KoKoHs

Modeling and Measuring Competencies in Higher Education

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

KoKoHs Working Papers

No. 8

Corinna Lautenbach & Katharina Schulz (Eds.)

Developing International Research Projects in the Field of Academic Competency Assessment

Report from the "KoKoHs-Autumn Academy" from October 6-10, 2014 in Berlin

SPONSORED BY THE

Federal Ministry of Education and Research

Johannes Gutenberg University Mainz

Humboldt University of Berlin

***KoKoHs Working Papers* from the BMBF-funded research initiative „Modeling and Measuring Competencies in Higher Education“**

The *KoKoHs Working Papers* series publishes articles from the funding initiative „Modeling and Measuring Competencies in Higher Education (KoKoHs)“. These may be conceptual papers or preliminary results intended for rapid dissemination or public discussion. Publication in *KoKoHs Working Papers* does not preclude publication of the texts elsewhere. The responsibility for the content lies with the authors. The content does not necessarily reflect the views of the publishers of *KoKoHs Working Papers*.

Contact:

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

The research initiative is funded by the German Federal Ministry of Education and Research under grant no. 01PK11100A and 01PK11100B.

The *KoKoHs Working Papers* are also available for download:

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

Developing International Research Projects in the Field of Academic Competency Assessment

Report from the "KoKoHs-Autumn Academy" from October 6-10, 2014 in Berlin

Lautenbach, C. & Schulz, K. (Eds.)

Contact:

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

Bibliographical references:

Lautenbach, C., Schulz, K. (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.

Developing International Research Projects in the Field of Academic Competency Assessment - Report from the "KoKoHs-Autumn Academy" from October 6-10, 2014 in Berlin

Abstract:

A main task within the research program "Modeling and Measuring Competencies in Higher Education (KoKoHs)", which is funded by the Federal Ministry of Education and Research (BMBF), is the systematic and sustainable advancement of young researchers. To meet this challenge as well as to gain higher international visibility and to maintain and enhance existing international cooperations, the KoKoHs coordination office has organized the "KoKoHs-Autumn Academy" from October 6-10, 2014 in Berlin. The present working paper documents insights into the workshops and the research projects developed by the young researchers as well as impressions and reflections on the conference.

Keywords:

Competence assessment, higher education, advancement of young researchers, Autumn Academy

Contents

1 Introduction

1.1 About KoKoHs

Over the past decade, there has been growing interest in various issues related to the provision of higher education. Policy-driven outcome-oriented reform strategies (e.g., the Bologna Reform) have changed the higher education sector in sustainable ways. Findings from empirical research on the effectiveness of higher education programs can serve as a basis for development and reforms on a structural, organizational, and individual level.

An analysis of the international state of research in higher education by Kuhn & Troitschanskaia (2010) provided a sound basis for the conceptualization and implementation of the German research program Modeling and Measuring Competencies in Higher Education (KoKoHs). The purpose of this research program funded by the German Ministry of Education and Research (BMBF) is to carry out fundamental, systematic, and internationally compatible research on competency development and assessment in higher education in Germany. Thus, its aims can be described as follows:

- modeling domain-specific and generic competencies in selected subjects while taking into account specific curricular and job-related features,
- transforming theoretical models into suitable measuring instruments,
- validating test score interpretations.

From 2011 to 2015, the program included 24 cross-university project alliances encompassing 70 projects with 220 researchers at more than 50 institutions of higher education in Germany and Austria. Selected in an external review process, each project alliance brought together domain experts, teaching methodology experts, and research methodology experts from at least two universities. The first research phase ran from 2011 to 2015, and, after positive external evaluations, the program will continue for another five years (from 2016 to 2020).

The general purpose of the KoKoHs program is to systematically model and assess domain-specific and generic competencies of students in HE[1]. KoKoHs projects take into account curricular and job-related requirements, transform theoretical competency models into suitable measuring instruments, and validate test score interpretations. To enable meaningful cooperation and promote cross-

[1] For more information on the research program see: http://www.kompetenzen-im-hochschulsektor.de/index_ENG.php

project synergies during the first phase, KoKoHs focused on student competencies in one generic cluster (self-regulation and general research competencies) and four domain-specific clusters comprising some of the most popular fields of study in Germany:

- engineering, including electrical engineering and mechanical engineering;
- economics and social sciences, including teacher training in economics and social sciences;
- educational sciences, including psychology; and
- teacher training in science, technology, engineering, and mathematics (STEM subjects).

1.2 Previous Activities for Young Researchers in the KoKoHs Program

A main objective within the KoKoHs program is the systematic and sustainable advancement of young researchers to help to build up capacities for research in higher education. To meet this challenge, as well as to gain international visibility, several activities to promote young researchers and their PhD projects have been organized throughout the years. These activities included workshops, possibilities for networking and exchange as well as internationalization events. In three workshops in 2012 and 2013, young researchers learned about important tools and methods for their research projects. In the context of networking and exchange, a Kick-Off Conference and a Roundtable Discussion on validation and validity were organized in 2012 and 2013. As one of the main purposes of the KoKoHs program is to bring together young researchers from Germany and international researchers from the field of competency assessment, young researchers were invited to attend an International Colloquium in 2013. At the colloquium, young researchers from the KoKoHs projects had the chance to present their planned doctoral and post-doctoral projects to well-renowned experts of different fields (statistics, psychology and empirical education research) as well as to other young researchers. This also provided them with the opportunity to discuss their projects and receive feedback.

1.3 KoKoHs Autumn Academy - Issue

The KoKoHs Autumn Academy, which took place in October 2014 in Berlin, was another cornerstone in the process of strengthening international collaboration and supporting international research projects in the field of assessing academic learning outcomes. At the sessions, young researchers were taught how to cooperate and collaborate in international research. The workshops covered examples, challenges and advantages of joint international research projects and papers. Moreover, the Autumn Academy provided an opportunity for outstanding junior researchers from the KoKoHs projects to meet selected international junior researchers specializing in research on the assessment

of higher-education competencies. The participating groups of young researchers, comprised of international as well as of German researchers, developed first drafts for their joint projects in advance of the conference. A group of international experts reviewed the drafts and selected the high-quality projects which were then admitted to the Autumn Academy.

1.4 KoKoHs Autumn Academy - Program

The KoKoHs Autumn Academy took place from October 6th to 10th, 2014, at the Humboldt University of Berlin. The first two days of the Autumn Academy included a two-day training on how to cooperate and collaborate in international research and on the challenges of competence measurement in higher education. These sessions were open to all KoKoHs researchers as they included topics relevant for all of the participants. Edward Wiley from the University of Colorado Boulder (USA) opened the conference with his keynote on "Challenges of international research on competencies". He was followed by Alicia Alonzo from the Michigan State University (USA) and her keynote on "How to write an international paper/project proposal - from a reviewer's perspective". Following the keynotes, extensive plenary discussions were held. The second day of the Autumn Academy started with presentations on international research projects on competencies held by Samuel Rikoon and Katrina Roohr, both from the Educational Testing Service (ETS) in New Jersey (USA). While Samuel Rikoon's presentation focused on the ETS Work Readiness Strength Assessment and Training System (WRSATS), Katrina Roohr focused on the efforts of developing next-generation student learning outcomes assessments for higher education institutions in the United States. The day ended with a workshop on "Theories and concepts of research on competencies" chaired by Alicia Alonzo and Edward Wiley.

The next three days (October 8th to 10th) of the Autumn Academy were dedicated to the members of the four research groups and consisted of workshop phases and presentations on the groups' work in progress. The research groups worked on their individual projects and where advised by international experts and peer counselors. On the final day of the Autumn Academy, all groups of researchers presented their projects and further plans.

1.5 KoKoHs Autumn Academy – Participants

1.5.1 The Project Groups

Competency-Based Model of Educator Development for Culturally and Linguistically Responsive Teachers: The Beginnings

Kara Mitchell Viesca, University of Colorado Denver, USA

Svenja Hammer, Leuphana University of Lüneburg, Germany

Modeling and Measuring the Competence of Teachers to Teach Computational Thinking

Aman Yadav, Michigan State University, USA

Jonathan Good, Michigan State University, USA

Leigh Graves Wolf, Michigan State University, USA

Marc Berges, TUM School of Education, Germany

Melanie Margaritis, University of Paderborn, Germany

Evidence-based Reasoning in Higher Education. Development and Validation of a Problem-oriented Vignettes Test to Access Educational Research Literacy

Christina Haberfellner, BIFIE Salzburg, Austria

Jana Groß Ophoff, University of Education Freiburg, Germany

Raffaela Wolf, Council for Aid to Education, USA

Professional Competencies of Prospective Kindergarten Teachers within Mathematics in Germany, Switzerland and Norway

Anne Nakken, Queen Maud's University College, Norway

Lars Jenßen, Humboldt University of Berlin, Germany

Michael Link, University of Education St. Gallen, Switzerland

Simone Dunekacke, Humboldt University of Berlin, Germany).

1.5.2 The International Experts and Peer Councellors

Alicia C. Alonzo

Institution: Michigan State University (USA)

Position: Associate Professor

Expertise: science education, physics education, teacher knowledge, formative assessment, learning progressions, measurement

Edward W. Wiley

Institution: National Education Policy Center, School of Education, University of Colorado Boulder (USA)

Position: Director of Big Data Analytics, Seagate Technology

Expertise: Big Data and advanced statistical analytics, systems of school accountability, teacher quality and compensation, school choice

Katrina C. Roohr

Institution: Educational Testing Service (USA)

Position: Associate Research Scientist

Samuel Rikoon

Institution: Educational Testing Service (USA)

Position: Ph.D. student in the Quantitative Methods program at the University of Pennsylvania's Graduate School of Education

2 Presentations

2.1 How to Write an International Paper/Project Proposal – A Reviewer's Perspective | Alicia Alonzo

There are several points you should pay attention to when you write an international paper or project proposal. Below you find a list of tips that can be helpful and which you should consider before, while and after your writing work.

Tip 1

- Don't wait until the end of the project to discuss authorship of project-related papers.
- Consider establishing "rules" to determine whether a project member is included on a conference paper, on a manuscript to be submitted to a journal, etc.
- Consider mapping out a set of papers that might result from the work and who will take responsibility for (and, thus, first authorship of) each.
- Have ongoing conversations about what makes sense, given each project member's contributions to a particular paper.

Tip 2

- Make the reviewer's job as easy as possible.
- Help the reviewer see how you are addressing the review criteria.
- If there are specific review criteria, consider using those as section headings to help the reviewer know where to find the information you want them to evaluate.
- Gaining access to the review criteria (especially by serving as a reviewer) can be very helpful for understanding the criteria.
- Provide sufficient information about your work.
- Particularly for proposals, if you are not using the whole page limit, you probably have not provided enough detail.
- Make the text as easy to read.
- Use a font that is large enough and clear to read.
- Take the time to learn and apply appropriate style guidelines.
- Leave yourself time for careful editing.
- US reviewers are used to seeing larger demarcation of the beginning/end of paragraphs (e.g., indenting the first line of a new paragraph of extra space between paragraphs).

Tip 3

- Answer the "so what" question.

- Why should someone care about the research you are doing?
- This should not just be a general argument for the importance of the topic, but a specific argument for the study that you have undertaken (or will undertake).
- It is not enough to say that there has not been work done in this particular area… Why should work be done in that area?
- How does your work both build on and contribute to existing work in the field?
- Make specific claims (i.e., do not just claim that your study has implications for a certain field but describe what those implications might be).
- Situate your work with respect to the field broadly (i.e., do not cite only your own work or only the work that you directly draw upon).
- Be clear about what your work adds to what is already known.

Tip 4

- Pay attention to alignment of all of the pieces of the paper/proposal.
- Purpose of the work:
- Problem you hope to address/answer to the "so what" question.
- Connections to existing literature.
- Research questions
- Study design:
- Including both data collection and analysis.
- (Anticipated) results
- (Anticipated) conclusions
- (Anticipated) implications
- If the work is part of a larger project, the focus should be on the specific work being reported/ proposed.

Tip 5

- Be aware that things that seem clear to you may not be as clear to a reviewer.
- Explain any contextual factors that may be unfamiliar.
- Consider the audience and provide extra explanation about jargon, etc. that may be common knowledge in one scholarly community but unfamiliar to another.
- Be as clear as possible about what you will actually do (in a proposal) or what you actually did (in a paper).

Tip 6

- Pay attention to theoretical clarity and psychometric quality.

- Clearly define any constructs (and explain how your definitions are similar to/different from other definitions in the field).
- Examples of items, coding schemes, or other means of data collection/analysis can help to illustrate what you mean by a particular construct.
- It should be very clear that the data collection/analysis you are proposing/have used is well-aligned with the construct(s) of interest.
- Be sure to discuss the technical quality (i.e., evidence of validity and reliability) of the instruments.

Tip 7

- For papers, the discussion/conclusion section should go beyond a summary of the results.
- Don't leave the work of making sense of the results to the reader. Clearly explain the results in terms of:
- Why they are important (the "so what")?
- How they relate to other work in the field?
- How they contribute to the field?
- For papers, be careful about overstating claims or implications based on your results. Be especially careful about making claims about causality.
- Consider practical (as well as statistical) significance.

Tip 8

- For proposals, demonstrating that you have the capacity to carry out the work is important.
- Demonstrate a "track record" of prior work.
- Include evidence of pilot work.
- Include evidence of capacity to seek advice from appropriate experts (e.g., advisory board members).
- For collaborative proposals, be clear about what expertise each person adds to the project and exactly who is responsible for what aspects of the work.

Tip 9

- Be persistent! Many manuscripts/proposals will not be accepted when first submitted.
- Carefully consider all of the feedback that you receive (even if it initially seems unhelpful).
- When resubmitting a manuscript, make it very clear how you have responded to each of the editor's comments (and as many of the reviewers' comments as possible).
- You don't have to follow every suggestion, but you should have a clear argument for any advice you have chosen to disregard.

2.2 Developing Next-Generation Student Learning Outcomes Assessments for Higher Education Institutions in the United States | Katrina Crotts Roohr and Ou Lydia Liu

Educational Testing Service

Influences and pressures from statewide governing boards, state mandates, regional and program accreditors, and a greater drive for accountability have resulted in an increase in the measurement of student learning outcomes (SLOs) across United States colleges and universities (Kuh, Jankowski, Ikenberry, & Kinzie, 2014; Toiv, 2013; Richman & Ariovich, 2013). SLOs are important skills, attitudes, or competencies (e.g., critical thinking skills, mathematics, and writing skills) that students are expected to acquire at higher education institutions (National Institute for Learning Outcomes Assessment [NILOA], 2011). SLO measures are typically used to satisfy accreditation and accountability requirements, conduct trend analysis, compare students' achievement levels across institutions, and offer feedback to students regarding their academic success (ACT, Inc., 2014; Council for Aid to Education, 2014; Educational Testing Service [ETS], 2014). SLO measures can also provide evidence of both student learning and institutional effectiveness, and provide feedback about the strengths and weaknesses of academic programs.

Measures of SLOs evaluate whether students are actually acquiring the skills important for life, work, and citizenship when attending higher education institutions. There are a variety of direct and indirect measures used to assess SLOs such as national and locally developed surveys, standardized measures, rubrics, performance assessments, and e-portfolios (Kuh et al., 2014). Each of these assessment tools has distinct advantages and disadvantages which can be found in Table 1.

Assessment Tool	Advantages	Disadvantages
National survey	Cost efficient; easy administration	No direct evidence of student learning
Locally developed survey	Aligned with instruction; meet institution's specific needs	No benchmark with other institutions; sometimes lack psychometric quality
Standardized measures	Comparable across institutions; sufficient validity and reliability evidence	Insufficient alignment with instruction
Rubrics	Flexibility for adaptation	Poor consistency among users
Performance assessment	Authentic	Expensive; difficult to implement; poor reliability
e-portfolio	Offer a range of data	Comparability is an issue

Table 1. Advantages and Disadvantages of Current SLO Assessment Tools

A distinct advantage of standardized measures that stands out among the various assessment tools is the ability to compare valid and reliable test scores across institutions. The disadvantage of these standardized measures, however, is their insufficient alignment with instruction. This disadvantage calls for the development of a next-generation standardized assessment of SLOs. A next-generation assessment should be aligned with instruction, involving faculty from multiple institutions in assessment development to ensure alignment across inter-institution curricula, and provide higher education institutions with the ability to customize the assessment to meet their institutional needs. These assessments should balance authenticity and psychometric quality, use multiple assessment formats, and consider the diversity of the higher education population (e.g., all students including those with disabilities and language learners).

