical and practical implications for teacher education.

1 Mareike Kunter and Olga Kunina-Habenicht are both first authors of this contribution.

O. Zlatkin-Troitschanskaia et al. (Hrsg.), *Student Learning in German Higher Education* https://doi.org/10.1007/978-3-658-27886-1_2

Keywords

Educational knowledge; pedagogical knowledge, assessment, knowledge test; validity, teacher education, test development

1 Teachers' Educational Knowledge as an Aspect of Their Professional Competence

In research on teachers' professional competence there is a consensus that teachers' success in providing high-quality instruction is connected to their profession-specific declarative and procedural knowledge (Baumert et al. 2010; Kennedy et al. 2008; Schleicher 2016; Shulman 1986). Thus, the foundations of professional competence are partly laid in the theoretical part of teacher education at university, which aims particularly at providing subject-specific and generic knowledge in formal learning settings. While the importance of a professional knowledge base per se is uncontested, it is less clear what content such a knowledge base should include (Zeichner 2005).

In his seminal work, Shulman (1986) distinguishes between content knowledge, pedagogical content knowledge, and pedagogical knowledge. The first two describe subject-specific knowledge, which has received much research attention in recent years. There is ample evidence that knowledge about the subject matter itself, subject-specific forms of instruction, and typical student thinking in a domain are important prerequisites for high-quality instruction and, thus, student learning (Abell 2008; Baumert et al. 2010; Depaepe et al. 2013; Hill et al. 2005). The part of teacher knowledge that transcends subject matter has received much less attention in current research. Shulman (1986) defines pedagogical knowledge as "knowledge of generic principles of classroom organization and management" and notes that "proper professional board examination would include other equally important sections as <…> knowledge of general pedagogy, knowledge of learners and their backgrounds, principles of school organization, finance and management, and the historical, social, and cultural foundations of education" (p. 14). In line with Shulman's argumentation we use the term "educational knowledge" that extends the narrow conception of pedagogical knowledge and define it as "teachers' subject-unspecific professional knowledge that comprises both classroom-related topics (instruction, learning and development, and assessment) as well as context-related topics (e.g., knowledge on the educational system, school development, or educational theory and history)" (Linninger et al. 2015, p. 73).

Voss et al. (2015) provided a comprehensive overview of recent developments and assessments measuring pedagogical teacher knowledge. This review revealed that previous research mainly focused on pedagogical topics that are closely related to instruction (e.g., classroom management, learning support) and largely ignored topics addressing matters outside the classroom such as principles of school organization, historical foundations of education, or knowledge about the teacher profession. Previous research revealed small significant correlations between pedagogical knowledge (in a narrow sense) and the instructional quality rated by school students (König and Pfanzl 2016; Voss et al. 2014). Nevertheless, it is not clear which benefit the broader generic educational knowledge might have both for teaching situations and for situations outside of classroom.

To close this gap, the BilWiss[2] research program aimed to investigate both the empirical structure of educational knowledge and its determinants and consequences. In this chapter, we summarize the theoretical background and the main results of the BilWiss project.

2 Background: Educational Foundation Courses in Teacher Education

The structure of teacher education varies across and within countries. University teacher training usually involves courses covering general educational topics and the study of one or more specific subjects. In general, two different models of teacher education are common: the concurrent model where subject courses, educational courses, and practical experiences are combined within one course of study and the consecutive model where a disciplinary degree (i.e., subject-specific) is followed by a degree in education (EURYDICE, 2002). The relative importance of subject-specific and generic parts (expressed in allocated credit points) differs both between and within different countries and institutions (Schmidt et al. 2011; 2008). However, in all systems, teacher students have to attend courses that aim at providing the generic educational knowledge seen as particular to the profession of teachers, the so-called "educational foundation courses".

