Portfolio of KoKoHs Assessments

Test Instruments for Modeling and Measuring Domain-specific and Generic Competencies of Higher Education Students and Graduates

KoKoHs

SPONSORED BY THE

Federal Ministry
of Education
and Research

Editorial

Portfolio of KoKoHs Assessments: Test Instruments for Modeling and Measuring Domain-specific and Generic Competencies of Higher Education Students and Graduates

Today's dynamically developing knowledge society requires university graduates to acquire domain-specific knowledge and interdisciplinary skills as well as transfer and apply them to social and professional challenges. To meet these far-reaching demands of the 21st century, it is indispensable that university practice is oriented towards fostering competence acquisition. The acquisition of domain-specific and generic competencies is a central teaching-and-learning objective in higher education. To design study programs, courses and examinations with competencies in mind and based on consistent measurement of students' learning progress, practitioners require objective, reliable, and valid assessments of the competencies acquired in higher education.

To this end, the national research initiative KoKoHs, Modeling and Measuring Competencies in Higher Education was established by the German Federal Ministry of Education and Research (BMBF). In two funding phases from 2011 to 2019, theoretical-conceptual competence models and corresponding test instruments and assessment procedures were developed and comprehensively validated. In some 40 collaborative large-scale projects, test instruments were developed or adapted to assess academic competencies in higher education in a valid and reliable manner. The targeted competencies included both domain-specific competencies in various disciplines (economics and social sciences, STEM, teacher training, psychology, educational sciences, medicine) and generic competencies (e.g., self-regulation, academic text comprehension). The models and instruments developed enable practitioners to validly assess the level of students' competencies at the beginning of studies as well as their development over the course of studies and in subsequent practical phases (such as in teacher training and medicine). KoKoHs instruments allow for the assessment of the following kinds of constructs: (i) knowledge-based and affective-volitional dispositions as prerequisites for performance in future professional situations; (ii) situation-specific skills and abilities acquired during studies in higher education; (iii) observed performance in realistic situations. For valid assessment using authentic, professional tasks, researchers developed various computer-based assessments (e.g., video-based teaching simulations), which are particularly suited for process diagnostics.

In the KoKoHs projects, more than 50 test instruments of various formats were developed or adapted (including innovative video-, computer- and simulation-based instruments) and were tested throughout Germany on more than 75.000 students at over 320 higher education institutions (as well as abroad in some adaptations). As a detailed overview, this publication provides a portfolio of assessments showcasing most test instruments developed in KoKoHs. For each assessment, key information is summarized at a glance, followed by a more comprehensive listing of all relevant information about the test instruments, including the conceptual background, test setup, validation, and information on its practical use and suitability.

This portfolio of KoKoHs assessments and additional information is intended to facilitate the implementation of the developed and validated test instruments in further research and transfer projects in higher education. For each KoKoHs assessment, various tried and tested as well as prospective use cases are indicated in the portfolio. Apart from test instruments for entry diagnostics, formative assessments are available to effectively support the teaching-and-learning processes and the acquisition of competencies, and summative assessments can serve as valid measures of students' learning outcomes.

Competence assessments can be used to inform the design of teaching-and-learning settings, to achieve an alignment of curriculum, instruction, and assessment, so that teaching-and-learning content, the design of instruction, and examinations are optimally aligned with one another. Competence assessments can also be included alongside other indicators to objectively evaluate degree courses and entire faculties, and contribute to quality improvement and evidence-based management in higher education. The assessment of academic student learning outcomes by means of competence tests can be a building block in the ongoing process of further developing teaching and learning in higher education. This publication contributes to promoting the test instruments to various levels of higher education research, practice and policy to help pave the way to successful transfer.

We would like to thank all test developers from the KoKoHs program for providing the instruments and additional project and test information for the preparation of this publication. We are indebted to the BMBF and in particular to Martina Diegelmann for the conceptual and financial support of this publication.

Olga Zlatkin-Troitschanskaia, Hans Anand Pant, Marie-Theres Nagel, Dimitri Molerov, Corinna Lautenbach, Miriam Toepper
Mainz, Berlin, January 2020

Content

Academic Text Competencies (AkaTex): Scientific Phrasing and Reasoning

SUMMARY

Name: AkaTex Rating Process

Domain: Academic text production, language didactics

Assessed competencies: Academic text competence in scientific phrasing and reasoning

Target group: Bachelor students at the beginning of their studies

Test type: Evaluation of a discussion paper (miniature format of a seminar paper)

Modality: Computer-based

Duration: Writing time: 2 weeks; average evaluation time (per discussion paper): 20 minutes

Test structure: 13 items total: 7 rating items for the category "scientific phrasing"; 6 rating items for the category "professional content and reasoning"

General test purpose: Assessing the state and development of competence regarding the production of academic texts (scientific phrasing and reasoning)

Application scenarios: Assessing the state and development of first-year students' performance

Not suitable for: Below upper secondary education level

Note for practical use: —

Applied in projects: Academic Text Competencies of First-Year and Advanced Teacher Trainees with Particular Reference to their Pre-Conditions (AkaTex); http://www.uni-siegen.de/phil/akatex/

Contact / Location:
Dr. Lena Decker
University of Siegen
Email: decker@germanistik.uni-siegen.de

GENERAL INFORMATION

Name: AkaTex rating process
Applied in projects: Academic Text Competencies of First-Year and Advanced Teacher Trainees with Particular Reference to their Pre-Conditions (AkaTex); http://www.uni-siegen.de/phil/akatex/
Contact / Location:
Dr. Lena Decker
University of Siegen
Email: decker@germanistik.uni-siegen.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Academic text competence in scientific phrasing and reasoning
Theoretical model: "Academic text competencies in the narrower sense" based on the models of Ossner (2008), Becker-Mrotzek & Schindler (2007) and Schindler & Siebert-Ott (2013)
Test type: Evaluation of a discussion paper (miniature format of a seminar paper)
Modality: Computer-based
Test structure:
Item pool: 13 items; 7 rating items for the category "scientific phrasing", 6 rating items for the category "professional content and reasoning". Participants are asked to write a discussion paper on a topic in the field of language didactics by evaluating several texts from a given task, relating them to each other and developing their own position. The papers are scored by trained raters using the AkaTex rating procedure.

TEST QUALITY CRITERIA

Reliability: Interrater reliability with 2 independent raters* "scientific phrasing" (Cronbach's α = 0. 703–0.902); "professional content and reasoning" (Cronbach's α = 0. 829–0.937)
Validity:
Test content: During expert validation (by experts in writing research), the operationalization of the competence criteria was deemed appropriate and complete.
Response processes: —
Internal test structure: The competence facets "professional content" and "scientific phrasing" show a statistically significant correlation.
Relationships with other competence indicators: —
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: Writing time: 2 weeks; average evaluation time (per discussion paper): 20 minutes
Testing materials: Computer or laptop
Special features: Scorer training required
Practical example: In introductory seminars on language didactics, first-year students wrote a discussion paper in which they evaluated several texts from a given task, relating them to one another and developing their own position. These were assessed by trained raters using the AkaTex rating procedure. The rating scheme was used in an enhanced form for the evaluation of seminar papers and academic theses in teacher training courses at the University of Siegen.