An Approach to Developing Next-Generation SLO Assessment

When thinking about next-generation SLO assessment, it is critical to first consider the needs of higher education institutions and identify which SLOs are important. The Association of American Colleges and Universities (AAC&U) and the Lumina Foundation have both made efforts to identify important SLOs for higher education institutions. The AAC&U launched the Liberal Education and American's Promise (LEAP; AAC&U, 2011) initiative and the Lumina Foundation developed the Degree Qualifications Profile (DQP; Adelman, Ewell, Gaston, & Schneider, 2011, 2014). The workforce has also identified important skills students should have upon graduating from a higher education institution that overlap considerably with the skills identified by the AAC&U and Lumina (see Casner-Lotto & Barrington, 2006; Hart Research Associates, 2013). Markle and colleagues (2013) researched these existing frameworks across the higher education and workforce communities to determine their level of overlap. Markle et al.'s paper, along with quantitative and qualitative market research (as well as input from higher education institutions) resulted in the identification of seven core competencies for higher education: (a) critical thinking, (b) written communication, (c) quantitative literacy, (d) digital information literacy, (e) civic competency and engagement, (f) intercultural competency and diversity, and (g) oral communication.

It is critical that each of these seven competencies be clearly defined by an operational definition to inform the development of a next-generation SLO assessment. Development of a next-generation assessment should follow the evidence-centered design framework (see Mislevy, Almond, & Lukas, 2003) or other appropriate methods for constructing assessments (e.g., Luecht, 2013) and should leverage existing literature. Developing an operational framework with assessment considerations can be useful for institutions developing in-house assessments, providing key dimensions and defini-

tions of SLOs that can easily translated into an assessment. In the next section, we provide the steps to developing a proposed framework for quantitative literacy, one of the seven core competencies.

In-Depth Discussion of Framework Development – Quantitative Literacy

To develop an operational definition of a competency, such as quantitative literacy, a considerable amount of research is needed. For a detailed review of the framework for quantitative literacy, see Roohr, Graf, and Liu (2014).

The first step in developing an operational definition of a competency is to recognize the synonymous terminology for your construct of interest. For quantitative literacy, terms such as quantitative reasoning, numeracy, mathematical literacy, and even mathematics have all been used synonymously. In some instances the definitions for these terms have considerable overlap. It is critical to recognize both the commonalities and differences between these terms to develop a concrete operational definition. Using these various terms, existing frameworks, definitions, and assessments were closely reviewed and synthesized. Existing frameworks included national and international organizations such as the AAC&U's LEAP (AAC&U, 2011), Lumina's DQP (Adelman et al., 2014), American Mathematical Association of Two-Year Colleges (AMATYC; Cohen, 1995), Mathematical Association of America (MAA; Steen, 2004), and the Organisation for Economic Co-Operation and Development (OECD; PIAAC Numeracy Expert Group, 2009). Workforce initiative definitions such as the Employment and Training Administration's (ETA) Industry Competency Model (U.S. Department of Labor, 2013), higher education institutions and researchers, and definitions from K-12 theorists and practitioners (e.g., the Common Core State Standards, Partnership for 21st Century Skills) were also reviewed. In reviewing these existing frameworks and definitions, key terms and sub-dimensions were identified. Specifically, we identified the most common themes across existing frameworks and definitions.

A similar approach was also taken when reviewing existing assessments such as the Collegiate Assessment of Academic Proficiency (CAAP) Mathematics, Collegiate Learning Assessment+ (CLA+) Scientific and Quantitative Reasoning, ETS Proficiency Profile (EPP) Mathematics, and others. Across assessments we examined the content assessed, contexts of test questions, item formats, and reliability and validity evidence, including fairness. This review helped to identify important issues when assessing quantitative literacy such as understanding the difference between assessing mathematics and quantitative literacy (e.g., Steen, 2001), creating a general assessment versus a domain-specific assessment (e.g., Ewell, 2001), recognizing the psychometric challenges of reporting multiple subscores (e.g., Sinharay, Puhan, & Haberman, 2011), and motivating students to take a low-stakes SLO assessment (e.g., Liu, Bridgeman, & Adler, 2012).

Based on the above review, we were able to create an operation definition and theoretical framework of quantitative literacy. This framework includes five mathematical problem-solving skills (interpretation, strategic knowledge and reasoning, modeling, computation, and communication), four mathematical content areas (number and operations, algebra, geometry and measurement, probability and statistics), and three real-world contexts (personal/everyday life, workplace, society). This proposed framework is intended to inform the development of an assessment of quantitative literacy that is intended for all college students, regardless of their major, to be mainly used for college accreditation, instructional improvement, and assessing individual students' proficiency levels (see Roohr et al., 2014 for more information).

The proposed assessment structure (based on the operational definition and theoretical framework) will be computer-administered and take 45 minutes or less, enabling the assessment to be completed during a single college class session. Assessment items will cover a primary problem-solving skill and content in a variety of real-world contexts. An on-screen, four-function calculator will be provided to reduce the computational load on students and allow them to focus on the problem-solving skills. Possible item formats include selected response (e.g., single- and multiple-selection multiple-choice, drop-down menu, hot spot, table grid) or open-ended items (e.g., numeric entry, fraction entry, create/edit a graph/table, short constructed response). In addition to item formats, a number of tasks can be used throughout the assessment to measure dimensions in the framework such as data accuracy, data sufficiency, draw conclusions, evidence, quantitative comparison, recognize inconsistency, and representational equivalence.

When thinking about assessment structure, it is also important to consider potential sources of construct-irrelevant variance. An assessment should be developed with all students in mind (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014) meaning the assessment should be accessible to all students, including students with disabilities and second language learners. This means careful consideration when thinking about delivery modes. Technology-enhanced items should have clear directions and not be over used. Additionally, with a computer-administered assessment we need to consider the potential barrier, and thus increased complexity, of completing quantitative items on a computer versus on paper. Lastly, with a quantitative assessment, we also need to consider the cognitive reading load. An assessment should measure quantitative skills, not reading ability.

A framework paper plays an important role in the development of a next-generation assessment of SLOs. This paper can provide a working operational definition to move forward with assessment development. Assessment development involves "coordination between content experts, assessment developers, measurement experts, and potential score users at higher education institutions" (Roohr

et al., 2014, p. 23). Before creating an operational assessment, a pilot or field test will be conducted to ensure that the assessment is psychometrically sound. Validation studies will also be completed to ensure that the assessment is measuring what it is intended to measure.

Important Considerations

Developing an operational definition for an SLO in higher education involves extensive research and coordination efforts. Similar approaches should be taken when thinking about other competencies, such as recognizing the various challenges. Although there are some challenges that may be similar across competencies (e.g., general versus domain specific, student motivation, etc.), there are unique challenges that should also be identified. For instance, a unique challenge for the oral communication competency is measuring someone's oral communication behavior (i.e., the act of giving a presentation or holding a conversation) in a standardized way. One proposed method to resolve this challenge is to video record students giving a presentation on the same task. With this proposed resolution comes other challenges such as scoring considerations (e.g., human scoring of videos), scorer bias, camera quality, etc. Feng et al. (under preparation) has proposed an alternative method to human scoring involving the use of transcribed speeches and watching clips of the video presentation rather than the full presentation, that has the potential to increase feasibility and lower costs. Recognizing the challenges of measuring SLOs might mean more research is needed, but is a critical step in developing operational definitions and proposing assessment considerations.

Conclusion

In summary, developing an operational definition should be the first step in building a next-generation SLO assessment. This task is a critical step in determining the appropriateness and feasibility of measuring the construct of interest. The targeted population (e.g., all college students attending a higher education institution) will play an important role in narrowing such an operational definition. For more examples of SLO frameworks for various competencies in higher education, see Liu, Frankel, and Roohr (2014), which discusses developing a framework for critical thinking, or Sparks, Song, Brantley, and Liu (2014) on developing a framework for written communication.

References

ACT, Inc. (2014). Collegiate assessment of academic proficiency. Retrieved from http://www.act.org/caap/index.html.

Adelman, C., Ewell, P., Gaston, P. & Schneider, C. G. (2011). The Degree Qualifications Profile. Defining degrees: A new direction for American higher education to be tested and developed in

partnership with faculty, students, leaders, and stakeholders. Indianapolis, IN: Lumina Foundation for Education.

Adelman, C., Ewell, P., Gaston, P. & Schneider, C. G. (2014).The Degree Qualifications Profile: a learning-centered framework for what college graduates should know and are able to do to earn the associate, bachelor's or master's degree. Indianapolis, IN: Lumina Foundation for Education.

Association of American Colleges and Universities (2011). The LEAP vision for learning: Outcomes, practices, impact, and employers' view. Washington, DC: Author.

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

Casner-Lotto, J. & Barrington, L. (2006). Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce. New York, NY: Conference Board, Partnership for 21st Century Skills, Corporate Voices for Working Families, Society for Human Resource Management.

Cohen, D. (1995). Crossroads in mathematics: Standards for introductory college mathematics before calculus. Memphis, TN: American Mathematical Association of Two-Year Colleges.

Council for Aid to Education (2014). CLA+ overview. Retrieved from http://cae.org/performance-assessment/category/cla-overview/.

Educational Testing Service (2014). ETS Proficiency profile. Retrieved from http://www.ets.org/proficiencyprofile/about.

Ewell, P. T. (2001). Numeracy, mathematics, and general education. In L. A. Steen (Ed.), Mathematics and democracy: The case for quantitative literacy (pp. 37–48). Princeton, NJ: Woodrow Wilson National Fellowship Foundation.

Feng, G., Joe, J., Kitchen, C., Mao, L., Roohr, K. & Chen, Lei (under preparation). A proof-of-concept study on scoring oral presentations videos. Manuscript under preparation.

Hart Research Associates (2013). It takes more than a major: Employer priorities for college learning and student success. Washington, DC: Author.

Kuh, G. D., Jankowski, N., Ikenberry, S. O. & Kinzie, J. (2014). Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning

Outcomes Assessment. Retrieved from http://www.learningoutcomeassessment.org/documents/2013%20Abridged%20Survey%20Report%20Final.pdf.

Liu, O. L., Bridgeman, B. & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. Educational Researcher, 41(9), 352–362.

Liu, O. L., Frankel, L. & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment (ETS RR-14-10). Princeton, NJ: Educational Testing Service.

Luecht, R. M. (2013). Assessment engineering task model maps, task models and templates as a new way to develop and implement test specifications. Journal of Applied Testing Technology, 14(1).

Markle, R., Brenneman, M., Jackson, T., Burrus, J. & Robbins, S. (2013). Synthesizing frameworks of higher education student learning outcomes (Research Report No. RR-13-22). Princeton, NJ: Educational Testing Service.

Mislevy, R. J., Almond, R. G. & Lukas, J. F. (2003). A brief introduction to evidence-centered design (RR-03-16). Princeton, NJ: Educational Testing Service.

National Institute for Learning Outcomes Assessment (2011). Transparency framework. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from http://www.learningoutcomessassessment.org/TransparencyFrameworkIntro.html.

PIAAC Numeracy Expert Group (2009). PIAAC numeracy: A conceptual framework (OECD Education Working Paper No. 35). Paris, France: OECD.

Richman, W. A. & Ariovich, L. (2013). All-in-one: Combining grading, course, program, and general education outcomes assessment (Occasional Paper #19). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Roohr, K. C., Graf, E. A. & Liu, O. L. (2014). Assessing quantitative literacy in higher education: An overview of existing research and assessments with recommendations for next-generation assessment (ETS Research Report No. RR-14-22). Princeton, NJ: Educational Testing Service.

Sinharay, S., Puhan, G. & Haberman, S. J. (2011). An NCME instructional module on subscores. Educational Measurement: Issues and Practice, 30(3), 29–40.

Sparks, J. R., Song, Y., Brantley, W., & Liu, O. L. (2014). Assessing written communication in higher education: Review and recommendations for next-generation assessment (ETS RR-14-37). Princeton, NJ: Educational Testing Service.

Steen, L. A. (2001). The case for quantitative literacy. In L. A. Steen (Ed.), Mathematics and democracy (pp. 1–22). Princeton, NJ: Woodrow Wilson National Fellowship Foundation.

Steen, L. A. (2004). Achieving quantitative literacy: An urgent challenge for higher education. Washington, DC: Mathematical Association of America.

Toiv, B. (2013). Top research university expanding efforts to assess, improve undergraduate student learning. Washington, DC: Association of American Universities. Retrieved from http://www.aau.edu/WorkArea/DownloadAsset.aspx?id=14849.

U.S. Department of Labor (2013). Competency model clearinghouse: Mathematics. Retrieved from http://www.careeronestop.org/competencymodel/blockModel.aspx?tier_id=2&block_id=9.

2.3 The ETS Work Readiness Strength Assessment & Training System | Samuel Rikoon

Educational Testing Service

Note: The below represents only a broad overview. For further detail on the development and research underlying the ETS Work Readiness Strength Assessment and Training System (WRSATS), please see the associated ETS Technical Research Report (currently under review for publication in early 2015):

Shore, J., Noeth, R., Inglese, P., Lentini, J., Naemi, B., Rikoon, S. & Robbins, S. (2014). The ETS Work Readiness Strength Assessment & Training System: Technical research report. Manuscript submitted for publication.

With increased globalization, the American workforce has evolved to find employers placing a premium on educated, highly-skilled workers. In 1973, workers with education and training beyond a high school diploma held only 28% of the jobs, whereas in 2010 they held 59% of the jobs, and will hold nearly two-thirds of the jobs by 2020 (Carnevale, Smith, & Strohl, 2013). The largest component of these jobs (over 50%) is represented by positions that are described as middle skills jobs (Carnevale et al., 2013). This type of job requires more than a high school education but less than a bachelor's degree (e.g., associate's degree, postsecondary certificate, apprenticeship, formal on-the-job training). Middle skills jobs are differentiated from high skills jobs (typically filled by individuals who have bachelor's degrees or higher) and low skills jobs (generally filled by those with a high school diploma

or less). In the past, middle skills jobs required only a high school diploma, but changes in workplace requirements, production, and technology make additional training and education beyond high school the new norm for access to these types of jobs (Achieve, 2012).

The Workforce Skills Gap

There are two important skills gaps to be addressed in meeting the need for a productive middle skills workforce. One is comprised of traditional cognitive and academic proficiencies, such as reading and mathematical abilities. The second skills gap includes a broad array of non-cognitive and personal workforce competences (e.g., teamwork, time management, resilience) that have been demonstrated to be instrumental to workplace success and job performance (America's Edge, 2011; Casner-Lotto & Barrington, 2006; National Research Council, 2011). These non-cognitive elements are the applied skills and behaviors that employers indicate are becoming more and more essential to developing, maintaining, and enhancing a productive, accountable, and learning-ready workforce

Numerous large-scale efforts to define and guide training towards improving non-cognitive workforce competencies have clarified the behaviors, skills, and knowledge crucial for workplace success. These efforts have addressed the fact that a large percentage of adults – including both first time employees and experienced workers — often lack the non-cognitive workplace skills that will not only help them to function as effective employees, but can also lead to effective management and leadership career pathways. For instance, as early as the U.S. Department of Labor's (1991) groundbreaking Secretary's Commission on Achieving Necessary Skills Report, a large number of national and regional research and policy studies have been conducted to validate and illustrate those particular measurable noncognitive skills and behaviors that are vital to strengthening the U.S. workforce and contributing to workplace productivity (America's Edge, 2011; Bailey, 1997; Carnevale, Gainer, & Meltzer, 1990; Casner-Lotto & Barrington, 2006; National Research Council, 2011; Society for Human Resources Management, 2008).