2 BilWiss stands for „Bildungswissenschaftliches Wissen", which is the German translation of "educational knowledge". The full name of the research program is "Bildungswissenschaftliches Wissen als Teil professioneller Kompetenz in der Lehramtsausbildung [Educational Knowledge as a Part of Professional Competence in Teacher Education]". For more information see: https://bilwiss.paedpsych.de/

Educational foundation courses have a long tradition as a relevant part of teacher education (for a historical review, see Tozer and McAninch 1986) and are assumed to provide important learning opportunities for the acquisition and construction of an educational knowledge base. Educational foundations are defined as a "broadly-conceived field of educational study that derives its <…> methods from a number of academic disciplines <…> including: history, philosophy, sociology, <…> psychology, <…>, educational studies <…>" (Council for Social Foundations of Education (CSFE) 1996, p. 3). In the last decades, there has been a vivid discussion about the nature of knowledge that should be taught within educational foundation courses in university teacher education programs (Hollins 2011; Patrick et al. 2011). Wilson and colleagues have reviewed studies on the impact of "pedagogical knowledge" on teacher effectiveness (Wilson et al. 2001; 2002) and conclude that "the impact of pedagogical knowledge or preparation was spotty and inconclusive" (2003, p. 16). Furthermore, the number of courses offered varies across institutions within single countries in terms of course sequencing and course content (Wilson et al. 2001, p. 12). Given the miscellany of topics from various disciplines (e.g., psychology, educational studies, sociology), it might be difficult for students to develop educational pedagogical knowledge in a sense of a coherent theoretical construct. The perceived fragmentation of educational courses occurs partly due to the high degree of freedom in the choice of educational courses (Terhart et al. 2010).

Furthermore, both teacher students and in-service teachers have often criticized university education and particularly, the educational foundation part, for providing insufficient practical preparation (Cochran-Smith and Zeichner 2005; Darling-Hammond et al. 2002; Veenman 1984; Zeichner 2006) and for the "absence of a set of organizing themes, shared standards, and clear goals"(Hollins 2011, p. 395). In contrast to this argumentation, we argue that – given the high complexity of teachers' actions in rapidly changing situations with high demands on reflection ability (Leinhardt and Greeno 1986) – a thorough conceptual understanding of the domain and of educational topics is a key to improving teachers' professional mastery of their job (Hollins 2011; Rittle-Johnson et al. 2001). To test this hypothesis, we refer to a theoretical framework which can serve as a foundation for the development of a research instrument that allows for objective and reliable assessment of educational knowledge. In line with this argumentation, there is empirical evidence that, despite the general criticism toward educational foundations, students and beginning teachers perceive these parts of university education as relevant for practical work (Alles et al. 2018; Dawson et al. 1984; Grossman and Richert 1988; Rösler et al. 2013).

In sum, although educational knowledge is deemed an important aspect of teachers' professional competence there is little consensus about how this knowledge could best be fostered during teacher education. One reason for this is insufficient empirical evidence on the structure of educational knowledge and its development during teacher education which is mainly due to the fact that to date, most studies have been based on self-report measures only (Choy et al. 2013; Wong et al. 2008). Very few researchers have attempted to measure generic educational knowledge directly via standardized assessments (Guerriero 2017; Sonmark et al. 2017; see also contributions in this volume).

The BilWiss research project started in 2009 to investigate the empirical structure, development, and impact of beginning teachers' educational knowledge. It was one of the first studies to develop a standardized knowledge test covering relevant content of educational foundation courses in academic teacher education. In the next section, we give an overview of the BilWiss research program including important findings.

3 The BilWiss Research Program

The BilWiss research program started in 2009 as a cooperation project between the Max Planck Institute for Human Development (Principal investigator: Jürgen Baumert), the Goethe University Frankfurt (Mareike Kunter), the University of Duisburg-Essen (Detlev Leutner), and the University of Münster (Ewald Terhart), joined in 2012 by the Technical University of Munich (Tina Seidel). The program, consisting of three consecutive funding periods, was supported by the German Federal Ministry of Education and Research (BMBF) over the course of ten years ending in spring 2019. The project team united of researchers who were all involved[3] with teacher education and combined expertise in educational sciences and psychology, the main disciplines in educational foundation courses. In 2012, the project was complemented by an additional study funded by the Ministry of Education in the German state of North-Rhine Westphalia where the study took place. The aim of this study was the evaluation of a newly implemented induction program in this state.