DOMAIN

Tested for: Production of academic texts, language didactics
Suitable for: All domains
Not suitable for: —

TARGET GROUP

Tested for: Bachelor students at the beginning of their studies
Suitable for: Upper secondary education students; advanced bachelor students
Not suitable for: Below upper secondary education level

GENERAL TEST PURPOSE

Tested for: Competence assessment
Suitable for: Assessment of competence development
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: —
Suitable for: Cognitive competencies, grades, educational achievement
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for courses over time. The analyses were carried out at the individual and group level. The sample was collected exclusively at one university.
Suitable for: Individual students over time; course level
Not suitable for: University level, state and national level, international level

APPLICATION SCENARIOS

Tested for: Assessment of the level and development of pre-service teachers' writing skills
Suitable for: Prediction of study success, course comparisons, course improvement
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Assessment of academic performance
Suitable for: Examination grades, final grades
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

Master's degree in connection with the practical semester ("writing in the disciplines" in connection with "teaching writing and assessing writing")

ADDITIONAL INFORMATION

The text type 'discussion paper' is already taught at upper secondary school level as a preliminary form of the academic paper.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Decker, L. & Siebert-Ott, G. (2019). Schreibend an fachlichen Diskursen partizipieren: Ergebnisse einer Interventionsstudie zur Förderung der Textkompetenzen von Lehramtsstudierenden. In: Feilke, H.; Lehnen, K. & Steinseifer, M. (Eds.): *Eristische Literalität. Wissenschaftlich streiten – Wissenschaftlich schreiben*. Münster: Waxmann.
Decker, L. & Siebert-Ott, G. (2018). Wissenschaft als diskursive Praxis. Schreibend an fachlichen Diskursen partizipieren. In: Schmölzer-Eibinger, S.; Bushati, B.; Ebner, C. & Niederdorfer, L. (Eds.): *Wissenschaftliches Schreiben lehren und lernen. Diagnose und Förderung wissenschaftlicher Textkompetenz in Schule und Universität*. Münster: Waxmann, pp. 193–218.
Decker, L. (2016). Wissenschaft als diskursive Praxis. Schreibend an fachlichen Diskursen partizipieren. In: Becker-Mrotzek, M.; Jost, J. & Pohl, T. (eds.), *Reihe Kölner Beiträge zur Sprachdidaktik*, vol. 10. Duisburg: Gilles & Francke.

SECONDARY LITERATURE

Decker, L. & Siebert-Ott, G. (2019). Sprachsensible Bildungsräume gestalten. Eine Professionalisierungsaufgabe in der Lehrer/-innenbildung. In: Dirks, U. & Siebold, K. (Eds.): *Sprachsensible DaF-/DaZ-/DaM-Bildungsräume: Sprech- & Textformen im Fokus*. Münster: Waxmann.
Decker, L.; Oehme, V. (2018). Inhalte angeben – sachtext- und literaturbezogen. In: *Der Deutschunterricht*, 3/2018, pp. 44–55.

Action-related Competence of Mathematics Teachers (Math-AC)

SUMMARY

Name: Action-related Competence of Mathematics Teachers (Math-AC)

Domain: Teacher training in mathematics

Assessed competencies: Action-related competence (AC) of mathematics teachers

Target group: (Pre-service) mathematics teachers at (upper) secondary level (student teachers in mathematics (Bachelor, Master), trainee teachers and mathematics teachers); analogue tests were also developed by the authors for use in primary and elementary education.

Test type: Performance test; video-based tasks, open response format

Modality: Computer- and video-based; audio recorded for assessing action-related competence (AC) with speed component

Duration: 30 minutes

Test structure: 9 tasks; subject areas: algebra (6 items) and analysis (3 items) for secondary education. One video per item with a prototypical teaching scenario (30-60 seconds) including a short context description, to which test participants must react immediately. The answers are saved as audio assessments and evaluated by trained experts using a coding manual (from 0 to 2).

General test purpose: Assessing the level and development of competence (with other indicators of professional teaching competence: tests of domain-specific content knowledge (CK), pedagogical content knowledge (PCK); see also the tests: Math-RC; ARKOM)

Application scenarios: Comparisons of the level of competence between the training phases of teacher education (bachelor, master, practical training phase, teaching service); comparisons between pre-service and in-service teachers in the subjects of mathematics and/or business and economics

Not suitable for: Non-German-speaking international level; initial diagnostics; first-year students

Note for practical use: The test administrator manual includes organizational and technical information to ensure standardized test administration; a coding manual is provided for objective scoring; scorer training is required.

Applied in projects: Assessing Subject-specific Competencies in Teacher Education in Mathematics and Business and Economics – a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi); https://www.elmawi.de

Contact / Location:
Dr. Colin Jeschke
IPN – Leibniz Institute for Science and Mathematics Education Kiel
Email: jeschke@leibniz-ipn.de

GENERAL INFORMATION

Name: Mathematics Teachers' Action-related Competence (Math-AC)
Subject domain: Teacher training mathematics
Applied in projects: Assessing Subject-specific Competencies in Teacher Education in Mathematics and Economics – a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi); https://www.elmawi.de
Contact / Location:
Dr. Colin Jeschke
IPN – Leibniz Institute for Science and Mathematics Education Kiel
Email: jeschke@leibniz-ipn.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Mathematics teachers' action-related competence
Theory model: According to the competence structure model with the two facets AC and RC (see the Math-RC test) based on Lindmeier (2011) and Kuhn (2014)
Test type: Performance test; video-based tasks with open-response format
Modality: Computer- and video-based; audio recorded for AC assessment
Test structure:
Item pool: 9 items; subject areas: algebra (6 items) and analysis (3 items) for secondary level. One video per item with a prototypical teaching scenario (30-60 seconds) including a short context description, to which test participants react immediately. The answers are saved as audio recordings and scored according to qualitative content and formal criteria using a tested coding manual (from 0 to 2 points).

TEST QUALITY CRITERIA

Reliability: Cronbach's alpha for AC =.6 (9 items)
Validity:
Test content: —
Response processes: —
Internal test structure: The one-dimensional structure of AC is confirmed by CFA (Jeschke et al. 2019).
Relationships with other competence indicators: As assumed, weak yet significant relationships with CK and PCK in mathematics and moderate correlations with RC in mathematics (Jeschke et al. 2019).
Consequences of testing: Theoretical differences in test performance between different education levels (bachelor, master, practical training phase, teaching service) (Jeschke et al. 2019).
Test fairness: Due to the language intensity of the test, a language bias has to be assumed.

PRACTICAL USE

GENERAL INFORMATION

Duration: 30 minutes
Testing materials: Computers or laptops with the test program installed, headsets with microphones
Special features: The test administrator manual includes organizational and technical information to ensure a standardized test administration. A coding manual is provided for objective scoring; scorer training is required.
Practical example: The test was used with groups at all education levels at several institutions nationwide. The test contains an introduction to the use of the software and a context description for each video item. For editing, the video clip (30-60 seconds) is first viewed once. It shows a prototypical teaching situation from the mathematics lessons at secondary level I or II in which a mathematics-specific teaching requirement occurs (e.g. a student asks a technical question, a student makes a typical error). Participants are asked to react directly to the situation under time pressure in natural language (e.g.: "Give an explanation!"; "Give a helpful hint so that the learner can solve the task themselves!") and address the student in the process. The statements are recorded as an audio file. Responses are scored using a coding manual.

DOMAIN

Tested for: Mathematical didactics, business and economics education
Suitable for: Other teaching domains
Not suitable for: —

TARGET GROUP

Tested for: Pre-service and in-service teachers of mathematics at upper secondary level (student teachers in mathematics [bachelor, master], trainee teachers and mathematics teachers)
Suitable for: Pre-service and in-service teachers in mathematics at secondary education level, possibly as a short scale (corresponding tests were also developed by the authors for the primary and elementary level)
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessment of the level of competence in the two facets AC and RC (see the Math-RC test) and comparisons of the level of competence between the training phases of teacher education (bachelor, master, practical training phase, teaching service); comparisons between pre-service and in-service teachers in the subjects of mathematics and/or business and economics
Suitable for: Assessment of competence development in all training phases; comparisons with other school subjects
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Video-based tests for assessing AC and RC in business and economics for domain comparisons; tests for assessing PCK and CK in business and economics and mathematics; tests for assessing generic competencies (e.g. intelligence, situational awareness, ambiguity tolerance) for analyses of domain specificity (Jeschke et al. 2019)
Suitable for: Further tests for assessing subject-related or generic competencies
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at the individual level, group or course level, university level, international level (German-speaking countries).
Suitable for: State and national level
Not suitable for: International level (non-German-speaking countries)

APPLICATION SCENARIOS

Tested for: Comparisons of competence levels between training phases in the teacher education (bachelor, master, practical training phase, teaching service); comparisons between business and economics and mathematics (for students, trainees, teachers)
Suitable for: Teaching-learning tool for estimations of individual learning needs; assessment of competence development in all phases of education and professional development training; comparisons with other school subjects
Not suitable for: Entry diagnostics

PERMITTED CONSEQUENCES

Tested for: Assessment of competence in university, practical training phase, teaching service; implications for the improvement of course offers on the tested competencies
Suitable for: Assessment of competence development over the various phases of teacher education; statistical prediction of competence level at a later point in time
Not suitable for: Entry diagnostics

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test can be expanded for further content dimensions (e.g. in the form of additional tasks). The aim is to further develop the survey software to keep it up to date in the long term. The test for assessing Math-AC can be requested from the authors at any time. Test development is documented in detail in Lindmeier (2011).
In the future, the test will also be used in the project "Promotion of Subject-specific Competencies of Pre-service Teachers in Mathematics and Business & Economics Using Video-based ELMaWi Tools (ELMaWi-Transfer)".