The assessment and development of traditional cognitive and academic skills are, of course, foundational elements of the nation's formal education systems. Yet, while non-cognitive, applied workforce skills and behaviors are equally critical to both corporate and employee success, the means, programs, and venues for assessing, developing and supporting these necessary workplace competencies are far less systemic and systematic.

Assessing and Developing Non-cognitive Workforce Skills

The ETS Work Readiness Strength Assessment & Training System (WRSATS) has been designed to assess and develop a wide array of non-cognitive workforce skills found to be essential for workplace success. Its creation has been guided by the recognized need to enhance the readiness skills of a

broad range of individuals either entering or transitioning in the workforce, and to meet the significant demands of the U.S.'s rapidly changing economy.

The WRSATS focuses on six major workforce success skill sets: (a) Initiative and Perseverance, (b) Responsibility, (c) Flexibility and Resilience, (d) Teamwork and Citizenship, (e) Customer Service Orientation, and (f) Problem Solving and Ingenuity. Across these six areas are arrayed thirteen WRSATS subskills. These thirteen subskills further define and delineate the six broad skill sets. All WRSATS skill areas and subskills have been shown to be significantly correlated with workplace readiness and success.

The WRSATS system encompasses three major interfacing components: The (a) Strength Assessment, (b) Strength Profile, and (c) Training System.

- The Work Readiness Strength Assessment is a 35-45 minute computer-adaptive test, consisting of 104 questions. It is designed to measure the test taker's levels of strength and weakness across the non-cognitive skills and behaviors encompassed within each of the six focal WRSATS domains and 13 WRSATS subskills. The Strength Assessment accomplishes these goals by means of a forced-choice test comprised of a series of paired statements, where each statement is recognized as similar in social desirability. For each question (i.e., set of paired statements), individuals must choose between the two statements, each representing a different non-cognitive workforce skill or behavior, and decide which statement is more like them. The adaptive nature of the Strength Assessment and the fact that paired answer statements are of similar social desirability result in the test being "fake resistant." Additionally, since the statements are selected adaptively within the Strength Assessment program structure for each individual test taker and there are no right answers, the need for a secure testing environment is reduced opening options for a range of user-friendly test administration venues.

- The four-page Work Readiness Strength Profile provides the detailed report (i.e., formal score report) of the test taker's performance on the Strength Assessment. As such, it provides the test taker's measured level of strength and weakness for each of the six focal WRSATS domains and thirteen WRSATS subskills. All Strength Assessment scores are reported at one of three levels (High, Moderate, Low) referenced to a normative sample. Additionally, the Strength Profile provides information and links that support more detailed score interpretation, skills development, and training information. Each Interpretive Statement also offers a link directly to the Module in the Training System relevant to the particular score being reviewed by the test taker.

- The Work Readiness Training System is a flexible, modularized set of training materials that are organized based on the six major WRSATS skill areas. Drawing from best practices in adult and workforce education programming (Eyster et al., 2011), each module begins with an introduction, and presents measurable and attainable objectives. Users are offered guiding questions to think about while working through activities in the module. The modules are designed to systematically guide learners through a series of activities, building one upon the other, and are aligned with a central goal that motivates a set of content materials and associated tasks. Each module is contextualized within scenarios that guide learning and build readiness skills, and incorporates activities that scaffold adults' vast and varied life, educational, and workforce experiences. The Training System is intended for self-guided use, classroom application, and/or the workforce training environment, and can be incorporated into another integrated learning environment.

Multiple psychometric studies support the development of the WRSATS (e.g., Drasgow et al., 2012; Shore et al., 2014), and there are several ongoing research projects to evaluate its application and impact. Foundational among the latter are several pilot studies to determine and improve the extent to which the WRSATS may be feasibly implemented within a real-world training environment. Future research will focus on how effective the WRSATS is as an intervention in terms of its effects on individual work readiness skill levels. While the Training System materials were designed primarily with adult learning and student workforce skills training in mind, ETS also plans to formally evaluate whether significant impact(s) on work readiness skill levels or observable workplace behaviors should be expected after exposure to the WRSATS.

References

Achieve (2012). The future of the U.S. workforce. Washington, DC: Author.

America's Edge (2011). Can California compete? Reducing the skills gap and creating a skilled workforce through linked learning. Washington, DC: Author.

Bailey, T. (1997). Changes in the nature of work: Implications for skills and assessment. In H. O'Neil Jr. (Ed.), Workforce readiness: Competencies and assessment. Mahwah, NJ: Lawrence Erlbaum Associates.

Carnevale, A. P., Gainer, L. J. & Meltzer, A. S. (1990). Workplace basics: The Essential skills employers want. San Francisco: Jossey-Bass.

Carnevale, A.P., Smith, N. & Strohl, J. (2013). Recovery: Job growth and education requirements through 2020. Washington, DC: Georgetown Public Policy Institute, Center on Education and the Workforce.

Casner-Lotto, J. & Barrington, L. (2006). Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce. New York, NY: The Conference Board, Partnership for 21st Century Skills, Corporate Voices for Working Families, & society for Human Resources Management.

Drasgow, F., Stark, S, Chernyshenko, O. S., Nye, C. D., Hulin, C. L. & White, L. A. (2012). Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions (Technical Report No. 1311). Fort Belvoir, VA: Army Research Institute for the Behavioral and Social Sciences.

Eyster, L., Smith Nightingale, D., Barnow, B. S., O'Brien, C. T., Trutko, J. & Kuehn, D. (2011). Implementation and early training outcomes of the High Growth Job Training Initiative: Final report. Washington, DC: The Urban Institute. Retrieved from http://www.urban.org/publications/412476.html.

Society for Human Resource Management (2008). Critical skills needs and resources for the changing workforce. Alexandria, VA: Author.

Shore, J., Noeth, R., Inglese, P., Lentini, J., Naemi, B., Rikoon, S. & Robbins, S. (2014). The ETS Work Readiness Strength Assessment & Training System: Technical research report. Manuscript submitted for publication.

National Research Council (2011). Assessing 21st century skills: Summary of a workshop. Washington, DC: National Academies Press.

U.S. Department of Labor (1991). What work requires of schools: A SCANS report for America 2000. Washington, DC: Author.

3 Presentation of the Young Researchers Projects' - Theoretical Consideration and Work in Progress

The four research groups focused on different topic from the field of higher education learning outcomes. In the following, the four groups present their projects – their theoretical considerations and work-in-progress. All groups received the following guiding questions in advance:

- What competence/competency(ies) is/are the focus of your work? How do you define the boundaries of what is included within the scope of your work on competence/competency/(ies)? What is/is not included?
- Where do you situate your work with respect to the "continuum" laid out by Blömeke, Gustafsson, and Shavelson (p. 8)?
- Given your position on Blömeke et al.'s continuum, what argument would you like to make about the interpretation of your test/measurement strategy?
    - What evidence would most convincingly support this argument (ignoring resource constraints)?
    - What evidence is most crucial to support this argument (i.e., what evidence is required to make a convincing validity argument)?
    - c) Taking into account the methods and findings from the Förster et al. paper, what special validity concerns might exist as you either develop a multi-country test or move from single- to multi-country uses of an existing test? Try to express your ideas using the language of the AERA/APA/NCME Standards.
- Considering Blömeke et al.'s call for combinations of approaches, how might you incorporate other aspects of the "continuum" into your work? (What additional questions and/or critiques might this allow you to address?)

Literature:

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education [AERA, APA & NCME] (2014). Standards for educational and psychological testing (2nd ed.). Washington, DC: American Educational Research Association.
- Blömeke, S., Gustafsson, J.-E & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. In S. Blömeke, J.-E. Gustafsson & R. Shavelson (2015) (Eds.). Assessment of Competencies in Higher Education. Topical Issue of the Journal for Psychology.
- Förster, M., Zlatkin-Troitschanskaia, O., Brückner, S., Happ, R., Hambelton, R., Walstad, B.W., Asano, T. & Yamaoka, M. (submitted). Validating Test Score Interpretations by Comparing the

Results of Students from the United States, Japan and Germany on a Test of Economic Knowledge in Higher Education. In S. Blömeke, J.-E. Gustafsson & R. Shavelson (2015) (Eds.). Assessment of Competencies in Higher Education. Topical Issue of the Journal for Psychology.

3.1 Competency-Based Model of Educator Development for Culturally and Linguistically Responsive Teachers: The Beginnings | Svenja Hammer and Kara Mitchell Viesca

Work in progress of this project

Children with another language beside the language of assessment have an increasing influence on the classroom discourse in Germany and the US. Due to the rising number of students with an immigration background teachers need increased expertise in teaching both second language acquisition and grade level content. As the National Center for Education Statistics stated in 2011, one out of five students in the U.S. speak a language other than English at home and furthermore, referring to the US Department of Education, nearly ten percent of students nationwide are designated as English Language Learners (ELLs) (de Cohen & Clewell, 2007; Gifford & Valdés, 2006). Looking at the PISA data from 2012, 5% of all 15 year-old students in Germany are immigrants who do not speak the language of assessment at home, in the USA the percentage is even higher with 12% (OECD, 2013). The results of the PISA study 2012 have shown that there are significant differences in the performance between immigrant[2] and non-immigrant students.

Due to the lack of preparation of regular content teachers to work with second language multilingual learners (Lucas, 2011) as well as the challenge of being assessed and participating in schooling processes in a language a student is still learning (Abedi, 2002; Menken, 2008; Solano-Flores & Li, 2008), multilingual learners who are still learning the dominant language of society consistently perform at lower levels than their monolingual peers. Research has also shown that although multilingual learners acquire basic interpersonal communication without much difficulty, they experience considerable

[2] The term “immigrant students” used here includes students whose two parents were born abroad but who, themselves, were born in the country of assessment or in another country. Additionally, while the term “ELL” is commonly used in the U.S. context, we will refer to students across both the US and German context as multilingual learners when they are students whose daily life constitutes the use of multiple languages, often comprised of the society’s dominant language at school and different language(s) at home and in their community.

difficulties when acquiring academic language skills in elementary and secondary schooling (e.g., Cummins, 2000). Consequentially multilingual learners have to face a double task in school: They have to meet high expectations regarding grade level subject matter knowledge while they are also developing the literacy and academic language competencies associated with that content. For this reason, Short and Fitzsimmons (2007) argue that multilingual learners have to do "double the work" of their native-speaking peers.

Accordingly, international researchers from reference disciplines call for systematic language facilitation integrated into the mainstream classroom (e.g., Leung, 2007). Correspondingly, teachers must be able to support multilingual learners' academic language development in reference to subject-specific requirements (e.g., Schleppegrell, 2010). As a result more and more mainstream teachers see themselves also as language teachers, even though they are not familiar with any language learning concepts (de Jong, 2013). They suddenly have to know how to combine language, literacy, and content instruction to make effective learning possible and to reach the rigorous subject matter expectations. However, with the increased expertise required of teachers to work with multilingual learners, there is also a lack of research on preparing mainstream content teachers to work with multilingual students (Freeman & Freeman, 2014). As an international field, we still have a great deal to learn about preparing teachers to meaningfully support the language and academic development of multilingual learners. Therefore, our project seeks to add significantly the knowledge of our field while also creating valuable tools and resources to support multilingual students and their teachers in Germany and the US.

Cochran-Smith and Villegas (2015) conducted a comprehensive literature review of around 1500 studies from across the world on teacher preparation and framed their review in terms of teacher preparation research as historically situated social practice. Based on their framing of the current and historical context, Cochran-Smith and Villegas argue that there are currently three larger programs of research in teacher preparation research: research on teacher preparation accountability, effectiveness, and policy; research on teacher preparation for the knowledge society; and research on teacher preparation for diversity and equity. Clearly, our project falls into the category of research on teacher preparation for diversity and equity. However, we have a long-term vision for our collaboration that we see spanning the scope of many years that will meaningfully intersect with the other two programs of research. While it is not within the scope of this paper to fully define our long-term vision and goals, we will briefly describe them here as the context for our current work and immediate collaborative plans.

In order to meaningfully support teachers to develop the skills they need to achieve equity for multilingual learners, we suggest that any model of teacher development will have to take into account

three things: teacher/teacher candidate competencies, teacher/teacher candidates learning opportunities, and K-12 multilingual student outcomes. However, at the moment, our field does not have a clear empirical sense around what any of these three things should be, or how they should intersect to support high quality learning for multilingual learners and their teachers. Therefore, our program of research will explore each of these components in depth as well as how they relate to one another and impact equity for multilingual students and their teachers. Through this extensive research our end goal is to create a competency-based model of educator development for culturally and linguistically responsive teachers. We will have a battery of qualitative and quantitative assessment/data collection points for teachers to take and engage with that will determine a profile of their strengths and opportunities for learning as equitable culturally and linguistically responsive teachers. With that profile, teachers will be provided options for supportive, innovative, flexible and differentiated learning opportunities related to their areas of potential growth that builds on their strengths. Once teachers have completed the learning opportunities provided to them, they will once again engage with data collection protocols and assessments to determine their growth as well as next steps (by getting a new profile and recommendations for ongoing learning), a cycle that can be repeated until a teacher reaches high levels of competency demonstration as a culturally and linguistically responsive teacher.

In order to meet this long-term goal of developing a competency-based model of educator development, we are building on two current projects to conduct research and expand our empirical foundation towards meeting the goals described above. The next sections will briefly describe our two projects, one in the US (eCALLMS) and one in Germany (DaZKom). Then the following sections will describe our current collaborations across these two projects that are the result of our time at the Autumn Academy.

eCALLMS

In 2011, eCALLMS (e-Learning Communities for Academic Language Learning in Mathematics and Science), a $1.9 million National Professional Development grant funded by the U.S. Office of English Language Acquisition (T365Z110177), was launched at the University of Colorado Denver (University of Colorado Denver, 2014). The overarching goal of this teaching grant is to support teachers and teacher educators in developing linguistically responsive practices that support multilingual learners' language and content development as well as bilingualism/biliteracy development. eCALLMS seeks to accomplish this through various initiatives that focus on teaching and learning opportunities at both the pre-service teacher and well as in-service teacher level. The major initiative for this grant, and

the focus of our collaboration across international contexts, is the development and launch of online professional learning opportunities for in-service teachers that we refer to as eCALLMS modules.

The eCALLMS modules have been designed in three strands, academic language learning in mathematics, academic language learning in science, and culturally and linguistically diverse education. To design the modules, teacher education faculty from mathematics education, science education, culturally and linguistically diverse education, literacy, educational foundations and instructional technology were recruited and formed into a module design team. As a team, eCALLMS authors work collaboratively across disciplines in the module design process. Together, the eCALLMS Module Design Team decided that the modules would last six weeks long and require two hours of a teacher's time each week. The team also decided on a consistent learning cycle in the modules where each week, module participants explore new content then try activities to make those new ideas work in their teaching practice. To end the week, module users come back online and share their learning with their module colleagues. Essentially, eCALLMS modules are intended to be high quality resources to support meaningful professional learning for groups of teachers working in professional learning communities (DuFour, 2004). We have been testing our modules with teachers for years and have consistently received positive feedback (Viesca, Russell, Gutierrez, Leonard, & Nocon, 2014), however, our research agenda around the impact and outcomes of these modules is just beginning. Up to date we have six modules launched, four modules that will soon be launched and around 18 more in various stages of development. By the time our grant ends in 2016, we anticipate having launched around 30 modules for teachers to learn about improving their practice with multilingual learners. Currently, close to 150 teachers, including teachers in Finland, have used our modules. By the end of our project, we anticipate that number rising to around 500, including teachers in Germany.