3 Post doc and pre doc researchers involved in the BilWiss project: Andreas Dick, Theresa Dicke, Nora Hein, Olga Kunina-Habenicht, Hendrik Lohse-Bossenz, Christina Maurer, Nadine Schlomske-Bodenstein, Maria Schmidt, Franziska Schulze-Stocker, Kathleen Stürmer, Ziwen Teuber, Katharina Willis, Kristin Wolf.

The BilWiss research program addressed several research questions concerning the nature and meaning of teachers' educational knowledge. Empirically, these research questions were addressed with diverse methodological approaches as outlined below.

1) Conceptualization of educational knowledge: What is the central subject-unspecific content that prospective teachers should know at the end of their university studies? What are the central topics that educational foundational courses should address? What are important areas of knowledge and how can this knowledge be theoretically structured?

The empirical base for answering these research questions was literature work, curriculum analyses, and an expert Delphi study (Section 4.1).

2) Assessment of educational knowledge: Can educational knowledge be measured with a standardized knowledge test and can this test be used to describe knowledge differences between prospective teachers?

As outlined in Section 4.2, we constructed a comprehensive knowledge test that tapped all the topics identified as important by our expert study. After a series of pilot studies this test was administered to a representative sample of German teacher candidates after university completion and was later used in two longitudinal studies. In addition, various smaller validation studies were carried out.

3) Relevance of educational knowledge: To what degree does teachers' educational knowledge influence their later practice? What is the relationship between educational knowledge and teachers' instructional quality, their professional vision, other aspects of professional behavior, and teachers' professional well-being?

To investigate the practical relevance and long-term effect of teachers' educational knowledge on the successful mastery of their job, our research program included a longitudinal study that followed a sample of teacher candidates from completion of their university studies to their entrance into teaching practice and up to seven years beyond. To assess teachers' professional behavior, we used and developed various instruments such as video-based professional vision assessment, student ratings of instruction, vignette tests, and behavioral checklists (for results, see Section 4).

4) Fostering educational knowledge: How does educational knowledge develop during teacher education and beyond? How effective are educational foundation courses in fostering educational knowledge in students? Can tailored interventions support the growth of educational knowledge?

In addition to our first longitudinal study that investigated prospective teachers' development after university completion, our second longitudinal study targeted teacher students at university and investigated how their educational knowledge developed during the course of their university studies (covering a period of two years).

The next section provides a short overview of our core findings. Following the topics of the present volume, we then summarize the findings of the third and final research phase which focused on revision and validation of the BilWiss test.

4 Summary of Important Results from the BilWiss Research Program

4.1 Conceptualization of Educational Knowledge

Before the BilWiss project, the educational foundation courses in German teacher education where a matter of great debate. To tackle the much-debated heterogeneity of content and lack of consistence especially in this part of teacher education, the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (KMK) had passed their "Standards for Teacher Training in the Educational Sciences". These standards specify the abilities and skills prospective teachers should acquire in the course of teacher education, specifically in the educational foundation courses (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) 2014). However, curriculum analyses carried out in BilWiss (Schulze-Stocker 2016; 2017) and by others (Bauer et al. 2012; Hohenstein et al. 2014) showed that even several years later there was still much heterogeneity in terms of the content addressed in educational foundation courses. Further, as seen in these studies, many topics outlined in the Standards were not covered at all, and that universities varied substantially in the courses offered.