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation. If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Jeschke, C.; Kuhn, C.; Lindmeier, A.; Zlatkin-Troitschanskaia, O.; Saas, H. & Heinze, A. (2019). Performance assessment to investigate the domain-specificity of instructional skills among pre-service and in-service teachers of mathematics and economics. *British Journal of Educational Psychology*, 1–13. https://doi.org/10.1111/bjep.12277

Jeschke, C.; Lindmeier, A. & Heinze, A. (2018). Wie fachspezifisch sind „fachspezifische Kompetenzen" von Lehrkräften? In: Kortenkamp, U. & Kuzle, A. (Eds.): *Beiträge zum Mathematikunterricht 2017 WTM*, pp. 1119–1122.

Jeschke, C.; Lindmeier, A. & Heinze, A. (2017). What do mathematics pre-service teachers lack for mastering instructional demands? In: Kaur, B.; Ho, W. K.; Toh, T. L. & Choy, B. H. (Eds.): *Proceedings of the 41st Conference of the International Group for the Psychology of Mathematics Education*, 3, pp. 33–40. Singapore: PME.

SECONDARY LITERATURE

Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung* (Empirische Berufsbildungs- und Hochschulforschung, vol. 2). Landau: Verlag Empirische Pädagogik.

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

Advanced Design Project (ADP) Skills in Mechanical Engineering

SUMMARY

Name: Advanced Design Project (ADP) Skills

Domain: Chemistry, process engineering

Assessed competencies: Competencies of students in interdisciplinary design projects

Target group: Master students, graduates

Test type: Ratings and open-response questions

Modality: Paper-pencil test

Duration: 60 minutes

Test structure: 36 items; 2 partial scales with 16 respectively 20 items with 3 respectively 4 facets

General test purpose: One-time testing of social skills for design assignments

Application scenarios: Teaching-learning tool

Not suitable for: Admission to studies, choice and recommendation of study domain, evaluation of events, longitudinal studies

Note for practical use: The test is based on the model for Team Member Effectiveness by Loughry et al. (2007) and is adapted for a specific assessment situation. Some items assess content-related competencies needed for the overall technical design task.

Applied in projects: Modelling of Competencies of Students of Mechanical Engineering (MokoMasch)

Contact / Location:
Oliver Watteler
GESIS Leibniz Institute for the Social Sciences
Email: oliver.watteler@gesis.org

GENERAL INFORMATION

Name: Advanced Design Project (ADP) Skills
Applied in projects: Modelling of Competencies of Students of Mechanical Engineering (MokoMasch)
Contact / Location:
Oliver Watteler
GESIS Leibniz Institute for the Social Sciences
Email: oliver.watteler@gesis.org

THE INSTRUMENT

Domain: Chemistry, process engineering
Target group: Master students, graduates
Assessed competencies: Competencies of students in interdisciplinary design projects

CONTENT AND STRUCTURE

Theoretical model: Three areas of competence are distinguished in the design of plants (e.g. bio-ethanol): "general" professional competencies, social competencies and technical-specific competencies. "General competencies" include abstract scientific skills for integrating technical solutions or alternatives. In the case of "social competencies and team effectiveness", particular attention is given to the ability to work in a team (Loughry, Ohland & Moore, 2007). "Specific competencies" include task-related aspects.
Test type: Ratings and open-response questions
Modality: Paper-pencil test
Test structure:
Item pool: 36 items: 2 partial scales with 3 facets for part A (16 items) and 4 facets for part B (20 items). Part A can be used to assess general competence facets: team competencies (7 items), solution evaluation skills (3 items) and the identification of complex plant design problems (6 items). Part B serves to assess (domain-specific) technical and scientific competencies: general technical understanding (in the context of the plant problem to be solved; 5 items), process knowledge (5 items), fundamental domain-specific knowledge (5 items) and advanced domain-specific knowledge (5 items).

TEST QUALITY CRITERIA

Reliability: Cronbach's α of the scales: .61-.63
Validity:
Test content: —
Response processes: —
Internal test structure: For the English test (The Comprehensive Assessment of Team Member Development (CATME), see Loughry, Ohland & Moore, 2007).

Relationships with other competence indicators: Expectation-compliant relations to social competencies (ISK, Kanning, 2009) and Ten-Item-Personality-Inventory (TIPI, Gosling, Rentfrow, & Swann, 2003; German version: Muck, Hell and Gosling, 2007)
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 60 minutes
Testing materials: Templates of the test items
Special features: The test is based on the model for Team Member Effectiveness by Loughry et al. (2007) and is adapted for a specific assessment situation. Some items assess content-related competencies needed for the overall technical design task.
Practical example: The test was used as an accompanying instrument in the Advanced Design Project at 3 institutions (N=63) – a university course in which a professional situation is simulated that requires students to construct a plant, e.g. for the production of bio-ethanol. The subtasks of the overall task need to be distributed within the team and completed. For successful completion, the team members need to integrate their individual performance (professional performance). The task can only be successfully completed through teamwork. With regard to the predictive power or explanatory power of the instrument, there is a relation between performance and outcome in the team, especially with the assessed "soft skills" such as competencies for the acquisition and application of knowledge in the group situation.
A competence-oriented teaching-learning instrument is being developed as part of the course.

DOMAIN

Tested for: Chemistry, process engineering
Suitable for: Engineering sciences, plant design
Not suitable for: —

TARGET GROUP

Tested for: Master students
Suitable for: Graduates
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing social competencies for design assignments
Suitable for: One-time testing; repeated testing
Not suitable for: Longitudinal studies

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Grades, scientific (short) tests, personality factors
Suitable for: Validation (self-report-related) of subjective measures (soft skills in group situations)
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for individuals (according to a validated measurement model structure).
Suitable for: Group level (if the measuring instrument is adapted because it is currently available in the Round Robin design)
Not suitable for: University level, state and national level, international level

APPLICATION SCENARIOS

Tested for: Teaching-learning tool
Suitable for: —
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Examination grades or assessment of study or work performance
Suitable for: —
Not suitable for: Admission to studies, choice of studies, recommendation, evaluation of events

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test was used once and was not used again after the project had ended. Connections to established tests for assessing social competence and personality factors were investigated. The test will be made available for future design projects as an assessment tool (Anders, Pinkelman, Hampe, & Kelava, 2014). Use will be organized via contact to the project leader. The application is free of charge. However, any items must be carefully adapted to the application context.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Anders, B.; Pinkelman, R. J.; Hampe, M. J. & Kelava, A. (2015). Development, assessment, and comparison of social, technical, and general (professional) competencies in a university engineering advanced design project-a case study. *Competence in Higher Education and the Working Environment. National and International Approaches for Assessing Engineering Competence*, pp. 217–238.

SECONDARY LITERATURE

Kanning, U. P. (2009). *Inventar Sozialer Kompetenzen (ISK).* Göttingen: Hogrefe.
Loughry, M. L.; Ohland, M. W. & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement, 67*, p. 505.
Muck, P. M.; Hell, B. & Gosling, S. D. (2007). Construct validation of a short Five-Factor Model instrument: A self-peer study on the German adaptation of the Ten-item Personality Inventory (TIPI-G). *European Journal of Personality Assessment, 23*(3), pp. 166–175.