DaZKom

The research study *DaZKom* (Professional competencies of prospective teachers (secondary schools) for German as a Second Language - GSL) is an interdisciplinary and inter-university project, funded by the German Federal Ministry of Education and Research (BMBF) program "Modeling and Measuring Competencies in Higher Education" (KoKoHs; Koch-Priewe et al. 2014). The objective of the project *DaZKom* is to theoretically model (Köker et al. 2015) and empirically validate (Hammer et al. 2015) pre-service teacher competencies in the field of German as a Second Language. By means of a test instrument that was developed within the project and measures pre-service teacher competencies. It is supposed to give empirically based insights on how learning opportunities in academic teacher training must be designed for enabling the acquisition of substantiated and standardized GSL-

competencies. Firstly, a competency model was developed on the basis of an analysis of 60 university curricula, which was validated by experts of the field. We discerned three main content areas, also in reference to relevant literature (Köker et al. 2015):

1. Subject-specific registers, which includes knowledge of lexical and grammatical structures as well as of semiotic systems, and the ability to recognize these in texts and student oral and written productions.
2. Multilingualism, which includes knowledge of the processes of second language acquisition and migration, and the ability to recognize these processes.
3. Didactics, which includes diagnosis of students' competencies and difficulties as well as language facilitation in the mainstream classroom.

Each sub-competency is further divided into three levels: (1) Novice, (2) Advanced Beginner and (3) Competent Actor. So GSL competency is supposed to be understood as a generic competency. Based on this a test instrument with 51 items was constructed which start with one out of four types of stimulus material (word problems, case studies, classroom interactions and written student productions). The items were finally tested in a study with N = 500 pre-service teachers from all subject areas. The test properties show good fit values for the Rasch model; as well as a good overlap for the difficulty of the items and the ability of the persons. Concerning learning opportunities and beliefs, questionnaires were developed that make it feasible to show relations between high GSL competency, having a lot of learning opportunities and valuing second language learners. Results of the test development process were published in Köker et al. (2015) and Hammer et al. (2015).

Working program

The planned joint project pursues three objectives that will be completed in 2015. We plan to have our collaboration extend over several years, but will focus here on what we have already started and plan to accomplish this year. Our *first* objective is to build off of the DaZKom project to translate and test the competency model in the US context. Our *second* objective is to build off of the eCALLMS project to translate a module on second language acquisition into German and research the learning outcomes it produces in the German context. Finally, our *third* objective will be to link the research and use of the DaZKom competency model and test instrument with an eCALLMS learning module to explore how the competency model intersects with learning opportunities and outcomes from eCALLMS modules in both the US and Germany. Additionally, we plan to present at multiple conferences, collaboratively write at least one major research grant, and submit a book prospectus to Routledge for an international book on the education of teachers to work with multilingual learners.

The following sections will describe each of the three objectives listed above in more detail as well as our overall collaborative plans for 2015.

Objective 1: Translate DaZKom Test Instrument into English

During our collaborative time in Berlin, we examined the DaZKom competency model and cross-walked the elements of the model with the major standards and frameworks impacting the education of teachers of multilingual students in the US and Europe. We focused on the TESOL/NCATE Standards (Teachers of English as a Second Language/National Council for Accreditation of Teacher Education), CAEP Standards (Council for Accreditation of Educator Preparation), CREDE Standards of Effective Pedagogy (Center for Research on Education, Diversity and Excellence), EUCIM-TE standards (European Core Curriculum for teacher education), BAMF standards (Bundesamt für Migration und Flüchtlinge). By examining the DaZKom competency model from the perspective of these various national and international standards, we solidified that the DaZKom competency model is sufficiently well aligned with US standards to likely be quite useful in the US context. For this reason, we have begun the work of translating the DaZKom test instrument into English and will test and validate it in the US context through the collaborations already underway between DaZKom and eCALLMS. We will follow similar procedures to how the instrument was tested in Germany, including the use of "cognitive labs" (Paulsen & Levine, 1999) to gain a deeper understanding of how the items are interpreted by participants. The translated test instrument will be tested in a pilot study with teachers/pre-service teachers of all subject areas. To develop a deeper understanding of the content dimensions of the test instrument, a validation study will be conducted in which external constructs will be tested.

Objective 2: Translate eCALLMS Second Language Acquisition Module into German

Because of the substantive overlap between the competency model and the frameworks guiding the eCALLMS project (something we determined during our collaborative time in Berlin), we believe there is value in making the learning opportunities from eCALLMS modules available to German pre-service teachers as well. We are currently translating one of the eCALLMS modules titled, "Second Language Acquisition" into German to be utilized in a course during the summer of 2015 with pre-service teachers at Leuphana University in Lüneburg. We are also designing a study to collect data on the impacts and outcomes of the use of the eCALLMS module in the German context.

Objective 3: Research the relationship between test instrument and learning module in both Germany and the US.

This final objective will require that we make great progress on objectives 1 and 2 in a timely matter. For this reason, our work for objectives 1 and 2 has long been underway and will continue for the next several months. We hope that by summer 2015, when the eCALLMS module is ready to be used in Germany, we will be far enough along to conduct research in both Germany and the US to examine the relationships between the DaZKom test instrument and the eCALLMS learning opportunities. Studying these relationships will help us move towards our larger vision described above as well as help us to conduct comparative research in the two contexts to learn about similarities and differences that can, in turn, impact the education of teachers of multilingual learners across even more national contexts.

In addition to these three objectives, we are presenting a paper with our initial investigations into the education of teachers of multilingual learners in the US and Germany at the American Educational Research Association Annual Conference in Chicago in April 2015. Our work is part of a symposium that we developed that includes research from other US researchers as well as from Finland. In June 2015, we are presenting our collaborations between DaZKom and eCALLMS at the International Society for Language Studies conference to be held in Albuquerque, New Mexico. During that time we also plan to take several days to collaboratively write a large research grant to be submitted around August 1st 2015 to the US Institute of Education Sciences. We also hope to submit a book prospectus to Routledge about international perspectives on the preparation of teachers to work with multilingual students by July and have already recruited chapters by leading researchers in the US, Finland, Australia and Germany. In September 2015 we are planning a research collaboration in Colorado where we can collaboratively collect data and conduct analyses to further our collaborations across DaZKom and eCALLMS as well as reach closer to our long-term vision of developing a competency-based model of educator development for culturally and linguistically responsive educators in both Germany and the US. Because of our ongoing collaborations, it was possible to invite Kara Mitchell Viesca to give some international insights into teacher preparation programs in the U.S. for the German journal *Die Deutsche Schule* which will be published in the end of 2015. We are also currently working on a proposal for the conference of the Deutsche Gesellschaft für Erziehungswissenschaft in Kassel in 2016.

Theoretical consideration

1) What competence/competency(ies) is/are the focus of your work? How do you define the boundaries of what is included within the scope of your work on competence/competency/(ies)? What is/is not included?

Our work focuses on the competency of teachers and pre-service teachers regarding multilingual learners. In our theoretical approach the teachers' competency consists of cognitive, conative, affective and motivational resources and is considered learnable. In our recent working program we are focusing on the cognitive and affective resources of teachers and pre-service teachers. In our study we draw on the theoretical framework from Villegas and Lucas (2011). They provide a systematic overview on special types of expertise needed for teaching multilingual learners and distinguish between *orientations* and *knowledge and skills* of linguistically responsive teachers.

We regard competency as a person's characteristic that drives his or her behavior. In specific, we see the competency of a linguistically and culturally responsive teacher as a conglomerate of knowledge about multilingual learning (Lucas, Villegas, & Freedson-Gonzalez, 2008) and student facilitation (Tharp, Estrada, Dalton, & Yamauchi, 2000) and the person's positive beliefs about multilingual students and responsibility for facilitating those students (Huerta, 2011). We assume that this mixture of resources leads to a competent linguistically and culturally responsive teacher who is able to facilitate multilingual learners in an adequate way to decrease inequity in the educational system.

2) Where do you situate your work with respect to the "continuum" laid out by Blömeke, Gustafsson, and Shavelson (p. 8)?

Regarding the "continuum" laid out by Blömeke, Gustafsson, and Shavelson (2015, p. 8), which differentiates between *disposition*, *situation-specific skills*, and *performance*, the concept of linguistically responsive teachers can be seen in the areas of disposition and situation-specific skills. Pre-service teachers with Second Language Competency mean to have knowledge about the meaning of language in pedagogical situations; especially knowledge about subject-specific registers, multilingualism, and didactics. Therefore, this theoretical construct can be situated in the area of disposition under *cognition* on the continuum. Besides the assessment instrument we are developing a questionnaire measuring teachers' beliefs concerning multilingual learners. This construct can be seen in the area of disposition under *affect-motivation* on the continuum.

Our second objective which has the aim to offer tailored learning opportunities for teachers and pre-service teachers can be seen as situation-specific skills. The eCALLMS e-modules give them the opportunity to not only extend their knowledge regarding multilingual learners and their facilitation but they also get the chance to evaluate their own perception, interpretation, and decision making of teaching situations by getting stimuli that set those processes in motion.

The learning opportunities that are presented in the eCALLMS modules are to increase the cultural and linguistic responsiveness in teachers and teacher candidates.

Our third objective, where we are planning to offer results about the relationship between the test instrument and the learning opportunities, is to be situated between disposition and situation-specific skills.

3) Given your position on Blömeke et al.'s continuum, what argument would you like to make about the interpretation of your test/measurement strategy?
a) What evidence would most convincingly support this argument (ignoring resource constraints)?
b) What evidence is most crucial to support this argument (i.e., what evidence is required to make a convincing validity argument)?

To focus just on one of our objectives, we will describe the validity argument for the German version of the DaZKom test instrument mentioned above that can be transferred for the English version. To get *evidence based on test content* the DaZKom test items were given to experts of the field to be rated. By this the fit between the items and the construct's dimensions was ensured. During the item development, cognitive labs were conducted where participants were asked to think aloud while completing the tasks. Through these efforts, evidence based on response processes was gained.
The correlations between the test items' dimensions showed expected results concerning the construct's dimensions; showing *evidence based on internal structure*. To get a deeper understanding of the content dimensions of the test and evidence based on relations to other variables, we tested three external constructs: linguistic knowledge, pedagogical knowledge, and mathematical content knowledge. The test instrument for pedagogical knowledge and mathematical content knowledge already exist in English, an English version for the linguistics test needs to be found. Furthermore, it would be interesting to test external constructs like intelligence and reading ability.
Ignoring resource constraints, testing a large cohort of pre-service and in-service teachers of different subjects and from different school districts around the US and Germany would support our validity argument.
c) Taking into account the methods and findings from the Förster et al. paper, what special validity concerns might exist as you either develop a multi-country test or move from single- to multi-country uses of an existing test?

The test instruments will have the same focus of competency in mind but will be adjusted to the differing educational system demands and legal requirements. Therefore a one-to-one translation will not be possible. But in the long run results from the German and English assessments could be used to enrich each other's learning opportunities.

4) Considering Blömeke et al.'s call for combinations of approaches, how might you incorporate other aspects of the "continuum" into your work? (What additional questions and/or critiques might this allow you to address?)

Our working program has a focus on the disposition and situation-specific skills. The last part of the continuum *performance* was not mentioned so far. In regard to our plans some crucial questions follow, e.g. 'Are teachers with high test scores in disposition better in facilitating multilingual learners when observing their teaching?' or 'Are the outcomes of low achieving students and multilingual learners higher if they are taught by highly competent linguistically and culturally responsive teachers?' Those questions will need to be addressed to fully understand if *disposition* leads to *performance* in the end. And over the long-term course of our collaboration, we plan to explore these types of questions.

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, *8*(3), 231-257.

Cochran-Smith, M. & Villegas, A.M. (2015). Framing teacher preparation research: An overview of the field, part 1. *Journal of Teacher Education*, *66*(1), 7-20.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire.*

Clevedon, England: Multilingual Matters.

de Cohen, C. C. & Clewell, B. C. (2007). Putting English language learners on the educational map. The No Child Left Behind Act Implemented. In The Urban institute: *Education in Focus. Urban institute policy brief.* Retrieved from: http://files.eric.ed.gov/fulltext/ED498664.pdf.

de Jong, E. (2013). Preparing Mainstream Teachers for Multilingual Classrooms. *Association of Mexican-American Educators (AMAE) Special Invited Issue, 7(2).*

DuFour, R. (2004). What is a professional learning community? *School as Learning Communities, 61*(8), 6-11.

Freeman, Y.S. & Freeman, D. (Eds.) (2014). *Research on preparing preservice teachers to work effectively with emergent bilinguals*. Emerald Group Publishing Limited, Bingley, UK.

Gifford, B. & Valdés, G. (2006).The Linguistic Isolation of Hispanic Students in California's Public Schools: The Challenge of Reintegration. *Yearbook of the National Society for the Study of Education. 105(2). 125-154.*

Hammer, S., Carlson, S. A., Ehmke, T., Koch-Priewe, B., Köker, A., Ohm, U., Rosenbrock, S. & Schulze, N. (2015). Kompetenz von Lehramtsstudierenden in Deutsch als Zweitsprache: Validierung des GSL-TC-Testinstruments. In S. Blömeke & O. Zlatkin-Troitschanskaia (2015) (Hrsg.). Kompetenzen von Studierenden. Zeitschrift für Pädagogik (Beiheft).

Huerta, T. M. (2011). Humanizing pedagogy: Beliefs and practices on the teaching of Latino children. *Bilingual Research Journal*, *34*(1), 38–57.

Koch-Priewe, B., Köker, A., Ohm, U., Ehmke, T., Carlson, S. A., Hammer, S., Rosenbrock, S. & Schulze, N. (2014). Professional competencies of prospective teachers (secondary schools) for German as a Second Language (GSL). Retrieved from: http://www.kompetenzen-im-hochschulsektor.de/177_ENG_HTML.php.

Köker, A., Rosenbrock, S., Ohm, U., Carlson, S. A., Ehmke, T., Hammer, S., Koch-Priewe, B. & Schulze, N. (2015). DaZKom- Ein Modell von Lehrerkompetenz im Bereich Deutsch als Zweitsprache. In B. Koch-Priewe, A. Köker, J. Seifried, & E. Wuttke (Hrsg.). Welche Kompetenzen brauchen Lehramtsstudierende und angehende ErzieherInnen? Theoretische und empirische Zugänge. Bad Heilbrunn: Klinkhardt.

Lucas, T. (2011). *Teacher preparation for linguistically diverse classrooms: A resource for teacher educators.* New York, NY: Routledge.

Lucas, T., Villegas, A. M. & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education: Preparing classroom teachers to teach English language learners. *Journal of Teacher Education*, *29*(4), 361-373.

Leung, C. (2007). English as an additional language policy: issues of inclusive access and language learning in the mainstream. *NALDIC Quarterly, 3 (1),* 16-26.

Menken, K. (2008). *English learners left behind: Standardized testing as language policy.*

Tonawanda, NY: Multilingual Matters.

OECD (2013), *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (Volume II),* PISA, OECD Publishing.

Schleppegrell, M. (2010). Language in Mathematics Teaching and Learning. In J. Moschkovich (Ed.), *Language and Mathematics Education* (pp. 73-112). Charlotte: Information Age Publishing.

Short, D. & Fitzsimmons, S. (2007). Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners - A report to Carnegie Corpo-

ration of New York. Retrieved from: http://carnegie.org/fileadmin/Media/Publications/PDF/DoubletheWork.pdf.

Solano-Flores, G. & Li, Min. (2008). Examining the dependability of academic achievement measures for English language learners. *Assessment for Effective Intervention*, *33*(3), 135-144.

Tharp, R. G., Estrada, P., Dalton, S. S. & Yamauchi, L. A. (2000). Teaching transformed: Achieving excellence, fairness, inclusion, and harmony. Boulder, CO: Westview Press.

University of Colorado Denver (2014). eCALLMS: Supporting Linguistically Responsive Teaching. Retrieved from: http://ecallms.ucdsehd.net/about.

Viesca, K. M., Russell, N. M., Gutierrez, C., Leonard, J. & Nocon, H. (April 2014). *Teacher perspectives of linguistically responsive teaching.* Paper presented at the American Education Researchers Association. Philadelphia, PA.

Villegas, A. M. & Lucas, T. (2011). Preparing classroom teachers for English language learners: The policy context. In Lucas, T. (Ed.), *Teacher preparation for linguistically diverse classrooms. A resource for teacher educators* (pp. 55-72). New York: Routledge.