In a quest to identify an "ideal curriculum" that would include all the educational topics deemed important for prospective teachers, we carried out an expert survey using a Delphi-technique (Linstone and Turoff 1975). Forty-nine teacher education experts from heterogeneous fields (psychology, educational science, sociology, educators in teacher professional development) participated in a paper-pencil Delphi study in which they rated the importance of 213 potentially relevant educational topics (identified through curriculum analyses and from literature) in three consecutive rounds. These topics covered nine theoretically predefined content areas: instruction, educational theory, educational system, teacher profession, developmental processes, socialization processes, learning processes, dealing with diversity, and assessment. In spite of the heterogeneity found in educational foundation courses offered at university, experts substantially agreed which topics would be the most important for future teachers, leading to a list of 104 topics that were chosen by the majority of experts (for more detailed informa-

tion on specific instructions and consensus development within the Delphi Study, see Kunina-Habenicht et al. 2012; Lohse-Bossenz et al. 2013).

4.2 Assessment of Educational Knowledge: The Development of the BilWiss Test

The Delphi study provided us with a theoretical systematization of the generic educational knowledge that experts agreed should be imparted by university studies. This systematization was the foundation for our test construction. For each of the core Delphi topics we constructed at least one item. The constructed items capture either declarative or conceptual knowledge, with declarative items requiring the mere recall of certain facts or theories, and conceptual items requiring students to connect several theories or to apply certain theories to case examples. Experts from the disciplines of educational science and psychology were involved into the item construction process.

We conducted three pilot studies during the process of test development between June 2010 and November 2010. We optimized or excluded several items based on the examination of item difficulties, item discriminations, and frequencies of incorrect answer alternatives. The pilot studies showed that six dimensions of educational knowledge could be reliably distinguished: classroom teaching, learning and development (subsuming developmental, learning, and socialisation processes), creating school environments, theoretical educational foundations, assessment and evaluation, and teaching as a profession. The final test version consisted of 289 items and was administered in spring 2011 to 3298 persons in one federal German state (North-Rhine Westphalia), representing 87 per cent of the full cohort of recently-graduated teacher candidates at the beginning of their induction phase. To cover the whole breadth of educational knowledge within a reasonable test time, we used a multiple-matric booklet design with anchoring items and estimated person scores for the six dimensions using unidimensional 2-PL partial-credit IRT models. The reliabilities for all six scales were moderate or satisfactory, and the measurement models were replicated in a second independent sample (for further details, see Linninger et al. 2015).

In addition to the comprehensive long test version that distinguishes between six dimensions of educational knowledge, we created a short test version, which includes 57 items and provides a general score for educational knowledge (for an overview of the two test versions with item contents, reliabilities, and examples, see Linninger et al. 2015).

4.3 Relevance of Educational Knowledge: Does It Matter to Teacher Practice?

One of the main guidelines of the BilWiss project was the assumption that a good theoretical knowledge base is a necessary prerequisite for the professional behavior of teachers. The most important arena for teachers is undoubtedly their classroom and the classroom instruction. There is already some evidence that the quality of instruction and students' learning success may be influenced by teachers' pedagogical knowledge which we expected to confirm in our studies. However, following the broad conception of educational knowledge, we were also interested in teachers' professional activities apart from teaching, such as assessment, counselling, and engagement in school development.

Overall, our findings highlight the relevance of educational knowledge to teachers' practice, although our findings were not as clear-cut as expected. Regarding our assumption that educational knowledge works as a theoretical frame that allows for a functional analysis of professional situations, we found that teachers who scored high in our BilWiss test showed significantly more productive reflection when watching teaching scenes by other teachers than those who scored lower (Linninger et al. 2016). However, with regard to professional vision as assessed with the Observer Research Tool (Stürmer and Seidel 2015), no systematic relationships between the overall BilWiss test score and professional vision skills were found. This finding might be attributed to the fact that the Observer Research Tool is quite a focused tool, measuring professional vision skills in the context of three generic teaching and learning principles: clarity of learning goals, teacher support, and learning climate (Seidel et al. 2017). This strong focus might have led to a situation in which the two measurements are not validly linked to find systematic relationships.