Argument Structure Test (AST)

SUMMARY

Name: Argument Structure Test (AST)

Domain: Humanities and social sciences, sciences requiring empirical or evidence-based reasoning

Assessed competencies: Epistemic-systematic reading skills: Recognition and assignment of functional argument components

Target group: Bachelor's and master's students at the beginning and over the course of their studies

Test type: Performance test; assignment tasks

Modality: Computer-based

Duration: 30 minutes

Test structure: 40 items; correct recognition and assignment of 5 argument components in 8 argument examples according to Toulmin (1985). Difficulty variation due to different number and placement of the argument components

General test purpose: Assessing the collective competence level and development of groups

Application scenarios: Group diagnostics, evaluation of teaching-learning and training measures, prediction of study success

Not suitable for: Individual diagnostics, international comparisons, candidate selection, non-native speakers

Note for practical use: Control for mother tongue or language competence

Applied in projects: Students' competence in dealing with original scientific literature (KOSWO), Assessment and training of Scientific Literacy (AstraLite), https://www.uni-regensburg.de/psychologie-paedagogik-sport/paedagogik-1/projekte/index.html

Contact / Location:
Prof. Dr. Tobias Richter
Institute for Psychology, University of Würzburg
Website: https://go.uniwue.de/edpsych

Dr. Sebastian Schmid
Institute for Pedagogy, University of Regensburg
Website: https://www.uni-regensburg.de/psychologie-paedagogik-sport/paedagogik-1/team/dr-sebastian-schmid/index.html

GENERAL INFORMATION

Name: Argument Structure Test (AST)
Applied in projects: Students' competence in dealing with original scientific literature (KOSWO)
Assessment and training of Scientific Literacy (AstraLite)
(https://www.uni-regensburg.de/psychologie-paedagogik-sport/paedagogik-1/projekte/index.html)
Contact / Location:
Prof. Dr. Tobias Richter, Institute for Psychology, University of Würzburg
Website: https://go.uniwue.de/edpsych
Dr. Sebastian Schmid, Institute for Pedagogy, University of Regensburg
Website: https://www.uni-regensburg.de/psychologie-paedagogik-sport/paedagogik-1/team/dr-sebastian-schmid/index.html

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Epistemic-systematic reading skills: Recognition and assignment of functional argument components
Theoretical model: Toulmin's argumentation model (1958)
Test type: Performance test; assignment tasks
Modality: Computer-based
Test structure:
Item pool: 8 argument examples, correct assignment of the 5 argument components according to Toulmin (1958); 40 items available, which differ according to: Number of components (3 = simple argument vs. all 5 = complex argument), placement of components (argument begins with assertion vs. reason).

TEST QUALITY CRITERIA

Reliability: Cronbach's α = .76 (40 items)
Validity:
Test content: The AST measures competencies in reading original scientific literature. These competencies are relevant to curricula and required in almost all academic study programs.
Response processes: Eye-tracking analyses (N=60) show that more complex arguments take longer to read and are read more frequently.
Internal test structure: In line with theory, simple arguments were solved correctly more often than complex arguments; arguments which begin with assertions are solved correctly more often than arguments which begin with justifications.

Relationships with other competence indicators: Moderate positive relationships with verbal intelligence (r = .40) and school leaving grade (r = .17)
Consequences of testing: Test results can indicate the need for teaching and learning epistemic-systematic reading skills.
Test fairness: No gender differences. Language differences: The tasks contain a lot of academic language. The language level may affect the test result for non-native speakers.

PRACTICAL USE

GENERAL INFORMATION
Duration: 30 minutes
Testing materials: Computer or laptop; test control program
Special features: Test result is the sum value of correctly solved items.
Practical example: The AST was used at two German universities in three different parallel forms during five assessments, which showed a satisfactory reliability and validity of the test instrument. In addition, the AST was used in two studies with a further test that assesses the plausibility of scientific arguments before and after a teaching intervention to improve epistemic-systematic reading skills to test the effectiveness of the intervention (von der Mühlen, Richter, Schmid & Berthold, 2018).

DOMAIN
Tested for: Humanities and social sciences
Suitable for: Sciences requiring empirical or evidence-based reasoning
Not suitable for: Purely hermeneutic humanities

TARGET GROUP
Tested for: Students at the beginning of their studies
Suitable for: Students throughout the course of their studies
Not suitable for: Researchers

GENERAL TEST PURPOSE
Tested for: Research and evaluation purposes
Suitable for: Assessing students' competence level and development throughout their studies
Not suitable for: General reasoning competence

USE WITH OTHER TESTS AND QUESTIONNAIRES
Tested for: Tests to evaluate the plausibility of arguments, verbal intelligence, socio-demographic data (e.g. mother tongue, gender), grades
Suitable for: Other methods of competence measurement
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY
Tested for: Results are statistically meaningful for the comparison of students, also over the course of time.
Suitable for: Comparison of courses, universities, state and national level
Not suitable for: Individual diagnostics, international comparisons

APPLICATION SCENARIOS
Tested for: Assessing students' current level of competence; evaluation of teaching-learning methods (pre-post measurement)
Suitable for: Prediction of study success
Not suitable for: Candidate selection

PERMITTED CONSEQUENCES
Tested for: Assessing the need for (additional) teaching and learning opportunities
Suitable for: Promotion of students' handling of original scientific papers; improvement of scientific literacy
Not suitable for: Study admission, study progression recommendations

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE
To improve the reading skills of students, the AST is to be used as part of a training block at the beginning of the course and later established in the university curriculum.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS
Münchow, H.; Richter, T.; von der Mühlen, S.; Schmid, S.; Bruns, K. & Berthold, K. (2020). Verstehen von Argumenten in wissenschaftlichen Texten: Reliabilität und Validität des Argumentstrukturtests (AST). *Diagnostica*.

SECONDARY LITERATURE
von der Mühlen, S.; Richter, T.; Schmid, S. & Berthold, K. (2018). How to improve argumentation comprehension in university students: Experimental test of a training approach. *Instructional Science, 47*(2), 215-237. https://doi.org/10.1007/s11251-018-9471-3
Münchow, H.; Richter, T. & Schmid, S. (2020). What does it take to deal with academic literature? Epistemic components of scientific literacy. In: Zlatkin-Troitschanskaia, O.; Pant, H. A.; Toepper, M. & Lautenbach, C. (Eds.): *Student learning in German higher education: Innovative measurement approaches and research results.* Wiesbaden: Springer.

Berlin Test for the Determination of Fluid and Crystalline Intelligence (BEFKI)

SUMMARY

Name: Berlin Test for the Determination of Fluid and Crystalline Intelligence (BEFKI)

Domain: Non-domain-specific

Assessed competencies: Fluid and crystalline intelligence

Target group: Students in grades 3-13, adults of different age groups including higher education students

Test type: Performance test

Modality: Paper-pencil test; computer-based

Duration: 70 minutes (when using all subtests)

Test structure: 112 items on two superordinate dimensions, with 3 sub-dimensions each: 1. *fluid intelligence* (48 items on verbal, numerical and figural reasoning) and 2. *crystalline intelligence* (64 items on knowledge of the natural sciences, humanities and social sciences)

General test purpose: Assessing cognitive abilities (fluid and crystalline intelligence) one-time or over time

Application scenarios: Basic research; control variables in examinations in the school context; diagnostics in school trajectory and career counseling; prediction of study success

Not suitable for: —

Note for practical use: Developed for use with students; successfully used with higher education students or adults. Individual or group tests are possible.