3.2 Modeling and Measuring the Competence of Teachers to Teach Computational Thinking | Marc Berges, Jonathan Good, Melanie Margaritis, Leigh Graves Wolf, Aman Yadav

Work in Progress of this Project

In the fall of 2014, we had an opportunity to attend the KoKoHs Autumn Academy in Berlin, Germany. In total 12 participants from different countries around the world met to discuss important topics and to create new research projects. A team of experts helped us to frame our ideas and to write a proposal for a project. Our project group consists of three people from Michigan State University, USA, one research assistant from Technische Universität München (TUM School of Education), Germany, and one research assistant from the University of Paderborn, Germany. Our goal was to create an international project group on computational thinking. In the panel discussion at the SITE Conference 2015, we are going to discuss “Computer Science and Computational Thinking: International Perspectives on Developing Student and Teacher Competencies”.

This paper is about our work at the KoKoHs Autumn Academy in Berlin, Germany.

In the first part, we are going to describe what computational thinking is.

<u>Computational Thinking</u>

Computational Thinking (CT), as defined by Jeannette Wing, "involves solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science" (2006, p. 33). Her paper served as a call to action for educators in not just computer science courses, but for those in science, mathematics, engineering, and the humanities. Researchers in this area have come to agree that the availability of computing power in students' lives demands an additional type of literacy in their daily lives (Grover & Pea, 2013). Students need to be able to recognize problems that could be analyzed and solved using CT techniques, as these techniques have a utility far beyond computer programming. Students that become proficient in these techniques will be able to leverage the power of computing in almost any field, needing them to both survive and thrive in the future world of work. Grover and Pea (2013, pp. 39–40) offered a list of concepts included within CT: Abstractions and pattern generalizations (including models and simulations), Systematic processing of information, Symbol systems and representations, Algorithmic notions of flow of control, Structured problem decomposition (modularizing), Iterative, recursive, and parallel thinking, Conditional logic, Efficiency and performance constraints, and debugging and systematic error detection.

Since Wing's first article on CT was published, differing definitions have been developed around the term 'computational thinking' (Grover & Pea, 2013, p. 39). There are some common notions, though, that we can identify throughout these competing definitions.

First, learning computational thinking skills is not dependent on learning computer programming skills. Learning how to program can be helpful in developing CT skills, but is not required. In fact, many CT proponents explicitly try to divorce CT lessons from the learning of coding. One such example, the "Computer Science Unplugged" curriculum ("Computer Science Unplugged", n.d.), developed for use without a computer, provides lessons for teaching Computer Science topics such as information theory, error detection, and sorting algorithms.

Second, computational thinking is an attempt to leverage computing power, but not an attempt to restrict thinking to only computational methods. Wing was very clear on this in saying "Computational thinking is a way humans solve problems; it is not trying to get humans to think like computers" (2006, p. 35). Computational thinking simply recognizes that students have and will continue to have access to large amounts of computing power and data. What CT is trying to do is give students the skills to approach problems with these possibilities in mind, using computing to enhance their work, and use this within multiple contexts.

Third, computational thinking includes a set of problem solving skills but is also a way in which to view the world. CT is an attempt to distil and then transfer the perceptions and skills that computer scientists develop in practice. Rather than require years of experience in coding to gain these capabil-

ities, we are attempting to integrate their use into non-coding tasks within other domains. Should a student decide to become a programmer later, they will be able to use these skills. However, what is possibly more powerful is that the non-programmer can approach problems and prepare their solutions with computing tools in mind.

Research idea(s)

The idea of our research team is to research in the area of computational thinking in combination with competence development. Our target groups are teachers in the United States and Germany (maybe Europe). We do not focus only on Computer Science teachers but also on teachers from other areas, such as mathematics or languages.

The research questions, which accompanied our work, are: (1) What competencies on computational thinking do teachers need to successfully teach computational thinking in class? and (2) What competencies on computational thinking do teachers already have?

To achieve our first research question we are planning to develop a competence model for computational thinking competencies by doing literature reviews in the first step to get the first categories and definitions for the model. To validate our model, the next step is to conduct expert interviews or expert ratings.

Once the competence model is validated we can start to fulfill our second research question.

For this reason, we will create a measurement instrument on the basis of our computational thinking competence model to test teachers on their existing computational thinking skills.

By means of this investigation we can see if, and in what way, teacher preparation has to be changed to successfully teach computational thinking in class.

Current Status

During our week at the KoKoHs Autumn Academy in Berlin Germany, we began a literature review to create our first categories for our computational thinking model. The following categories are a refined model based upon 2011 CSTA's CT teachers resources. We removed parallelization because it is not a CT skill.

We collapsed abstraction, modeling and simulation into abstraction as we see models/simulations as abstraction of real world. Algorithm & procedures collapsed to Algorithmic thinking.

data collection	problem decomposition	automation

data analysis	abstraction	simulation
data representation	algorithms & procedures	parallelization

Moreover, we applied for a panel discussion at the SITE conference in Las Vegas, USA, which will be on March 2015. During that panel, we would like to discuss our methodological approach and refine our next steps.

The biggest obstacle for our research project will be to find funding for the project.

Further Steps and Work

The next steps are finding funding for our computational research project and refining our research ideas after the panel discussion at the SITE 2015 conference in Las Vegas, USA.

Moreover, we are planning to write a paper for a journal with our ideas and the first results.

Theoretical Consideration

1) What competence/competency(ies) is/are the focus of your work? How do you define the boundaries of what is included within the scope of your work on competence/competency/(ies)? What is/is not included?

The main goal of our project is to investigate competencies teachers have to teach computational thinking in class. At this moment, not only computer science teachers are considered, but also teachers of different subjects. The basic notions of computational thinking are described by Barr et al. (2011). Here, the aspects of different subjects are included. Based on this theoretical framework, the subjects are investigated on their implementation of computational thinking. Our research focuses on the teaching process rather than the own application of the CT notions. The use of computational thinking ideas during the preparation process is another interesting question.

2) Where do you situate your work with respect to the “continuum” laid out by Blömeke, Gustafsson, and Shavelson (p. 8)?

We situate our work in the middle of the continuum presented by Blömeke et al. (2014). As mentioned above, we are mainly interested in the competencies teachers need for teaching computational thinking. As the investigated population is non-CS teachers, they won't plan to teach CT explicitly. Contrarily, the interpretation of the computer science related topic CT in other subjects is the interesting facet. The occurrence of these facets is meant to be in the interpretation and decision making during the teaching process. Nevertheless, the former dispositions mentioned in the continuum have to be considered as well. Furthermore, observable behavior or the performance, respectively, are of interest.

3) Given your position on Blömeke et al.'s continuum, what argument would you like to make about the interpretation of your test/measurement strategy?

Basically, there are two facets in the test strategy and two aspects of computational thinking in the teaching process. The interpretation and decision making during the teaching process can be measured by presenting typical or critical scenarios and asking the participants for their behaviour in such a situation. This can be approached by using critical incidents or written and/or video vignettes, respectively. Second, the use of computational thinking in the teaching process can be observed directly during classes. So, on the one hand, there is preparatory work that can be investigated in a survey and on the other hand, there is observable behaviour that has to be investigated directly.

a) What evidence would most convincingly support this argument (ignoring resource constraints)?

The most convincing evidence for the decision making during the teaching process would be an observation of complete teaching sequences. The preparatory aspect would be best observed by a complete summary of the preparatory work. This includes video observation and cognitive labs, as well as a kind of diary for the work.

b) What evidence is most crucial to support this argument (i.e., what evidence is required to make a convincing validity argument)?

The most crucial element in the measurement is the different naming of the concepts related to computational thinking. As we postulated in the beginning, most non-CS teachers still use computational thinking without knowing. They simply name the concepts in another way. So, part of the test development will be the finding of evidence for having synonyms for the concepts of computational thinking.

c) Taking into account the methods and findings from the Förster et al. paper, what special validity concerns might exist as you either develop a multi-country test or move from single- to multi-country uses of an existing test? Try to express your ideas using the language of the AERA/APA/NCME Standards.

Our test instrument has to be applied in Germany as well as in the United States, some language differences have to be faced. In a preliminary study, the concepts of computational thinking have to be proven to be the same in both languages. Furthermore, the differences in the teacher-education system have to be enlightened. More precisely, the explicit teaching of computational thinking facets during the educational process is of interest. Additionally, the applicability of vignettes of critical incidents has been proven for both counties. Again, the differences in the educational system might be challenging.

4) Considering Blömeke et al.'s call for combinations of approaches, how might you incorporate other aspects of the "continuum" into your work? (What additional questions and/or critiques might this allow you to address?)

As mentioned above, we want to consider the two main aspects of Blömeke et al.'s continuum. Nevertheless, the first cognitive aspects could provide interesting insights. For example, the application of cognitive labs could show the use of computational thinking concepts in the cognitive processes of teachers during their preparation or during classes.

References

Barr, V. & Stephenson, C. (2011). Bringing Computational Thinking to K-12: What is Involved and What is the Role of the Computer Science Education Community? ACM Inroads, 2(1), 48–54.

Blömeke, S., Gustafsson, J.-E. & Shavelson, R. (2014). Beyond dichotomies: Competence viewed as a continuum. Zeitschrift für Psychologie.

Grover, S., & Pea, R. (2013). Computational Thinking in K-12: A Review of the State of the Field. Educational Researcher, 42(1), 38–43.

Wing, J. M. (2006). Computational thinking. Commun. ACM, 49(3), 33–35.

3.3 Evidence-based Reasoning in Higher Education. Development and Validation of a Problem-oriented Vignettes Test to Access Educational Research Literacy | Raffaela Wolf Christina Haberfellner, Jana Groß Ophoff

Work in Progress of this Project

Introduction

Due to scientific progress, modern knowledge is constantly changing (Grundmann & Stehr, 2012; UNESCO, 2005). Skill sets that enable individuals to purposefully access, comprehend, reflect scientific evidence, and apply the resulting conclusions in complex situations are essential for successful participation in the 21st century workforce. With respect to educational decisions this ability is referred to as *Educational Research Literacy* (ERL, Groß Ophoff, Schladitz, Lohrmann, & Wirtz, 2014; Shank & Brown, 2007). Following Brown et al. (2010), only through systematic organization and processing of (re-)search, raw information or data turns into evidence. This allows distinguishing between basic and advanced levels of ERL: Whereas laymen support claims with arbitrary findings or experiences, experts base their conclusions on systematic and goal-oriented inquiry, analysis and interpretation of evidence.

Borg (2010) emphasizes that although current and future practitioners in education need to engage themselves *with research*, they do not necessarily have to engage themselves *in* research. Especially with view to continuous professionalization, the engagement with research in educational contexts is necessary, but not without difficulties: On the one hand, scientific evidence is formulated falsifiable and generalizable. On the other hand, educational practice aims at solving problems instantly and efficiently. It is for this gap between theory and practice that practitioners frequently comprehend research information as abstract, irrelevant factual knowledge that cannot be applied to practical problems (Astleitner & Herber, 2008; Borg, 2010; Hammersley, 2004; Harris & Brown, 2009; Patry, 2000; Weiss, 1998). Furthermore, the ability to reflect and use evidence is neither necessarily developed nor retrieved in an optimal manner in adulthood – not even in HE (Barchfeld & Sodian, 2009; Groth, 2007; Kuhn & Franklin, 2007; Schunn & Anderson, 2001). As students, graduates and professionals will be responsible for imparting relevant competencies to future generations, education plays a central role. Hence, future educators have to be trained to use research knowledge for practice (Shank & Brown, 2007). Higher Education institutions are particularly suitable for this as they provide research-based education. Thus, the focus of educational research in Higher Education (HE) has shifted to the requirement of evidence-based educational practice.

Although the notion of promoting competencies such as ERL in academic training is not new, the connotation of assessing and measuring such complex competencies with a valid and reliable instrument may prove psychometric challenges both on conceptual/curricular and on (inter)national comparative grounds. The challenge of assessing ERL may best be understood by situating the skills in a cognitive framework. For example, following Blömeke, Gustavson and Shavelson (in press), competence is defined as situation-specific skill that underlies observable performance (holistic approach, see Figure 1), but manifests itself through a combination of multiple facets such as cognition, conation, affect and motivation (dispositional approach). Divergent point of views (dispositional or holistic approach) lead to different interpretations of observable behaviours. The preponderance of instruments used for assessing ERL consist of selected response questions (SRQ) only. But problem-oriented performance tasks (PTs) are deemed suitable for assessing higher order thinking skills such as the ability of transferring research knowledge into practice (cf., Klein, Benjamin, Shavelson, & Bolus, 2007; Oakleaf, 2008). Consequently, it is questionable whether it is feasible to combine the dispositional and holistic approach to ensure appropriate measurement of ERL.

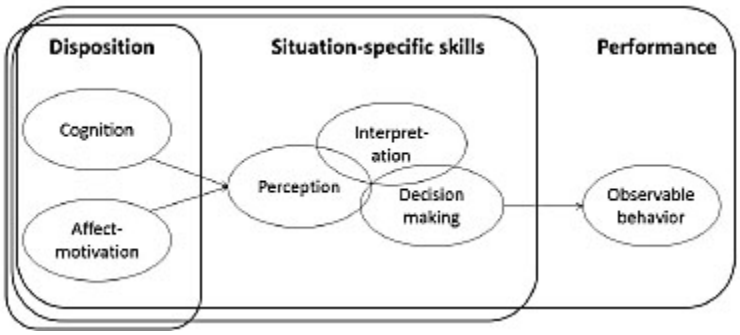

Figure 1. Cognitive Skills Continuum. (Adapted from Blömeke et al., in press)

Theoretical Background

The idea to promote the development of ERL is not novel (BAK, 1970). However, only after the so-called TIMSS astonishment and the PISA shock and the resulting empirical shift in the education system in Germany (Bos, Postlethwaite, & Gebauer, 2010; Hoffmann-Göttig, Eschmann, & Daumen, 2005) and some time later in Austria (Altrichter, Brüsemeister, & Heinrich, 2005), research literacy is not only included in general definitions of standards and objectives for Higher Education degrees (Kultusministerkonferenz, 2005; Wissenschaftsrat, 2000), but can also be found in the context of degree programs in Educational Science, e.g., in teacher education curricula (Kultusministerkonferenz, 2004), or as requirement of the accreditation of new degrees like Bachelor of Arts in Early Education (University of Education Freiburg, 2009). According to that, future educators should be able to reflect and evaluate evidence from educational research, whereas practitioners should also be able to use evidence-based insights for classroom and school practice. Closely related to this development

is a more distinct orientation towards learning outcomes (Ditton, 2010; Klieme et al., 2003) . Especially in the field of competence testing and large-scale assessment, Zlatkin-Troitschanskaia und Kuhn (2010) emphasize a great need for research in HE. Thus, within the funding initiative "Modeling and Measuring Competencies in Higher Education" (KoKoHs) by the German Federal Ministry of Education and Research (BMBF), test instruments are being developed to assess academic competences in different courses of studies (Blömeke & Zlatkin-Troitschanskaja, 2013).

In the context of this initiative, the German cross-university research project *Learning the Science of Education (LeScEd)*, a comprehensive ERL model was conceptualized and empirically validated (Groß Ophoff et al., 2014): The ability of students of educational sciences to *engage with research* (e.g. Borg, 2010) is defined as stepwise research cycle (cf., evidence-based medicine, Flores-Mateo & Argimon, 2007; Shaneyfelt et al., 2006): 1) question asked relative to a concrete problem, 2) search for information (e.g. via internet, data bases), 3) evaluate/appraise evidence, 4) integrate information, draw conclusion (e.g. instructional decision) and apply to solve the problem (knowledge application, cf., Greiff, Kretzschmar, & Leutner, 2014; Novick & Bassok, 2005). The ability to formulate appropriate (research) questions and to search and evaluate necessary information is usually investigated under the term *Information Literacy* (IL) in information science (e.g. Blixrud, 2003; Catts & Lau, 2008; Homeyer, 2008). Within the research cycle, this ERL subdimension can be situated between the Ask- and the (Re-)Search-step. Following this, it is necessary to be able to read and organize data, and interact with different data representations. This ability to understand and evaluate especially numerical information, is investigated as Statistical Literacy (SL) in the field of mathematics education (e.g. Ben-Zvi & Garfield, 2004; Groth, 2007; Watson & Callingham, 2003). The next progression involves an analysis of available evidence with view to patterns and rules (logic) and to substantiate reasoning, or to evaluate given conclusions with view to objectivity, reliability and validity (EB = Evidence-Based Reasoning). Approaches to investigate this subdimension can be found in research on Science Literacy (Brown, Nagashima, Fu, Timms, & Wilson, 2010; Kuhn, Iordanou, Pease, & Wirkala, 2010) or Critical Thinking (CT, z.B. Dunn, Halonen, & Smith, 2008).