Regarding our assumption that educational knowledge should also directly impact on teachers' behavior, we did find that teachers with greater educational knowledge reported greater improvement in teaching quality during their induction phase (Lohse-Bossenz et al. 2015); however, a direct link between teacher knowledge and instructional quality as perceived by students of their classes could not be established.

Going beyond classroom instruction, we also investigated the relevance of educational knowledge for other fields of teachers' work. An important finding was that educational knowledge, as measured by the BilWiss test (especially knowledge about classroom management or learning and development) works as a buffer against stress during the induction phase (Dicke et al. 2015a, b). Moreover, we found that teachers who scored higher in our subscale on knowledge about the

school system and educational policy at the end of the induction phase were more engaged in school development activities two years later than their less knowledgeable peers (Linninger et al. 2015). A pilot study has also shown that theoretical knowledge about counseling as measured within our BilWiss subscale "Teaching as a profession" was positively related to teachers' projected counseling behavior in a situational judgement test (Maurer et al. 2018), a finding which we seek to replicate in ongoing analyses of our longitudinal sample.

4.4 Fostering Educational Knowledge

We found strong evidence that the educational knowledge we assessed in our BilWiss test represents professional academic knowledge actually acquired during teacher education and can be empirically distinguished from everyday notions on education that laypersons may hold. A qualitative study where we conducted cognitive interviews with persons during test taking revealed that teacher education graduates and advanced teacher students were often familiar with item topics from their studies and that they solved items mostly by retrieving academic knowledge gained in teacher education (Linninger et al. 2015). A comparison of the test scores of teacher education graduates with scores of first-semester teacher students or persons without teacher education showed substantial advantages for those who had completed teacher education (Kunina-Habenicht et al. 2013; Linninger et al. 2015). Moreover, we found higher test scores in different dimensions for teacher education graduates who had taken more courses in the respective domain (Schulze-Stocker et al. 2016).

5 Further Development and Validation: The BilWiss 2.0 Test

The first two phases of the BilWiss research program (2009–2016) showed that it is possible to measure prospective teachers' educational knowledge via a standardized test, that this knowledge is actually a product of university teacher education, that beginning teachers differ substantially in their knowledge, and that these knowledge differences manifest in different qualities of behaviors. However, a number of issues remained open. First, although there was evidence of sufficient reliability and validity of the BilWiss test overall, the reliability of some scales was not satisfactory. In addition, item formats and the number of answer options in multiple-choice questions varied across the subscales of the test, which limited the

comparability of items and the scales, so that a further refinement and unification seemed warranted.

A second issue of concern was the lack of predictive validity of the subscale "instruction", which did not show any associations with teaching-related outcome measures. A closer content inspection revealed that some of items in this scale – which all were selected mainly based on empirical grounds – did not show optimal match with the ranking of topics from the Delphi study, so that it seemed necessary to revise the instruction scale substantially.

A third issue was the limited economy of the test. On the one hand, the breadth of educational topics typically addressed in educational foundation courses required a large number of items to adequately cover the heterogeneity of the construct. On the other hand, we were aware of the limited applicability for a test with almost three hundred items and sought to optimize our short test which at that stage did not adequately cover the topics judged most important in our Delphi study.

A fourth critical issue concerned the generalizability of our test and results: As our prior work had been carried out just in one federal state in Germany, it was not clear to what degree our instrument and findings were applicable to other teacher education contexts. Finally, apart from measurement issues, we wanted to learn more about the changes in educational knowledge during the teacher education studies at the university.

The third phase of the BilWiss research program was thus dedicated to further development and validation of the BilWiss test.