Applied in projects: Used to measure basic cognitive skills in studies on linguistic, mathematical and scientific competencies based on the educational standards of the German Conference of the Ministers of Education and Cultural Affairs (KMK) (Institute for Educational Quality Improvement (IQB)) pilot studies, standardization studies and educational trend studies); PISA (Programme for International Student Assessment) studies since 2012; scientific support for the Hector Children's Academies (Hector Institute for Empirical Educational Research / Leibniz Institute for Research and Information in Education (DIPF Frankfurt)); studies with students (Validation of an entrance examination in the study domain of business and economics (WiWiSET); https://www.wiwi-kompetenz.de/wiwiset-2016-2019/; Modeling and measuring competencies in business and economics among students and graduates (WiWiKom); https://www.wiwi-kompetenz.de/wiwikom-ii-2015-2019/)

Contact / Location:
Dr. Stefan Schipolowski
Humboldt University of Berlin
Institute for Educational Quality Improvement (IQB)
Email: stefan.schipolowski@iqb.hu-berlin.de

GENERAL INFORMATION

Name: Berlin Test for the Determination of Fluid and Crystalline Intelligence (BEFKI)
Applied in projects: Used to measure basic cognitive skills in studies on linguistic, mathematical and scientific competencies based on the educational standards of the German Conference of the Ministers of Education and Cultural Affairs (KMK) (Institute for Educational Quality Improvement (IQB)) pilot studies, standardization studies and educational trend studies); PISA (Programme for International Student Assessment) studies since 2012; scientific support for the Hector Children's Academies (Hector Institute for Empirical Educational Research / Leibniz Institute for Research and Information in Education (DIPF Frankfurt)); studies with students (Validation of an entrance examination in the study domain of business and economics (WiWiSET); https://www.wiwi-kompetenz.de/wiwiset-2016-2019/; Modeling and measuring competencies in business and economics among students and graduates (WiWiKom); https://www.wiwi-kompetenz.de/wiwikom-ii-2015-2019/)
Contact / Location:
Dr. Stefan Schipolowski
Humboldt University of Berlin
Institute for Educational Quality Improvement (IQB)
Email: stefan.schipolowski@iqb.hu-berlin.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Fluid intelligence (g) with verbal, numerical and figural reasoning; crystalline intelligence (gc) in the domains natural sciences, humanities and social sciences
Theoretical model: Intelligence structure theories by Carroll (1993; three-stratum theory) Horn and Cattell (Horn & Noll, 1997; gf-gc theory), and McGrew (2009; Cattell–Horn–Carroll theory of cognitive abilities—CHC theory)
Test type: Performance Test
Modality: Paper-pencil test; computer-based

Test structure:
112 items of two superordinate dimensions with 3 sub-dimensions each: figural intelligence/gf (48 items) with verbal, numerical and figural reasoning (16 items each); crystalline intelligence/gc (64 items) with scientific (6 content areas, 24 items), humanities (5 content areas, 20 items) and social science knowledge (5 content areas, 20 items)

TEST QUALITY CRITERIA

Reliability: Psychometric requirements for difficulty, selectivity and reliability are met. Internal consistency for adolescents in upper secondary education (test form BEFKI 11+): For the overall scale gf: α = .79; for the overall scale gc: α = .81. Internal consistency for students in psychology (test form BEFKI 11+): For the overall scale gf: α = .79; for the overall scale gc: α = .76

Validity:

Test content: The item types and contents were selected based on relevant literature and preliminary tests.

Response processes: —

Internal test structure: The testing of measurement models for the internal structure based on the standardization data of children and adolescents shows theoretically compliant results. For students, empirical studies prove the one-dimensionality of the scale for figural reasoning.

Relationships with other competence indicators: Discriminant and convergent validity were tested with other indicators of cognitive ability: Cognitive ability test (KFT 4-12+ R; Heller & Perleth, 2000); vocabulary test (WST; Schmidt & Metzler, 1992); competence tests based on the KMK educational standards; school grades; study grades; mean differences between tested groups correspond to expectations (e.g. different grades, type of school attended, gender).

Consequences of testing: Prediction of individual skill levels in career guidance or studies

Test fairness: Confirmatory factor analyses for measuring invariance analyses of gender and type of school attended largely confirm the assumption of strict measuring invariance and thus of a fair test with regard to the investigated intelligence constructs.

PRACTICAL USE

GENERAL INFORMATION

Duration: 70 minutes (when using all subtests)

Testing materials: Test item templates, evaluation key

Special features: Four different test versions for different age groups and two versions of the test booklet with reversed order of items and response options are available. Individual or group testing is possible. The test has been developed and standardized for use with school students; the test version for upper secondary education level has also been successfully used with adults (including university students).

Practical example: Standardization studies were carried out with a total of 18,000 students in grades 3 to 12 in all German federal states; in addition, trial studies were carried out with university students from various courses (including 9,000 students of social and economic sciences, 400 students of psychology and 240 students of chemistry teacher education).

DOMAIN

Tested for: School context; psychology, economics, social sciences, chemistry teacher education

Suitable for: Can be used across disciplines.

Not suitable for: —

TARGET GROUP

Tested for: Grades 3 to 12/13. The test version for the intermediate and advanced levels was also successfully used in several studies with adults (including university students).

Suitable for: Adults of different age groups including university students (interdisciplinary)

Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Cognitive ability assessment (fluid and crystalline intelligence) in cross-sectional and longitudinal designs

Suitable for: Cognitive abilities are predictively valid for numerous outcomes such as academic and professional success and are used in study and career guidance.

Not suitable for: Additional test procedures are required to assess further cognitive ability factors (e.g. mental speed); the general knowledge test does not assess domain-specific knowledge of certain subjects.

Use with Other* Test *Instruemts and Questionnaires

Tested for: Used to measure basic cognitive skills in studies on linguistic, mathematical and scientific competencies based on the educational standards of the German Conference of the Ministers of Education and Cultural Affairs (KMK) (Institute for Educational Quality Improvement (IQB)) pilot studies, standardization studies and educational trend studies); PISA (Programme for International Student Assessment) studies since 2012; scientific support for the Hector Children's Academies (Hector Institute for Empirical Educational Research / Leibniz Institute for Research and Information in Education (DIPF Frankfurt)); studies with students (Validation of an entrance examination in the study domain of business and economics (WiWiSET); https://www.wiwi-kompetenz.de/wiwiset-2016-2019/; Modeling and measuring competencies in business and economics among students and graduates (WiWiKom); https://www.wiwi-kompetenz.de/wiwikom-ii-2015-2019/)

Suitable for: No restrictions

Not suitable for: No restrictions

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Individual and group level; also over time
Suitable for: Versions in other languages (English, Arabic, Polish, Russian) are being tested.
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Basic research; reasoning as a control variable in studies in the school context (PISA studies and IQB education trends); prediction of study success
Suitable for: As control variable for numerous analyses; also predictive for various outcomes such as academic and professional success; use in study and career guidance
Not suitable for: Tests on general cognitive skills should be supplemented with tests on more specific constructs (e.g. domain-specific tests) depending on the purpose of use.

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Recommendations for school, study and career choice; selection of applicants (in each case in combination with further information and results of other test procedures)
Not suitable for: Differentiated assessment of subject-specific competencies (e.g. academic achievement)

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The instrument is being tested in other languages (English, Arabic, Polish, Russian). Some of the tests are published by Hogrefe publishing house. Test versions that have not yet been published may be used in consultation with the test developers.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Wilhelm, O.; Schroeders, U. & Schipolowski, S. (2014). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe (BEFKI 8-10).* Göttingen: Hogrefe.
Schipolowski, S.; Wilhelm, O. & Schroeders, U. (2020). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 11. und 12. Jahrgangsstufe (BEFKI 11+).* Göttingen: Hogrefe.
Schroeders, U.; Schipolowski, S. & Wilhelm, O. (2015). Age-related changes in the mean and covariance structure of fluid and crystalized intelligence in childhood and adolescence. *Intelligence, 48*, pp. 15–29. https://doi.org/10.1016/j.intell.2014.10.006

SECONDARY LITERATURE

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge University Press.
Horn, J. L. & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In: Flanagan, D. P.; Genshaft, J. L. & Harrison, P. L. (Eds.), *Contemporary intellectual assessment: Theories, tests and issues,* pp. 53–91. New York: Guilford.
McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, pp. 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Biology Teachers' Competencies in Explaining and Assessing Experiments

SUMMARY

Name: Test of teaching competence in experimental biology

Domain: Biology education

Assessed competencies: Analysis, planning and assessment skills of biology teachers at lower secondary level (analyzing and planning lessons, assessing students' experiments)

Target group: Biology education students

Test type: Performance test; open-response format

Modality: Paper-pencil-test

Duration: 120-140 minutes (lesson analysis and planning skills: 60-70 minutes; assessment competence: 60-70 minutes)

Test structure: Teaching skills experiments in biology class – analyzing lessons: 6 items; planning lessons: 6 items; expert knowledge in in-class experiments: 6 items. Assessment competence: assessing hypothesis formation, planning experiments and evaluating data: 7 open-response tasks, 22 assessment items. Example tasks from various teaching-learning topics: seed germination (5th/6th grade), photosynthesis (7th/8th grade), enzymology (9th/10th grade).