Within the context of ERL assessment, recent model fit analyses (MIRT, Hartig & Höhler, 2009) suggested a three-dimensional model with the competence facets *Information Literacy, Statistical Literacy* and *Evidence-based Reasoning* (Groß Ophoff, Schladitz, Leuders, Leuders, & Wirtz, submitted; Groß Ophoff, Schladitz, & Wirtz, 2014). But the final step of integrating multiple sources of evidence (in the sense of Problem Solving, e.g., Novick & Bassok, 2005; Phye, 2001) to make logical decisions is not addressed by the appointed test instrument, which mainly consists of SRQs. But whether practically relevant competences can be validly operationalized this way, is critically scrutinized (Oakleaf, 2008). Alternative assessment procedures rely on performance tasks (e.g. Collegiate Learning As-

sessment, Klein et al., 2007), for which test participants have to solve different tasks or write short essays (e.g. Baer et al., 2007). Authenticity and situatedness are the merits of such assessment procedures which may be useful to measure complex constructs such as ERL.

It is evident that both item formats, SRQs and PTs, have their own strength and weaknesses respectively. Employing both item formats in one assessment as planned in this research project affords the opportunity to balance the associated strength and weaknesses of each item format. In our research project we will combine both ways of operationalization by using PT and SRQ within one assessment. One part is a performance task (PT) consisting of a document library in combination with a "real life" task and represents the holistic view of competence. The second part consists of selected response questions (SRQ). Both parts cover the four theoretically assumed dimensions *Information Literacy, Statistical Literacy, Ecidence-Based Reasoning* and *Problem Solving*.

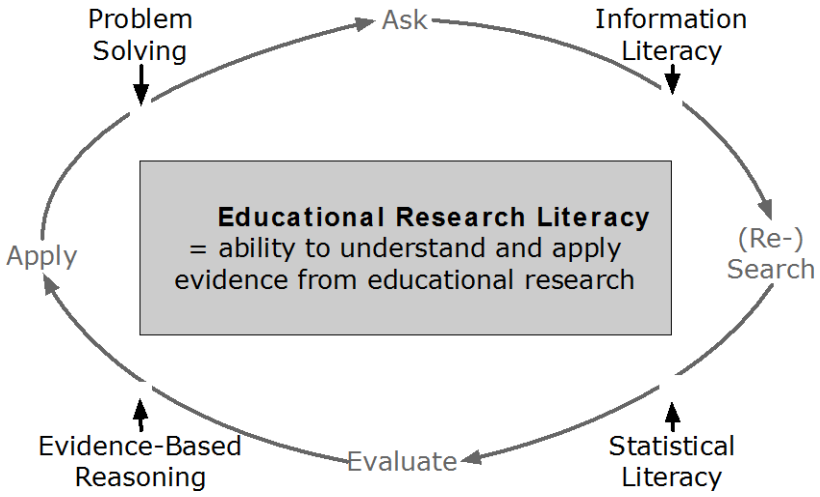

Figure 2. Research Cycle ERL.

Hypotheses

The purpose of the current project is to develop/refine, implement, and validate a mixed-format (i.e. PTs and SRQs) instrument that measures ERL within an international assessment framework. Participating countries may include the United States and traditionally more education-oriented states like Austria and Germany: For example, American teacher training can be characterized as "functional and market-oriented". The structure of German teacher training is by contrast "fragmented and education-oriented" (Blömeke, 2006). The term "fragmented" refers to the organization of teacher training in Germany that consists of two phases, for which the gap between theory and practice does not seem to be bridged easily (Blömeke, 2009): The period of academic education at university is being

followed by a practical teaching period that is supervised by so-called „Studienseminare“. Following Blömeke (2006), Austrian teacher training could also be described as education-oriented, but is based on one education period that integrates theory and practice courses in the same institution. However, in both countries different practical training phases (Schulpraktische Studien) are obligatory during university education. Another common aspect is that, until today, in the German federal state Baden-Wuerttemberg as well as in Austria *Universities of Education* (Pädagogische Hochschulen) still exists. These institutions emerged from pedagogical academies, and gained university status in Germany about 50 years ago, but only recently (2005) in Austria. It is now definite that in both countries teacher training will be converted to the Bachelor/Master system in 2015, which is expected to result in higher employability and international mobility (Blömeke, 2009). A more distinct evidence orientation of the disciplines involved in teacher training is hereby seen as a key element.

We anticipate that the comparison of the participating countries will afford us the opportunity to draw conclusions regarding system differences between countries and its impact on ERL performance.

The following research questions are driving this project in its current stage:

1. Is it feasible to develop (further) a reliable instrument that assesses ERL through a mixed-format test approach across Austria, Germany, and the United States?
2. Which test structure is most appropriate for assessing ERL? Is the test structure similar across countries? Stated differently, do countries ascribe the same meaning to the construct of ERL? Do items function similarly across countries?

Individuals in different countries may ascribe different meanings to the underlying construct of ERL due to differences in educational experiences among other factors. Moreover, the use of different item formats may also cause differences in the construct representativeness across countries. For example, it may be hypothesized that students in Europe may be more familiar with constructed response tasks, whereas students in the United States are more accustomed to multiple choice items.

3. Is it possible to combine the dispositional and holistic view of observing ERL performance?

The complexity of the underlying construct of ERL may present a challenge for statistical analysis and interpretation of results (a. o. local item dependency, within-item-multi-dimensionality etc.). For example, the debate on the structure of intelligence – a ERL related but empirically distinguishable construct (Schladitz, Groß Ophoff, & Wirtz, in press) offers further suggestions for understanding what it means to measure ERL. At the top of the hierarchy are theories of intelligence, with Spear-

man (1904) at one extreme postulating a single undifferentiated general intelligence and Guilford (1967) and Gardner (2006) at the other end of the spectrum postulating multiple abilities and different independent intelligences. But even though there are certain overlaps with concepts like academic or generic skills (Clanchy & Ballard, 1995), ERL is acquired within and influenced by its reference disciplines, and can be comprehended – at least in part – as domain-specific ability (Fischer et al., 2014; Lea & Street, 2006). Given the uncertainty of the construct, competing hypotheses (hypothesis 1 and hypothesis 2) regarding the underlying test structure were developed and are to be tested. In hypothesis 1 (Bifactor Model, see Figure 3), it is assumed that ERL consists of one dominant factor representing the generic aspect of this competence, and secondary factors standing for discipline-specific aspects in relation to the research cycle. Hypothesis 2 (Within-item IRT Model, see Figure 4) asserts that there is no generic component of the ERL construct, rather a complex test structure is assumed in which some items may measure more than one ability or discipline-specific aspects.

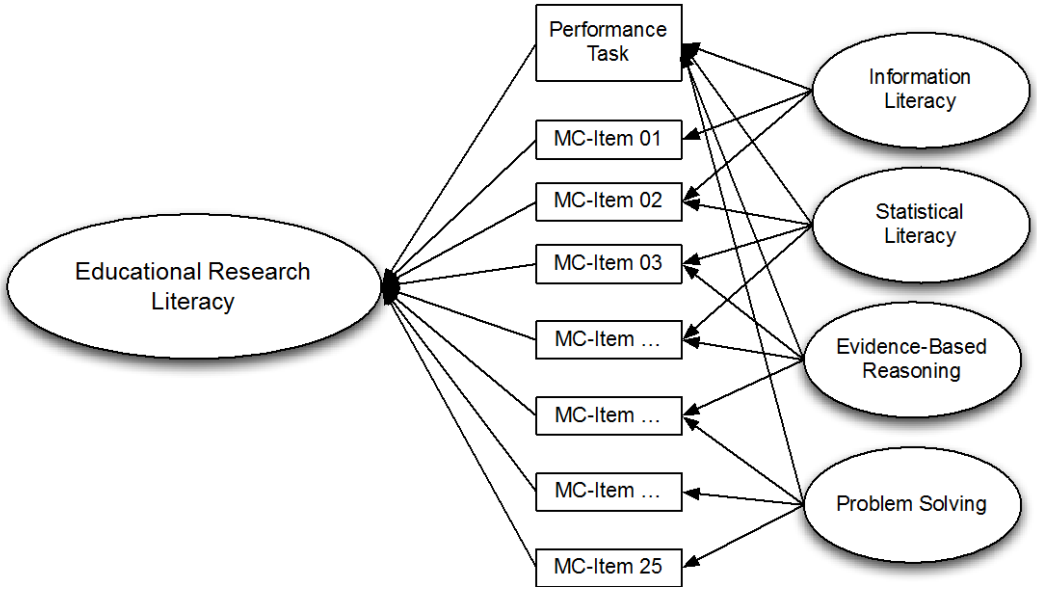

Figure 3: Bifactor Model (Hypothesis 1)

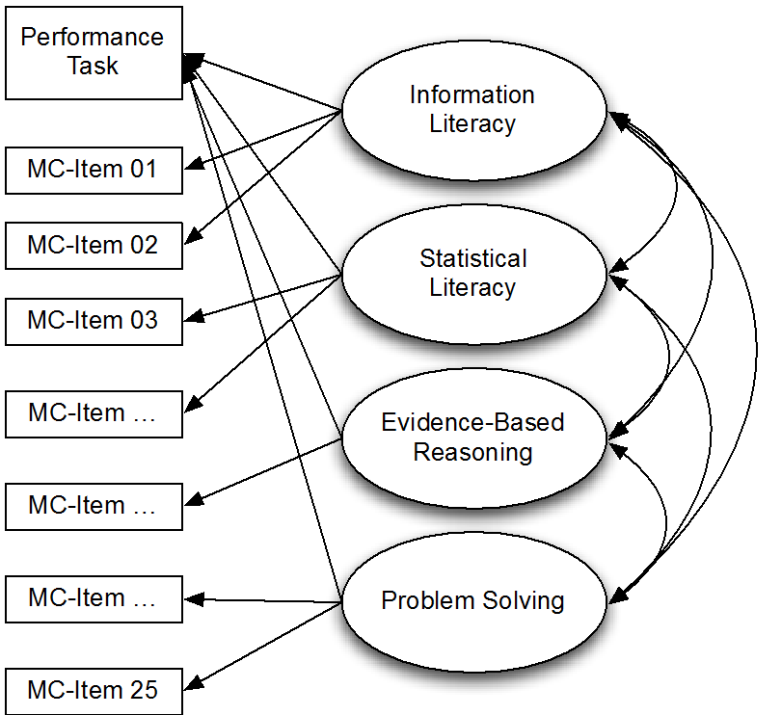

Figure 4: Within-Item Multidimensional Model (Hypothesis 2)

Methodological Background

The bifactor model was introduced by Holzinger and Swineford (1937) and it is typically used within the factor analysis and structural equation modeling communities. The model allows each item response to be explained by both a dominant factor as well as secondary orthogonal factors (Gibbons & Hedeker, 1992). The dominant trait is the factor of interest (i.e. ERL), whereas the secondary traits (i.e. information literacy, statistical literacy, evidence-based reasoning, and problem solving) may be considered as subdomains. In other words, the application of the bifactor model allows for retaining the goal of measuring a general latent construct such as ERL while controlling for the variance that arises due to the assessment of different cognitive skills by different subdomains.

The assumptions of the model include that each item loads on a dominant factor (ERL) in addition to only one of the subdomain factors. In addition, the subdomains are orthogonal to each other and to the dominant factor. For example, for a test that is composed of six items with two subdomains (Problem Solving and Evidence Based Reasoning), the model can be conceptualized in terms of a factor pattern as follows

$$\Lambda = \begin{pmatrix} \lambda_{10} & \lambda_{11} & 0 \\ \lambda_{20} & \lambda_{21} & 0 \\ \lambda_{30} & \lambda_{31} & 0 \\ \lambda_{40} & \lambda_{41} & 0 \\ \lambda_{50} & 0 & \lambda_{52} \\ \lambda_{60} & 0 & \lambda_{62} \end{pmatrix}$$

λ_{ij} represents the loading of item *i* (*i*=1,2,…6) on latent factor *j* (*j*=0,1,2).

In this structure matrix the general domain items will have a nonzero value of the item discriminations or slopes along with clusters of items that belong either to the Problem Solving or Evidence Based Reasoning subdomain. All other item discriminations are zero.

The parameters from factor analytic approaches do not directly correspond to the IRT item parameters. However, it is possible to transform the factor loadings (λ) and threshold values (τ) to obtain the item parameter estimates for the within-item multidimensional structure with uncorrelated dimensions as follows:

$$a_{ik} = \frac{(D)\lambda_{ik}}{\sqrt{1 - \sum \lambda_{ik}^2}}.$$

In the case of the bifactor model, each complex item would have two a_{ik} or λ_{ik} slope parameters. Similarly the item-category threshold parameters can be obtained as follows

$$d_{ik} = \frac{\tau_{ik}}{\sqrt{1 - \sum \lambda_{ik}^2}}$$

From an IRT perspective an examinee's response to each item will be determined by a dominant construct (ERL) and one of the specific subdimensions (information literacy, statistical literacy, evidence-based reasoning, and problem solving). Examinee's responses to SRQs can be modeled by a three-parameter bifactor model

$$\mathrm{P}\left(X_i = 1 \middle| \theta_j, a_i, c_i, d_i\right) = c_i + (1 - c_i)\frac{\exp[D(a_{iERL}\theta_{ERL} + a_{is}\theta_s - d_i)]}{1 + \exp[D(a_{iERL}\theta_{ERL} + a_{is}\theta_s + d_i)]},$$

where θ_{ERL} represents the dominant proficiency, while θ_s (s=1,2,…,*k*) represents one of the *k* subdomains that are orthogonal to each other as well as to the general or dominant proficiency. Additionally, a_{iERL} and a_{is} are item discrimination parameters for the dominant factor (ERL) and one of the *k* subdomains. Lastly, d_i can be conceptualized as a scalar parameter that is related to an overall multidimensional item difficulty as found in the typical MIRT model and *D* (1.7) is the scaling constant.

The scalar d_i from Equation can be calculated as follows

$$d_i = -b_i\sqrt{a_{ERL}^2 + a_{ij}^2}$$

using the b_i parameter and the discrimination parameters from the general dimension and one of the subdimensions (information literacy, statistical literacy, evidence-based reasoning, and problem solving).

Meanwhile, the PTs constructed response items can be modeled as a bifactor graded response model

$$P(X_{ij} \geq 1|\theta_{GLR}, \theta_S) = \frac{1}{1+\exp[-(d_1 - a_{GRL}\theta_{GRL} + a_S\theta_S)]},$$

$$P(X_{ij} \geq K-1|\theta_G, \theta_S) = \frac{1}{1+\exp[-(d_{K-1} - a_{GRL}\theta_{GRL} - a_S\theta_S)]},$$

where $d_1, \ldots, d_{K-1}$ are strictly ordered intercepts that are related to the MIRT item difficulty parameters, a_{GRL} θ_{GRL} are the item discrimination and proficiency estimates for the dominant factor, whereas a_S θ_S reflect the item discrimination and proficiency estimates for the specific factors (information literacy, statistical literacy, evidence-based reasoning, and problem solving).

In within-item multidimensionality or complex structure some items measure more than one ability. These types of models are useful for modeling interactions between different abilities and task demands; thus, the probability of solving an item can be modeled as a function of a combination of different dimensions of abilities. Hence, within-item multidimensional models imply explicit assumptions about the abilities required for the different items, which necessitate strong theoretical assumptions. Models with within-item multidimensionality are particularly interesting for modeling performance in complex tasks that cannot be explained by a single ability dimension for each task (Hartig & Hohler, 2009).