5.1 Revision of the Original Test

In the first step, we identified dodgy items using prior data in terms of psychometric indicators and conducted cognitive labs using very similar procedures as described by Linninger and colleagues (2015) to identify items with high task-unspecific variance (e.g., items that could be solved by guessing or ambiguous wording). We rephrased the identified items and created some new tasks. In particular, we substantially modified the instruction scale, which includes many completely new items in the revised version. Moreover, we unified the number of answer options in multiple-choice questions to four in the entire test and ensured that the highly rated topics from the Delphi study were covered by at least one item in the short test form. The modified and new items were tested in an iterative process in several field tests (Kunina-Habenicht et al. 2020).

The short version of the revised BilWiss-2.0 test includes 65 items, the long version contains 119 items, all from the following six knowledge domains: learning and development, instruction, assessment, educational theory (and history), school system and educational policy, and teacher profession. 2-PL partial-credit IRT models were applied to a data set collected from 788 teacher students from four different German universities in four different states (for details, see Kunina-Habenicht et al. 2020).

5.2 Content Validity

All items of the test were examined in a qualitative study where 40 teacher students worked through the test (on average five teacher students per item) and verbalized their solution approaches, difficulties, and sources of knowledge. These cognitive labs showed that the majority of items were congruent with the topics addressed in educational foundation courses at university and that they could be answered by drawing on the content learned during university teacher education. The interviews also gave hints to construct-irrelevant variance inherent in some items that were considered during the item revision described above.

In two additional validation studies, we investigated whether the test content corresponded to the intended academic teacher education curriculum at the federal level in Germany (Kunina-Habenicht et al. 2019). In the first study delegates from the Ministries of Education of most German federal states (besides North-Rhine Westfalia) rated the relevance of the Delphi topics for their specific state. In the second study the test content (i.e. individual test items) was matched to the federal standards on the generic, educational part of academic teacher education (KMK, 2014). Results from both studies indicated that the BilWiss-2.0-test can be used across German federal states (Kunina-Habenicht et al. 2019).

5.3 Internal Structure

With regard to the empirical structure of the BilWiss 2–0 test, structural equation models indicated a good fit for the model with six correlated latent factors (for details, see Kunina-Habenicht et al. 2020). Moreover, we could show that educational knowledge can be invariantly measured across three subject groups, i.e., science, languages/humanities, and a combination of these subjects (Lohse-Bossenz et al. 2018).

5.4 Relation with Other Variables

Small significant correlations between the BilWiss 2.0 test score and the number of relevant university courses attended and grades in university studies support the convergent and prognostic validity of the test score interpretations. Moreover, teacher students who had to re-sit at least one exam in their educational foundation courses showed significantly lower test performance than students who passed on the first try (Kunina-Habenicht et al. 2020). Moreover, we found that university students with a range of individual risk factors such as lower cognitive abilities, lower SES, immigration background or unfavorable personality traits (e.g., high neuroticism scores) showed significantly lower test scores in later semesters than students without these risk factors (Wolf 2019).

6 Theoretical and Practical Implications

The BilWiss research program aimed to investigate the emergence and relevance of theoretical educational knowledge as an important facet of (prospective) teachers' professional competence. Over the period of almost ten years, we have succeeded to constructing and validating a psychometrically solid standardized test to assess generic educational knowledge in a generic, broad sense. A distinctive feature of the BilWiss test is that it not only includes those teaching-related knowledge areas that are typically targeted in existing concepts of pedagogical knowledge such as instruction, assessment, learning and development (Voss et al. 2015). The BilWiss test also covers topics that go beyond classroom teaching and touches other fields of teacher activity, such as school development, counseling, and an understanding of teaching as a professional occupation.

We found that teacher students and graduates of teacher education differ substantially in this form of knowledge which is at least partially due to a differing uptake of learning opportunities during teacher education, and that these differences in knowledge are associated with differences in professional success.