General test purpose: Assessing the level of competence or competence development in pre-service teachers' analysis, planning and assessment competencies

Application scenarios: Basic research, evaluation of courses, assessment of study or work performance

Not suitable for: School subjects not involving scientific experiments

Note for practical use: A coding manual is provided for scoring the open-response tasks. 0–1 or 0-2 points are awarded for each item.

Applied in projects: Conveying and assessing competencies in experimental lessons: Modelling, validation and development of a test (EXMO)

Contact / Location:
Prof. Dr. Marcus Hammann
University of Münster
Email: hammann.m@uni-muenster.de
Website: https://www.uni-muenster.de/Biologie.Didaktik/

Prof. Dr. Susanne Bögeholz
University of Göttingen
Email: sboegeh@gwdg.de
Website: https://www.uni-goettingen.de/de/137982.html

Prof. Dr. Claus H. Carstensen
University of Bamberg
Email: claus.carstensen@uni-bamberg.de
Website: https://www.uni-bamberg.de/psymethodenbf/mitarbeiterinnen-und-mitarbeiter/prof-dr-claus-h-carstensen/

GENERAL INFORMATION

Name: Test of teaching competence in experimental biology
Applied in projects: Conveying and assessing competencies in experimental lessons: Modelling, validation and development of a test (EXMO)

Contact / Location:
Prof. Dr. Marcus Hammann
University of Münster
Email: hammann.m@uni-muenster.de
Website: https://www.uni-muenster.de/Biologie.Didaktik/

Prof. Dr. Susanne Bögeholz
University of Göttingen
Email: sboegeh@gwdg.de
Website: https://www.uni-goettingen.de/de/137982.html

Prof. Dr. Claus H. Carstensen
University of Bamberg
Email: claus.carstensen@uni-bamberg.de
Website: https://www.uni-bamberg.de/psymethodenbf/mitarbeiterinnen-und-mitarbeiter/prof-dr-claus-h-carstensen/

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Lesson analysis, lesson planning and assessment competencies for biology lessons at lower secondary level (analyzing and planning lessons, assessing students' experiments)
Theoretical model: The competence model "Teaching and Assessment Competencies for Biology Class Experiments" is based on the three phases of Klahr's Scientific Discovery as Dual Search (SDDS) model (2000; Search in the Hypothesis Space, Search in the Experimentation Space and Data Analysis; Hammann, 2004). Hammann, Phan & Bayrhuber (2006) describe students' deficits in these three phases when conducting experiments. For teacher training, Hasse et al. (2014) include student competencies from these three experimental phases to develop tests for assessing teachers' instructional and evaluation competencies.
Test type: Performance test; open-response format
Modality: Paper-pencil test
Test structure:
Item pool: The test "Teaching Skills in Experimental Biology" consists of three written vignettes and two partial scales, "Analyzing experimental biology lessons" (6 items) and "Planning experimental biology lessons" (6 items) as well as the scale "Expertise in class experiments" (6 items). The test for "Assessment competence" consists of seven tasks on three biology class topics for grades 5-10 (three for "seed germination", two for "photosynthesis", two for "enzymology") as well as the partial competencies "Assessment skills regarding students' hypothesis formation" (6 items), "Assessment skills regarding experiment planning" (8 items) and "Data evaluation" (8 items). The 22 items are scored on a scale of 0-1 or 0-2 points.

TEST QUALITY CRITERIA

Reliability: The EAP/PV reliability coefficients are 0.69 for the scale "Planning experimental biology lessons", 0.73 for the scale "Analyzing experimental biology lessons" and 0.72 for the scale "Expertise in class experiments". For the entire test on assessment skills, the inter-rater agreement is 0.82 for Cohen's Kappa, 0.65 for WLE and 0.63 for EAP/PV for reliability. The quality of the scoring by four scorers was controlled by a quadruple scoring of 10% of the data. Krippendorff's Alpha lies between 0.61 and 0.84 for 18 of the 22 items.
Validity:
Test content: Curricular validity was examined by means of an expert survey on "assessment competencies" surveying biology education researchers and biology education instructors actively involved in teacher training; topics of accordingly high curricular relevance were selected, including "enzymology", "photosynthesis" and "seed germination".
Response processes: The items on the scales "Planning experimental biology lessons", "Analyzing experimental biology lessons" and "Assessment competence" cover cognitive aspects according to the three phases of Klahr's SDDS model (2000): "formulating and assessing scientific hypotheses", "planning and assessing scientific experiments" and " assessing the evaluation of scientific experiments", which were assessed in studies using the think-aloud method
Internal test structure: According to IRT model comparisons, a two-dimensional model with the two dimensions "Expertise in class experiments" and "Planning and analyzing experimental biology lessons" (summarized) has the best fit. So far, "Assessment competence" has been modeled as one dimension. Multidimensional modeling, which focuses on both teaching and assessment competencies, is still lacking.
Relationships with other competence indicators: Relations to external variables are largely compliant with theory: The following short scales were used as external variables: a short scale of the KoWadis test (assesses knowledge about experimentation, Hartmann et al. 2015) and a scale from the Berlin intelligence structure test (Jäger et al. 1997) for assessing the ability of verbal reasoning. Evaluation competence, relevant facets of self-efficacy expectations for teaching biology (Mahler, 2014) and diagnostic competence (Dübbelde, 2013) were assessed.
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 120-140 minutes (lesson analysis and planning skills: 60-70 minutes; assessment competence: 60-70 minutes)
Testing materials: Test booklets, coding manual
Special features: A coding manual is provided for the evaluation of the open-response tasks. Each item is scored on a scale of 0-1 or 0-2 points.
Practical example: The scales for teaching competence were tested with 525 biology education students and the scale for assessment competence with 500 biology education students at 19 German universities. The participants studied to become teachers at Gymnasium schools (lower to upper secondary education), Realschule schools (lower to medium secondary education) or other types of school (primary school, vocational school, lower secondary school, special education). The findings from EXMO were incorporated into the design of the teacher training course for biology. Specifically, teaching in profession-oriented bachelor's and master's of education was improved with a supplementary design variant for the biology-didactic research internship with a focus on promoting experimental skills as a subject-didactic task. At another university, parts of various compulsory courses in the Bachelor and Master of Education programs were geared towards teaching skills in experimental biology lessons.

DOMAIN

Tested for: Biology education
Suitable for: Teaching chemistry, physics, geography (physical)
Not suitable for: School subjects not involving scientific experiments

TARGET GROUP

Tested for: Biology education students
Suitable for: Pre-service trainee teachers in their practical training phase
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessment of competence level
Suitable for: Assessment of competence development after sensitivity to change is validated
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Domain-specific knowledge test with experiments as an additional variable, scales for assessing self-efficacy expectations for teaching biology, scales for assessing diagnostic competencies
Suitable for: The test can be combined with other instruments.
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Group and course level (teacher training for the subject biology)
Suitable for: Institutional level
Not suitable for: Individual diagnostics, state and national level, international level

APPLICATION SCENARIOS

Tested for: The findings from the EXMO project were incorporated into the design of the teacher training course for biology in Bachelor and Master of Education programs at two universities.
Suitable for: Evaluation of courses and assessment of study or work performance, evaluation of modules or module components in teacher training which are dedicated to the corresponding competencies.
Not suitable for: Applications are limited according to the test focuses on specific aspects of teacher training. Use with assessments of additional dimensions is recommended.