Psychometric Considerations: Overcoming Challenges in Cross-National Comparisons

In order to substantiate test score interpretations, our evaluation of validity will be based on the following sources of evidence of validity:

a) Content validity: an analysis of the relationship between a test's content and the construct it is intended to measure"
b) Construct validity: "the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are made"
c) Criterion validity: "analyses of the relationship of test scores to variables external to the test"

Further, the following measures will be taken to overcome the related challenges:

ad a) Content validity: Evidence of validity based on test content is revealed by the extent to which the material on the test represents an appropriate sampling of skills, knowledge, and understanding of the domain tested. Content validity refers to the extent to which the assessment measures education research literacy. Content area experts from participating countries will examine the content of the test during the test development stage carefully to ensure that the test is in alignment with the intended standards and objectives. Professional judgment by staff and external consultants is an integral part in defining what the test is intended to measure, the breadth of content to be sampled, and specifying the item format and scoring system to be utilized in the process.

Language Unbiasedness

- Bilingual individuals who know the content area translate test items, and scoring rubrics
- Backtranslations
- Testing translations with bilingual students—generalization may be limited because he sample may not be representative of the population we are going to test
- Bilingual translators (Germany, Austria, US) work together (i.e. video conference) to translate items

Topic Relevance: Literature review on differences in curricula

System Differences: Analysis (literature review, document analysis) of

- Educational standards
- HE curricula
- Course specific contents

  in participating countries and HE institutions.

ad b) Construct validity: The following aspects regarding the internal structure of ERL have to be taken into account in the context of cross-national comparisons:

Language Unbiasedness: Test for DIF—flagged items need to be reevaluated for item biasedness

Sample Representativeness

- Ideally we would want to use simple random probability sampling to draw the samples. However, this may not be feasible due to many factors such as socio-economic factors, and motivation. If demographic variables are being collected, statistical procedures can be used to adjust for differences in the samples (case mix control, weighting procedures, propensity

score matching). Another approach could include attaching stakes to the assessment. For example, students should receive feedback about their performance on the assessment. This would facilitate criterion referenced interpretations.

- Sample size requirements: Ideally we would want to sample 200 students in each country to attain a minimum sample size between 100-200 of students in their second or third year of undergraduate studies within the field of educational sciences.

Motivation

- Test taking motivation
- Motivation behind competency
- Assess item response patterns and speediness in case we use computer administration
- Furthermore, we plan on including survey items that ask students about their engagement and motivation
- Expectancy Theory Model
- Examinee Persistence Model (IRT)
- Motivation filtering

Scoring Equivalency

- Statistical analysis to examine measurement equivalence at test and item levels. We plan on implementing G-theory and/ or a multi-facet Rasch model to examine rater effects.
- Another challenge arises due to the use of multiple item formats. More specifically, open-ended tasks or performance tasks may be influenced by differences in score leniency. Scorers, either within a country or across countries, may not score the same student response consistently or equivalently. However, score equivalency across countries is essential if scores are to be compared across countries. Scoring-equivalency analyses will be conducted to examine within- and between country scoring equivalence. A combination of different statistical analyses (G-theory, multi-facet Rasch model) will be implemented to determine the number of raters needed for reliable measures and to explore rater effects. Furthermore, statistical analyses will be conducted to assess measurement equivalence at the test and item levels.

ad c) Criterion validity: A multitrait-multimethod matrix (MTMM) (Campbell, 1959) will be analyzed to facilitate interpretations of construct validity and to examine the relationship of test scores to variables external to the test. Through the use of this methodology, convergent and discriminant validity will be assessed to determine the degree to which concepts that should relate theoretically

are interrelated in reality. Discriminant validity is the degree to which concepts that should not be related theoretically, are, in fact, not interrelated. Construct validity can be claimed if evidence of both convergence and discriminant validity exists.

Anticipated Implications

In summary, we foresee three major insinuations:

1. The aforementioned endeavor can contribute to the discussion of overcoming disagreements due to oversimplified dichotomies in the context of the assessment of competence because it integrates both aspects of competence measurement. Both, the analytic (dispositional) and the holistic approaches to assessment are afflicted by issues of validity and reliability. Following the suggestions from Blömeke et. al. (in press) we attempt to develop an approach which allows to combine the two views in a meaningful way to improve reliability and validity of competence assessment scores.
2. Based on the cross-national design we have the possibility to analyze country and curricular effects.
3. Furthermore we hope to provide an element for the discussion of generic versus domain-specific view of competence.

Project Outlook

It is evident that the aforementioned effort requires several stages, thus leading to short-term and long-term goals. With “short-term” we propose a first project phase (Figure 5) with a duration of approximately 2 years with the following stages: 1) Literature review, 2) development of education-specific case vignettes and coding criteria, 3) preliminary studies to try out the vignettes (thinking aloud method, content analysis of short essays), 4) large sample study (poss. multi-cohort-sequence-design), 5) data entry, criteria-based encoding and statistical analysis, 6) transfer study (pre-service, in-service).

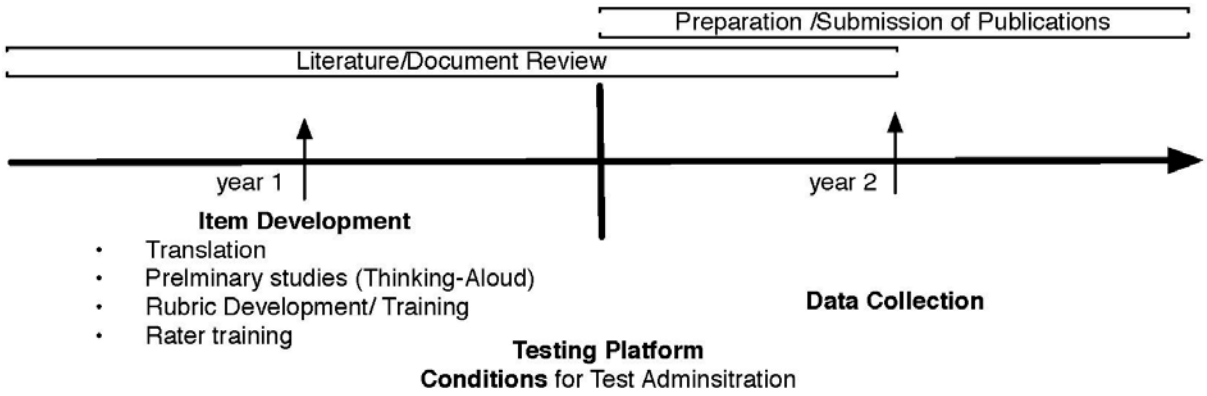

Figure 4: schematic representation of the first project phase

References

Altrichter, H., Brüsemeister, T. & Heinrich, M. (2005). Merkmale und Fragen einer Governance-Reform am Beispiel des österreichischen Schulwesens. *Östereichische Zeitschrift für Soziologie, 30*(4), 6-28.

Astleitner, H. & Herber, H.-J. (2008). Methoden und Probleme bei der Entwicklung instruktionspsychologischer Ansätze zur Unterrichtsgestaltung. In F. Hofmann, C. Schreiner & J. Thonhauser (Eds.), *Qualitative und quantitative Aspekte. Zu ihrer Komplementarität in der erziehungswissenschaftlichen Forschung* (pp. 359–370). Münster: Waxmann.

Baer, M., Dörr, G., Fraefel, U., Kocher, M., Küster, O., Larcher, S., . . . Wyss, C. (2007). Werden angehende Lehrpersonen durch das Studium kompetenter? Kompetenzaufbau und Standarderreichung in der berufswissenschaftlichen Ausbildung an drei Pädagogischen Hochschulen in der Schweiz und in Deutschland. [Competencies and Standards in Teacher Education in Switzerland and Germany]. *Unterrichtswissenschaft, 35*(1), 15-47.

Barchfeld, P. & Sodian, B. (2009). Differentiating Theories from Evidence: The Development of Argument Evaluation Abilities in Adolescence and Early Adulthood. *Informal Logic, 29*(4), 396–416.

Ben-Zvi, D. & Garfield, B. (Eds.). (2004). *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. New York, NY: Kluwer Academic Publishers.

Blixrud, J. C. (2003). Project SAILS: Standardized Assessment of Information Literacy Skills. *ARL Bimonthly Report.* Retrieved 14.02.2012, from http://www.arl.org/bm~doc/sails.pdf.

Blömeke, S. (2006). Struktur der Lehrerausbildung im internationalen Vergleich. Ergebnisse einer Untersuchung zu acht Ländern. *Zeitschrift für Pädagogik, 52*(3), 393-416.

Blömeke, S. (2009). Lehrerausbildung in Deutschland. *PäD-Forum: unterrichten erziehen, 37/28*(1), 5-8.

Blömeke, S., Gustafsson, J.-E. & Shavelson, R. (in press). Beyond dichotomies: Viewing competence as a continuum. Zeitschrift für Psychologie *Zeitschrift für Psychologie*.

Blömeke, S. & Zlatkin-Troitschanskaja, O. (2013). Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor: Ziele, theoretischer Rahmen, Design und Herausforderungen des BMBF-Forschungsprogramms KoKoHs. *KoKoHs Working Papers, 1.* Retrieved 30.09., 2013, from http://www.kompetenzen-im-hochschulsektor.de/Dateien/KoKoHs_WP1_Bloemeke_Zlatkin-Troitschanskaia_2013_.pdf.

Borg, S. (2010). Language Teacher Research Engagement. *Language Teaching, 43*(4), 391–429.

Bos, W., Postlethwaite, T. N. & Gebauer, M. M. (2010). Potenziale, Grenzen und Perspektiven internationaler Schulleistungsforschung. In R. Tippelt & B. Schmidt (Eds.), *Handbuch Bildungsforschung* (pp. 275-295). Wiesbaden: VS Verlag für Sozialwissenschaften.

Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O. & Wilson, M. (2010). The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning. *Educational Assessment, 15*(3/4), 123-141.

Brown, N. J. S., Nagashima, S. O., Fu, A., Timms, M. & Wilson, M. (2010). A Framework for Analyzing Scientific Reasoning in Assessments. *Educational Assessment, 15*(3/4), 142-174.

Bundesassistentenkonferenz. (1970). *Forschendes Lernen – wissenschaftliches Prüfen* (Neuaufl. nach der 2. ed.). Bielefeld: UVW Universitäts Verlag Webler.

Campbell, D. T. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Catts, R. & Lau, J. (2008). Towards Information Literacy Indicators: Conceptual framework paper. Paris.

Ditton, H. (2010). Evaluation und Qualitätssicherung. In R. Tippelt & B. Schmidt (Eds.), *Handbuch Bildungsforschung* (3. durchgesehene Auflage ed., pp. 607-623). Wiesbaden: VS Verlag für Sozialwissenschaften.

Dunn, D. S., Halonen, J. S. & Smith, R. A. (Eds.). (2008). *Teaching critical thinking in psychology: A handbook of best practices*. Malden, Mass.: Wiley-Blackwell.

Flores-Mateo, G. & Argimon, J. M. (2007). Evidence based practice in postgraduate healthcare education: a systematic review. *BMC health services research, 7*, 119.

Gibbons, R. D. & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*(3), 423-436.

Greiff, S., Kretzschmar, A. & Leutner, D. (2014). Problemlösen in der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie, 28*(4), 161-166.

Groß Ophoff, J., Schladitz, S., Leuders, J., Leuders, T. & Wirtz, M. (submitted). Assessing the development of Educational Research Literacy. The effect of courses on research methods in studies of Educational Science. *Peabody Journal of Education (Special Issue)*.

Groß Ophoff, J., Schladitz, S., Lohrmann, K. & Wirtz, M. (2014). Evidenzorientierung in bildungswissenschaftlichen Studiengängen: Entwicklung eines Strukturmodells zur Forschungskompetenz. In W. Bos, K. Drossel & R. Strietholt (Eds.), *Empirische Bildungsforschung und evidenzbasierte Reformen im Bildungswesen* (pp. 251-276). Münster: Waxmann.

Groß Ophoff, J., Schladitz, S. & Wirtz, M. (März 2014). *Struktur von Research Literacy in den Bildungswissenschaften.* Paper presented at the 24. Kongress der deutschen Gesellschaft für Erziehungswissenschaft (DGfE), Humboldt-Universität Berlin.

Groth, R. E. (2007). Toward a Conceptualization of Statistical Knowledge for Teaching. *Journal for Research in Mathematics Education, 38*(5), 427–437.

Grundmann, R. & Stehr, N. (2012). *The Power of Scientific Knowledge: From Research to Public Policy*: Cambridge University Press.

Hammersley, M. (2004). Some Questions about Evidence-based Practice in Education. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (Vol. 2004, pp. 133–149). Maidenhead: Open University Press.

Harris, L. R. & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice, 16*(3), 365-381.

Hartig, J. & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies In Educational Evaluation, 35*(2-3), 57–63.

Hoffmann-Göttig, J., Eschmann, W. & Daumen, C. (2005). Und sie bewegt sich doch... Vom Umgang mit den Ergebnissen externer Evaluation aus der Sicht von Bildungspolitik und Schulaufsicht. In G. Becker, A. Bremerich-Vos, M. Demmer, K. Maag Merki, B. Priebe, K. Schwippert, L. Stäudel & K.-J. Tillmann (Eds.), *Standards. Unterrichten zwischen Kompetenzen, zentralen Prüfungen und Vergleichsarbeiten* (pp. 32-36). Seelze: Friedrich Verlag.

Holzinger, K. J. & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*(1), 41-54.

Homeyer, E. (2008). *Informationskompetenz an Grundschulen. Probleme und Perspektiven für Schulen und Lehrer*. Berlin: Simon Verlag für Bibliothekswesen.

Klein, S., Benjamin, R., Shavelson, R. & Bolus, R. (2007). The Collegiate Learning Assessment. Facts and fantasies. *Evaluation Review, 31*(5), 415-439.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., . . . Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards* (Vol. 1). Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF).

Kuhn, D. & Franklin, S. (2007). *The Second Decade: What Develops (and How)*: John Wiley & Sons, Inc.

Kuhn, D., Iordanou, K., Pease, M. & Wirkala, C. (2010). Beyond control of variables: What needs to develop to achieve skilled scientific thinking?: Scientific reasoning -- Where are we now? *Cognitive Development, 23*(4), 435–451.

Kultusministerkonferenz. (2004). Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16.12.2004. Retrieved 15.09.2010, from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf.

Kultusministerkonferenz. (2005). Qualifikationsrahmen für deutsche Hochschulabschlüsse. Retrieved 09.01.2011, from http://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/Beschluesse_Veroeffentlichungen/Hochschule_Wissenschaft/BS_050421_Qualifikationsrahmen_AS_Ka.pdf.

Novick, L. R. & Bassok, M. (2005). Problem solving. In K. Holyoak & B. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 321-349): Cambridge Univ Pr.

Oakleaf, M. (2008). Dangers and Opportunities: A Conceptual Map of Information Literacy Assessment Approaches. *portal: Libraries and the Academy, 8*(3), 233–253.

Patry, J.-L. (2000). Schulunterricht ist komplex - Kann da Theorie noch praktisch sein? . *Salzburger Beiträge zur Erziehungswissenschaft*, 43-59.

Phye, G. D. (2001). Problem-solving instruction and problem-solving transfer: The correspondence issue. *Journal of Educational Psychology, 93*(3), 571-578.

Schladitz, S., Groß Ophoff, J. & Wirtz, M. (in press). Forschungskompetenz und allgemeine Intelligenz. Ein Beitrag zur Konstruktvalidierung von Kompetenztests in bildungswissenschaftlichen Studiengängen. *Zeitschrift für Pädagogik*.

Schunn, C. D. & Anderson, J. R. (2001). Acquiring Expertise in Science: Explorations of What, When, and How. In K. Crowley, C. D. Schunn & T. Okada (Eds.), *Designing for science* (pp. 83–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Shaneyfelt, T., Baum, K. D., Bell, D., Feldstein, D., Houston, T. K., Kaatz, S., . . . Green, M. (2006). Instruments for Evaluating Education in Evidence-Based Practice. *JAMA: The Journal of the American Medical Association, 296*(9), 1116-1127.

Shank, G. & Brown, L. (2007). *Exploring Educational Research Literacy*. New York, NY: Routledge.