6.1 Implications for Future Research

After the comprehensive revision and validation of the first BilWiss test, the revised BilWiss 2.0 test will now be made available for other researchers[4]. It consists of a long version with 14–24 items per scale, providing a comprehensive assessment of most topics teacher education experts consider essential for prospective teachers, and a short version with 65 items that can be used as a condensed indicator for teachers' generic educational knowledge. We recommend the use of the long version in situations where one is interested to what degree an intended teacher education curriculum is realized or where one wants to assess selected dimensions of educational knowledge in more depth. Typically, this would be studies within the context of educational monitoring or evaluations of specific courses or treatments within the context of teacher education. We recommend the use of the short version in situations where one needs just a general individual score of educational knowledge as a measure of an important aspect of teachers' individual professional competence, figuring as a predictor, outcome or control variable.

The BilWiss 2.0 test can be applied for further research on the effectiveness and long-term impact of educational foundations courses. It is our conviction that this part of teacher education is better than its reputation and we expect other studies to confirm the high relevance of a sound theoretical base for successful mastery of the teaching job (Hollins 2011; Patrick et al. 2011).

We also hope that our test is a useful tool for studies investigating the development of educational knowledge through teacher education or specifically tailored interventions. Although there is ample research on effects of certain formats in teacher education, many of these studies suffer from a shortage of convincing objective measures (Cochran-Smith and Zeichner 2005). Moreover, the mechanisms underlying the development of teachers' professional knowledge are not well understood yet. For instance, it is not clear whether consecutive approaches in teacher education are more effective than concurrent approaches or vice versa (Harr et al. 2015). Another important open issue refers to the question of how theoretical knowledge can transfer into better practice. While it is unlikely that theoretical knowledge itself directly transfers into better teaching (or other professional) behaviors, it can be assumed that it plays an important role in the interpretation of teaching (or other professional) situations (Blömeke et al. 2015; König et al. 2014).

4 The BilWiss 2.0 test (in German language) will be made available at the Research Data Centre Education (FDZ Bildung) at the Leibniz Institute for Research and Information on Education (DIPF), the BilWiss data is available at the Research Data Centre (FDZ) at the Institute for Educational Quality Improvement (IQB).

Thus, the effects are not necessarily direct, but more complex and can only be recovered using moderation and/or mediation analysis (e.g., Dicke et al. 2015).

6.2 Implications for Practice

The general message from BilWiss research program is that the theoretical generic educational knowledge is one important condition for successful mastery of practice. Our findings show that gap between theory and practice may not be as wide as often purported: Not only do academic and practice experts agree which theoretical issues should be covered in initial teacher education. Moreover, our findings give first evidence that a good theoretical foundation can ease beginning teachers' entry into practice. Which forms of teacher education may be best suited to gain this type of knowledge remains an open question. With the BilWiss test, there is now a tool available with which new developments in teacher education may be empirically validated. The test is available for studies aiming at monitoring and evaluation of specific teacher programs. One essential issue for school policy and school administration is, for example, whether there is a difference between teachers who finished regularly certified teacher education programs, and lateral entry employees, who enter school after attending only reduced or even none teacher preparation programs and possess only limited knowledge about theoretical foundations. Although the BilWiss test was developed in German, we are convinced that its content can be at least partly adapted to other languages and countries.

References

Abell, S. K. (2008). Twenty Years Later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education, 30*(10), pp. 1405–1416. doi:10.1080/09500690802187041

Alles, M., Apel, J., Seidel, T., & Stürmer, K. (2018). Candidate Teachers Experience Coherence in University Education and Teacher Induction: the Influence of Perceived Professional Preparation at University and Support during Teacher Induction. *Vocations and Learning, 6*(2). doi:10.1007/s12186–018-9211–5

Bauer, J., Diercks, U., Rösler, L., Möller, J., & Prenzel, M. (2012). Lehramtsausbildung in Deutschland: Wie groß ist die strukturelle Vielfalt? [Teacher Education in Germany: How Big is the Structual Variety?]. *Unterrichtswissenschaft, 40*(2), pp. 101–120.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, (pp. 133–180). doi:10.3102/0002831209345157