PERMITTED CONSEQUENCES

Tested for: Competence assessment
Suitable for: Evaluation of modules or module components
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The tests should be examined for sensitivity to change to be used for future evaluations of modules or module components in teacher training.
The tests can be requested from the test developers and used for the validated test purposes.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Bögeholz, S.*; Joachim, C.*; Hasse, S. & Hammann, M. (2016). Kompetenzen von (angehenden) Biologielehrkräften zur Beurteilung von Experimentierkompetenzen. *Unterrichtswissenschaft, 44*(1), pp. 40–54. [*lead authors]
Hasse, S.; Joachim, C.; Bögeholz, S. & Hammann, M. (2014). Assessing Teaching and Assessment Competences of Biology Teacher Trainees: Lessons from Item Development. *International Journal of Education in Mathematics, Science and Technology, 2*(3), pp. 191–205.

SECONDARY LITERATURE

Hammann, M. (2004). Kompetenzentwicklungsmodelle: Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. *MNU – Der mathematisch-naturwissenschaftliche Unterricht, 57*(4), pp. 196–203.
Hammann, M.; Phan, T. T. M.; Ehmer, M.; & Bayrhuber, H. (2006). Fehlerfrei Experimentieren. *MNU – Der mathematisch-naturwissenschaftliche Unterricht, 59*(5), pp. 292–299.
Klahr, D. (2000). *Exploring Science: The Cognition and Development of Discovery Processes.* Cambridge: MIT Press.

Business and Economics Competence: WiWiKom Test

SUMMARY

Name: WiWiKom test

Domain: Business and economics

Assessed competencies: Business and economics competence; knowledge and understanding in business and economics

Target group: Bachelor students during and at the end of their studies, master students at the beginning of their studies

Test type: Multiple-choice tasks, selection tasks, assignment tasks, sequence tasks

Modality: Paper-pencil test, also available as an online version

Duration: 30-40 minutes, depending on the length of the test booklet

Test setup: 204 items: 60 items on economics, comprising the subscales microeconomics (30 items) and macroeconomics (30 items); 144 items on business comprising the subscales organization (40 items), financing (24 items), marketing (40 items), human resources (20) and manegerial accounting (20 items). Test versions: Long version in test booklet design: 24-30 items per questionnaire and socio-demographic part; single (sub)scales; short version: 36 items

General test purpose: Assessing the level of business and economics competence at any time during the course of studies (one-off testing, e.g. with the short version); assessing the development of competence over several points in time during the course of studies (from the bachelor's program to the beginning of the master's program in business and economics)

Application scenarios: Formative assessment as a teaching-learning tool for students and lecturers; assessment of competence level as a supplement to examinations (also as self-assessment); assessment of strengths and deficits over the course of studies for the purpose of recommending courses and learning support offers; observation of competence development over the course of studies (e.g. annual growth); recognition of knowledge and understanding deficits by lecturers and appropriate design of teaching for the target group; further possible applications after consultation with the project team.

Not suitable for: Study entry and study suitability diagnostics; foreign students whose German proficiency is not advanced enough to understand domain-specific content

Notes for practical use: The long version is used in the booklet design with 43 questionnaire variants (24-30 items each). The scoring of the results therefore requires a statistical approach using IRT.

Applied in projects: Modelling and measurement of business and economic competence among students and university graduates (WiWiKom); https://www.wiwi-kompetenz.de/wiwikom-i-2011-2015/; National Education Panel (NEPS): "University studies and transition to work"; Product- and process-oriented modelling and assessment of competencies for self-regulated learning in the tertiary sector (PRO-SLR); Competence-oriented examination of communicative skills (KomPrü); Swiss Leading House "Teaching and learning processes in the commercial sector" (LINCA); used in further cooperation projects in Japan, Finland, Russia, and South Korea.

Contact / Location:
Prof. Dr. Olga Zlatkin-Troitschanskaia
Johannes Gutenberg University Mainz
Email: lstroitschanskaia@uni-mainz.de

GENERAL INFORMATION

Name: WiWiKom test
Applied in projects: Modelling and measurement of business and economic competence among students and university graduates (WiWiKom); https://www.wiwi-kompetenz.de/wiwikom-i-2011-2015/; National Education Panel (NEPS): "University studies and transition to work"; Product- and process-oriented modelling and assessment of competencies for self-regulated learning in the tertiary sector (PRO-SLR); Competence-oriented examination of communicative skills (KomPrü); Swiss Leading House "Teaching and learning processes in the commercial sector" (LINCA); used in further cooperation projects in Japan, Finland, Russia, and South Korea.
Contact / Location:
Prof. Dr. Olga Zlatkin-Troitschanskaia
Johannes Gutenberg University Mainz
Email: lstroitschanskaia@uni-mainz.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Business and economics competence; knowledge and understanding in business and economics

Theoretical model: The theoretical model (Zlatkin-Troitschanskaia et al. 2014) distinguishes 3 dimensions: a) Types of knowledge: propositional, case-related and strategic knowledge (Shavelson & Ruiz-Primo, 1999); b) Levels of expertise: remembering and understanding, applying and analyzing, creating and judging (taxonomy levels according to Anderson & Krathwohl, 2001, especially for economic contexts according to Walstad et al. 2007). c) Classification according to subject areas: economics: micro- and macroeconomics, business administration: organization, financing, marketing, human resources, internal accounting

Test type: Multiple-choice tasks, selection tasks, assignment tasks, sequence tasks

Modality: Paper-pencil test, also available as an online version

Test setup:

Item pool: 204 items: 60 items on economics were adapted from the fourth version of the US-American Test of Understanding College Economics (TUCE IV) and include the subscales microeconomics (30 items) and macroeconomics (30 items); 144 items on business administration originate from the German adaptation of the Spanish language exam General para el Egreso de la Licenciatura (EGEL) with the subscales organization (40 items), financing (24 items), marketing (40 items), human resources (20) and managerial accounting (20 items)

Test versions: Individual scales: see above; short version: 36 items from all areas; long version with 204 items in a booklet design for multiple group testing: 43 questionnaire variants with 24-30 items each and socio-demographic part

TEST QUALITY CRITERIA

Reliability: Cronbach's alpha on all scales at all measurement points (at the end of the 1st, 2nd, and 3rd academic year in bachelor's programs): 0. 6–0.8

Validity:

Test content: Curricular analysis of study descriptions and module manuals for 96 courses of study at 64 universities (curriculum analyses) as well as online expert surveys among 78 professors and lecturers of business and economics at German universities: The items represent a basic core curriculum of business and economics in bachelor studies.

Response processes: Cognitive interviews with students (N=120) using the think-aloud method during item responding as well as interviews with experts (N=32): During item responding, students verbalized the intended thought processes; item response strategies, e.g. guessing show no statistically significant correlation with successful task solutions.

Internal test structure: Statistical analyses (confirmatory factor analysis, IRT modelling) confirmed the theoretically assumed multi-dimensionality of the content; the disciplines business and economics are independent sub-dimensions

Relationships with other competence indicators: Relations to other variables in line with expectations: e.g. business-related prior knowledge (learning opportunities taken in the school subject of economics, completion of commercial vocational training) show a positive correlation with the test results; positive correlations with the general cognitive performance of the test takers as well (school leaving grade; intelligence) were also demonstrated.

Consequences of testing: Differences in test performance between the cohorts of students of different semesters were in line with expectations in cross-sectional studies: higher performance with increasing semester; average competence developments at group level in line with expectations in longitudinal studies, but also cohort and composition effects were found; various developments at individual level.