University of Education Freiburg. (2009). Studien- und Prüfungsordnung der Pädagogischen Hochschule Freiburg für die Bachelor-Studiengänge Pädagogik der frühen Kindheit. Amtliche Bekanntmachung Nr. 17 vom 03.11.2009. Retrieved 07. Oktober 2014, from https://www.ph-freiburg.de/fileadmin/dateien/zentral/webdoks/hochschule/bekanntmachungen/ab_0917_spo_bachelor_pfk.pdf.

Watson, J. M. & Callingham, R. A. (2003). Statistical Literacy: A Complex Hierarchical Construct. *Statistics Education Research Journal, 2*(2), 3–46.

Weiss, C. H. (1998). Have We Learned Anything New About the Use of Evaluation? *American Journal of Evaluation, 19*(1), 21-33.

Wissenschaftsrat. (2000). Empfehlungen zur Einführung neuer Studienstrukturen und -abschlüsse (Bakkalaureus/Bachelor – Magister/Master) in Deutschland. Retrieved 04.01.2011, from http://www.wissenschaftsrat.de/download/archiv/4418-00.pdf.

Zlatkin-Troitschanskaia, O. & Kuhn, C. (2010). Messung akademisch vermittelter Fertigkeiten und Kenntnisse von Studierenden bzw. Hochschulabsolventen – Analyse zum Forschungsstand. Mainz.

Theoretical Consideration

1) What competence/competency(ies) is/are the focus of your work? How do you define the boundaries of what is included within the scope of your work n competence/competency/(ies)? What is not included?

Our group is focused on assessing competencies that enable individuals to purposefully access, comprehend, reflect on scientific evidence, and apply the resulting conclusions in complex situations. Our construct definition does not entail for an individual to design and construct research independently.

The question of interest is whether individuals are able to read and interpret scientific evidence. The aforementioned competencies are deemed essential for successful contribution to the 21st century workforce. Within the context of educational research these types of skills may be broadly summarized as Educational Research Literacy (ERL). The four latent dimensions underlying ERL are the following: Information Literacy, Statistical Literacy, Evidence-Based Reasoning and Problem Solving.

One assumption in regards to these competencies is that they are developed over time by the interplay between discipline-oriented education, general education, and the general abilities that students have developed and bring to higher education. Furthermore, it may be assumed that complex competencies such as ERL are inherently intertwined in the task (e.g. performance task) and the response demands. In contrast a construct oriented approach may conceptualize these competencies as a multidimensional construct by taking pieces apart. More specifically, individual test items (e.g selected response questions), subtests or tests are developed to tap each component. Lastly, the components are put back together again to provide a total score. The construct oriented approach assumes that an accurate picture of the whole can be provided by stitching components together. However we question whether this approach neglects the fact that the whole is usually much greater than the sum of its parts.

2) Where do you situate your work with respect to the “continuum” laid out by Bloemeke, Gustafsson, and Shavelson (p.8)?

Given the complexity of the underlying construct we are interested in examining whether it is feasible to combine the dispositional or holistic approaches to assess ERL through a mixed-format test (e.g. performance tasks and selected response questions). One part of the assessment may consist of a performance task that is structured as a document library integrating a real life task scenario. The second part consists of selected response questions. Both parts of the assessment cover the four theoretically assumed dimensions Information Literacy, Statistical Literacy, Evidence-Based Reasoning and Problem Solving and thus may reflect a combination of the holistic and dispositional approaches.

3) Given your position on Blömeke et al.’s continuum, what argument would you like to make about the interpretation of your test/measurement strategy?

Measuring ERL may present a challenge for statistical analysis and interpretation of results (e.g. local item dependency, within-item-multi-dimensionality). Although research has demonstrated that there are certain overlaps with concepts like academic or generic skills, ERL is acquired within and influenced by its reference disciplines, and can be comprehended – at least in part – as domain-specific ability. Given the uncertainty of the construct, we are interested in examining two competing hy-

potheses. For example, one hypothesis is that ERL consists of one dominant factor representing the generic aspect of this competence, and secondary factors standing for discipline-specific aspects in relation to the research cycle. Another hypothesis to be tested asserts that there is no generic component of the ERL construct, rather a complex test structure is assumed in which some items may measure more than one ability or discipline-specific aspects.

a) What evidence would most convincingly support this argument (ignoring resource constraints)?

In order to substantiate test score interpretations, our evaluation of validity will be based on the following sources of evidence of validity:

a. Content validity: "an analysis of the relationship between a test's content and the construct it is intended to measure"

b. Construct validity: "the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are made"

c. Criterion validity: "analyses of the relationship of test scores to variables external to the test"

b) What evidence is most crucial to support this argument (i.e., what evidence is required to make a convincing validity argument)?

Given that validity does not pertain to the instrument itself but rather to the score interpretations drawn from the assessment we believe that all aspects of validity evidence (e.g. evidence based on test content, response processes, and the internal and external structure of the assessment) are important to foster a sound scientific foundation for the intended score interpretations.

c) Taking into account the methods and findings from the Förster et al. paper, what special validity concerns might exist as you either develop a multi-country test or move from single- to multi-country uses of an existing test?

From a psychometric perspective, measuring competencies within an international framework poses challenges that pertain to test development, scoring, and the validity of score interpretations. The following measures will be taken to overcome the related challenges with a validity framework:

Content validity: Evidence of validity based on test content is revealed by the extent to which the material on the test represents an appropriate sampling of skills, knowledge, and understanding of the domain tested. Content validity refers to the extent to which the assessment measures education research literacy. Content area experts from participating countries will examine the content of the test during the test development stage carefully to ensure that the test is in alignment with the intended standards and objectives. Professional judgment by staff and external consultants is an integral part in defining what the test is intended to measure, the breadth of content to be sampled, and specifying the item format and scoring system to be utilized in the process.

Language Unbiasedness

- Bilingual individuals who know the content area translate test items, and scoring rubrics
- Backtranslations
- Testing translations with bilingual students—generalization may be limited because he sample may not be representative of the population we are going to test
- Bilingual translators (Germany, Austria, US) work together (i.e. video conference) to translate items

Topic Relevance: Literature review on differences in curricula

System Differences: Analysis (literature review, document analysis) of

- Educational standards
- HE curricula
- Course specific contents

in participating countries and HE institutions.

Construct validity: The following aspects regarding the internal structure of ERL have to be taken into account in the context of cross-national comparisons:

Language Unbiasedness: Test for DIF—flagged items need to be reevaluated for item biasedness

Sample Representativeness

- Simple random probability sampling unlikely to attain
- Collect demographic variables that allow for statistical adjustments in the samples (e.g. case mix control, weighting procedures, propensity score matching)
- Attach stakes to the assessment (e.g. provide students with feedback about their performance to facilitate criterion referenced interpretations)

- Sample size requirements: Sample minimum of 200 students in each country to attain a minimum sample size between 100-200 of students in their second or third year of undergraduate studies within the field of educational sciences

Motivation

- Test taking motivation
- Motivation behind competency
- Assess item response patterns and speediness in case we use computer administration
- Furthermore, we plan on including survey items that ask students about their engagement and motivation
- Expectancy Theory Model
- Examinee Persistence Model (IRT)
- Motivation filtering

Scoring Equivalency

- Statistical analysis to examine measurement equivalence at test and item levels.
- Examine rater effects for performance task such as scoring equivalence within and between countries (e.g. G-theory and/ or a multi-facet Rasch model)

Criterion validity: A multitrait-multimethod matrix (MTMM) will be analyzed to facilitate interpretations of construct validity and to examine the relationship of test scores to variables external to the test. Through the use of this methodology, convergent and discriminant validity will be assessed to determine the degree to which concepts that should relate theoretically are interrelated in reality. Discriminant validity is the degree to which concepts that should not be related theoretically, are, in fact, not interrelated. Construct validity can be claimed if evidence of both convergence and discriminant validity exists.

4 Personal Reflection in the Autumn Academy

…of Svenja Hammer and Kara Mitchell Viesca

As our paper already illustrates, we have gained great value from participating in the Autumn Academy. Without the Autumn Academy, none of the collaborations that are already underway and are planned would be possible as we would not even be in contact. When the opportunity for participating in the Autumn Academy came out, Svenja spent a great deal of time reaching out to prominent researchers around the globe seeking a suitable junior researcher to collaborate with. Through Tamara Lucas, Svenja and Kara were introduced and our collaborations began right away. Once we were accepted to the Autumn Academy, we created a symposium proposal for AERA to bring together international research perspectives on the education of teachers to work with multilingual learners. Once our symposium was accepted, we began a conversation with an editor at Routledge regarding a book prospectus on the same topic. And because of the opportunity to spend a week together in Berlin collaborating at the Autumn Academy, we have been able to substantially expand our collaborative plans and even co-construct a long-term vision for collaboration around both teaching and research.

We owe KoKoHs and the opportunities provided through the Autumn Academy a great deal in terms of what we have gained both personally and professionally. As you can see from our plans for 2015, we have already started down a meaningful and important collaborative research path that we sincerely believe will positively impact teaching and learning for multilingual learners and their teachers across both the US and Germany.

…of Marc Berges, Jonathan Good, Melanie Margaritis, Leigh Graves Wolf and Aman Yadav

This section deals with some personal reflections and experiences of the KoKoHS Autumn Academy 2014 in Berlin, Germany. Our project group thanks the organizer and the teams for the nice conference, the opportunities we got during the week and the fun we had with all the other participants.

"The KoKoHS Autumn Academy was a nice experience for me and a great opportunity to meet other researchers from different countries and from different research areas. The conference was well organized, and we had the great opportunity to talk with experts about or topics. It was also very nice to meet our American friends Leigh, Jon, and Aman, who we only met because of the conference call. I highly recommend the Autumn Academy for other young researcher because it was a nice experience and a good way to connect researchers."

Melanie Margaritis

"I am sincerely grateful for the KoKoHS experience. The conference was a tremendous opportunity to collaborate with Jon, Melanie, Marc and Aman. Oftentimes we are very caught up in our day to day work and rarely have the luxury to spend concentrated time working on research. This conference gave us a chance to focus on starting up our research agenda and allowed us to connect to others who are doing the same."

Leigh Wolf

"I was honored to have been invited to the KoKoHS conference, especially as a graduate student in the early stages of my career. It was enlightening to discover how international scholars bring their own vocabulary, methods, and priorities to their research, yet we can all benefit from working together. The conference provided a depth of interaction I would not have had through only email and videoconferences. I would recommend this opportunity to any interested scholar."

Jonathan Good

…of Christina Haberfellner, Jana Groß-Ophoff and Raffaela Wolf

The following personal reflection summarizes the overall intentions and central issues of the conference and concludes with a critical reflection of the workshop.

Overviews of the conference's intentions and central issues

The KoKoHs Autumn Academy basically consisted of two parts:

- The first part (October 6th to 7th, 2014) included a two day introduction into the topic of competence measurement in higher education and was open to all KoKoHs researchers. The two keynotes with plenary discussion ("Challenges of international research on competencies" by Edward Wiley and "How to write an international paper/project proposal - from a reviewer's perspective" by Alicia Alonzo) were followed by two presentations of international research projects on competencies ("Using Psychosocial Skill Typologies to Identify Student Needs and Target Early Service Delivery in Community College Samples" by Samuel Rikoon; "Developing Next- Generation Student Learning

Outcomes Assessments for United States Higher Education Institutions" by Katrina Roohr) and a workshop on “Theories and concepts of research on competencies” by Alicia Alonzo and Edward Wiley.

- The second part (October 8th to 10th, 2014) was only accessible for the members of the four project groups and consisted of workshop phases and presentations. The members of the project groups worked on their individual projects and could rely on the expertise and support of the coaches if needed.

Our core group consisted of three people (alphabetical order):

1. Jana Groß-Ophoff, University of Education, Freiburg, Germany
2. Christina Haberfellner, University of Education Salzburg; Austrian Federal Institute for Educational Research, Innovation & Development of the Austrian School System, Salzburg, Austria
3. Raffaela Wolf, Council for Aid to Education, New York, USA

The purpose of our current project is to develop, refine, implement, and validate a mixed-format (i.e. PTs and SRQs) instrument that measures educational research literacy (ERL) within an international assessment framework. Participating countries may include the United States and traditionally more education-oriented states like Austria and Germany. Currently we use the following working title to describe our project: “Evidence-Based Reasoning in Higher Education: Development and Validation of a Mixed-Format Test Approach to Assess Educational Research Literacy”.

The following research questions are driving this project in its current stage:

1. Is it feasible to develop (further) a reliable instrument that assesses ERL through a mixed-format test approach across Austria, Germany and the United States?
2. Which test structure is most appropriate for assessing ERL? Is the test structure similar across countries? Stated differently, do countries ascribe the same meaning to the construct of ERL? Do items function similarly across countries?
3. Is it possible to combine the dispositional and holistic view of observing ERL performance?

Critical reflection

In order to meet the requirements of an international conference on competence measurement in higher education, the main focus lay on the challenges of cross-national comparisons. When realizing

such research projects, researchers have to assume that construct -competency – is comparable across contexts (countries. Operationally, this means that the scores across the countries must trace back to the same construct and assessments must be both reliable and valid (1) in and of themselves and (2) across educational institutions, countries, cultures, and disciplines. Six specific issues of which to remain mindful when approaching research associated with comparative international measurement were defined:

1. Language Unbiasedness
2. Sample Representativeness
3. Motivation
4. Topic Relevance
5. Scoring Equivalency
6. System Differences

While the desire for and the necessity of cross-national research projects in higher education remains, my previous assumption is confirmed that, compared to the United States, Europe (especially Austria) has to catch up regarding theoretical knowledge and practical experiences in this field. Therefore, cooperative events like this conference should sustain and be well-promoted. Further exchange of knowledge is necessary to create a basis for cooperation on equal terms. What is more, the long workshop phases of the second part of the conference were intensive and useful. However, more soundly based coachings would have been essential to increase the output of these phases. Further funding possibilities for the research projects were not discussed, which complicates continued work or might even make it impossible.

Summarizing, the KoKoHs Junior Faculty Research Conference (Autumn Academy) in Berlin, entitled ”Modelling and Measuring Competencies in Higher Education“, was a great opportunity to get to know other research groups who also work in cross-national projects, in order to discuss specific problems and to establish contacts in the scientific community. This is- from my point of view- of great importance, especially for young researchers.

…of Katrina Roohr and Samual Rikoon

We were honored to have been asked to participate in the 2014 Autumn Academy, and thank Drs. Zlatkin-Troitschanskaia and Pant for their generous invitation. The conference was highly informative, and provided us with the welcome opportunity to learn in depth about the work of the KoKoHs initiative and its many fascinating projects. In addition, the meeting provided an excellent forum for the exchange of ideas among researchers across many different stages of their careers surrounding

issues of how to rigorously conceptualize, operationalize, and assess competencies in higher education.

Our primary role in the conference was to serve as “Peer Counselors,” which resulted in productive meetings with each of the four project-focused groups in attendance. These groups were comprised of other young researchers at institutions across the U.S. and Western Europe, with a focus on conducting international comparisons of higher education competency models in varied disciplines. In addition, we met with Dr. Zlatkin-Troitschanskaia and one of her graduate students to discuss the “WiwiKom” project seeking to assess student business and economics expertise in higher education. This meeting provided both a helpful overview of the project and the opportunity for us to provide consultation on the some of the measurement strategies and technical models under consideration.

In addition to the above, Katrina delivered a presentation on the theoretical and practical development of competencies to be incorporated in the HEIghten assessments (a suite of next-generation student learning outcomes assessments in the U.S.), and Sam presented on work to build and field test the Workforce Readiness Strength Assessment and Training System (WRSATS).

Everyone we met at the conference was both thoughtful and friendly, and we expect to maintain the relationships established in Berlin through future collaboration on a number of fronts. We will reconnect with Dr. Zlatkin-Troitschanskaia and her team at the upcoming 2015 AERA conference in Chicago, Illinois, discuss the potential utility of large-scale ETS studies to facilitate international comparisons of student learning, and hopefully also remain in touch with one or more of the KoKoHs project teams to explore opportunities for future collaborative research.

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.