Test fairness: According to measurement invariance analyses (CFA, IRT), almost all items work equally well for different student groups, with the exception of a few items that have a medium-sized effect between subgroups: Effects of gender and migration background are evident: female students (M(female)=12.03, SD=4.196, n=3438; M(male)=14.28, SD=4.289, n=4125; t(7561)=-22.954, p<0.0001) and students who are non-native speakers of German (M(other mother tongue other than German)=13.34, SD=4.335, n=7294; M(German as mother tongue)=9.05, SD=3.936, n=228; t(244,527)=-16.379, p<0.0001) performed worse on average. It is not yet sufficiently clear whether the results are due to test characteristics or actual differences in competence.

PRACTICAL USE

GENERAL INFORMATION

Duration: 40 minutes

Testing materials: Test books, pencils, calculators (for the items of the partial scales financing and accounting)

Special features: Feedback: An online Wiki of the Humboldt University of Berlin is provided as score interpretation aid. When completing the test, students create an individual code and can anonymously view their results online at: https://mars.wiwi.hu-berlin.de/mediawiki/feedback/index.php/Hauptseite

Practical example: Items from the WiWiKom test were used at national level in the National Education Panel (NEPS) and in the PRO-SRL project (Product- and process-oriented modelling and assessment of competencies for self-regulated learning in the tertiary sector) in Austria, and in the KomPrü – KomFäh project (Competence-oriented test of communicative skills). In addition, the test was used in Switzerland in the Leading House LINCA (Teaching and learning processes in the commercial sector) as well as in an ongoing Finnish project at the University of Turku and a Russian project at the National Research University Higher School of Economics. The entire test and partial scales have been adapted and tested for other countries: Switzerland (German), Austria (German), Japan (Japanese), Finland (Finnish), South Korea (Korean), Russia (Russian) and the countries of origin of the tests (USA (English) and Mexico (Spanish)).

DOMAIN

Tested for: Business and economics; social sciences

Suitable for: All courses of study with business and economic contents

Not suitable for: —

TARGET GROUP

Tested for: Bachelor students during and at the end of their studies, master students at the beginning of their studies
Suitable for: Master students
Not suitable for: Beginning students, foreign students whose German proficiency is not advanced enough to understand domain-specific content

GENERAL TEST PURPOSE

Tested for: Assessing students' level of competence in business and economics knowledge at any time throughout the course of bachelor study (one-off testing, e.g. with the short version of 36 items); assessing competence development over several points throughout the course of studies (from the bachelor's program to the beginning of the master's program in business and economic sciences)
Suitable for: Self-assessment
Not suitable for: Admission and suitability diagnostics

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Grades from the attended courses in business and economics during studies (business studies, economics; methods); socio-demographic questionnaire; intelligence; further psychological scales
Suitable for: Further measures of competence
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for the individual and group level, also over time, at university level, state and national level, and at international level (with the same test and subscales).
Suitable for: —
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Formative assessment as a teaching-learning tool for students; assessment of competence level as a supplement to examinations; recognition of strengths and deficits in the course of studies to recommend courses and support offers; observation of competence development throughout the course of studies (e.g. annual increase)
Suitable for: Further application possibilities after consultation with the test developers
Not suitable for: Study entry and suitability diagnostics

PERMITTED CONSEQUENCES

Tested for: Individual and group feedback on the state of knowledge and knowledge development
Suitable for: Course recommendations (e.g. elective (compulsory) courses); evaluations of teaching-learning offers
Not suitable for: Selection decisions, assessment of academic performance

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The WiWiKom test will continue to be used in the annual surveys of the various ongoing national and international research projects during the course of the project: the long-term development of student competencies (in a real panel) and further international analyses will be continued in 2020.
The test is available on request for use at universities.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Brückner, S. & Pellegrino, J. W. (2016). Integrating the Analysis of Mental Operations into Multilevel Models to Validate an Assessment of Higher Education Students' Competency in Business and Economics. *Journal of Educational Measurement, 53*(3), 293–312.
Zlatkin-Troitschanskaia, O. Schmidt, S. Brückner, S. Förster, M. Yamaoka, M. & Asano, T. (2016). Macroeconomic Knowledge of Higher Education Students in Germany and Japan – A Multilevel Analysis of Contextual and Personal Effects. *Assessment & Evaluation in Higher Education, 41*(5), 787–801.
http://dx.doi.org/10.1080/02602938.2016.1162279
Zlatkin-Troitschanskaia, O. Jitomirski, J. Happ, R. Molerov, D. Schlax, J. Kühling-Thees, C. Förster, M. & Brückner, S. (2019). Validating a Test for Measuring Knowledge and Understanding of Economics Among University Students. *Zeitschrift für Pädagogische Psychologie, 33*(2), 119–133.

SECONDARY LITERATURE

Vidal, R.U. (2013). Measurement of learning outcomes in higher education: The case of Ceneval in Mexiko. In: Blömeke, S.; Zlatkin-Troitschanskaia, O.; Kuhn, C. & Fege, J. (Eds.), *Modeling and Measuring Competencies in Higher Education.* Rotterdam: Sense Publishers.
Walstad, W. B. & Rebeck, K. (2008). The Test of Understanding of College Economics. *American Economic Review*, 98, pp. 547–551.
Zlatkin-Troitschanskaia, O.; Förster, M.; Brückner, S. & Happ, R. (2014). Insights from a German Assessment of Business and Economics Competence. In: Coates, H. (ed.), *Higher Education Learning Outcomes Assessment: International Perspectives*, 175–197. Frankfurt am Main: Lang. http://dx.doi.org/10.3726/978-3-653-04632-8

Business and Economics Teachers' Pedagogical Content Knowledge: PCK-Business and Economics

SUMMARY

Name: Pedagogical Content Knowledge of Business and Economics Teachers: PCK-Business and Economics:

Domain: Business and economics education

Assessed competencies: Pedagogical content knowledge of (pre-service) teachers (PCK) of business and economics

Target group: (Pre-service) teachers of economics (bachelor and master students of business and economics education, trainees and teachers of economics at vocational study seminars and schools); (pre-service) economics teachers in the US and possibly other countries

Test type: Situation-based tasks in multiple-choice and open-response formats

Modality: Paper-pencil test

Duration: 45 minutes (long version); 30 minutes (short version)

Test structure: Long version: 17 items divided into the categories "practical requirements" (lesson planning, reacting to students' comments) and "cognitive processes" (application & analysis, creation). Short version: 11 items

General test purpose: Competence assessment

Application scenarios: Comparison of competence levels between groups in the various training phases of teacher education (bachelor, master, practical training phase, teaching service); comparison between the subjects of economics and mathematics; intervention study with pre-post measurement design to assess competence development in business and economics

Not suitable for: Entry diagnostics; beginning students

Note for practical use: Test administrator manual including organizational information to ensure standardized test administration; a coding manual is provided to ensure objective scoring; also contains documentation of the pedagogical content knowledge tasks and theoretical background

Applied in projects: Pedagogical Content Knowledge of (Future) Teachers in Business and Economics: Theoretical Modeling, Test Development, and Validation; Assessing Subject-specific Competencies in Teacher Education in Mathematics and Business and Economics – a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi); teaching project "Promotion of Action-oriented Competencies Using a Video-based Learning Tool for Simulating Authentic Teaching Situations Among Students of Business and Economics Education"; international cooperation project "Feasibility Study for Adaptation and Use of a German-language Instrument for Assessing the Subject-related Didactic Competencies of Economics Teachers in the USA"

Contact / Location:
Dr. Christiane Kuhn
Johannes Gutenberg University Mainz
Email: christiane.kuhn@uni-mainz.de
Websites: https://www.blogs.uni-mainz.de/fb03-business-education/team/research-assistants/dr-christiane-kuhn/
https://www.blogs.uni-mainz.de/fb03-business-education/dissertation-project-pedagogical-content-knowledge-of-future-teachers-in-business-and-economics-theoretical-modeling-test-development-and-validation/

GENERAL INFORMATION

Name: Pedagogical Content Knowledge of Business and Economics Teachers: PCK-Business and Economics
Domain: Business and economics education
Applied in projects: Pedagogical Content Knowledge of (Future) Teachers in Business and Economics: Theoretical Modeling, Test Development, and Validation; Assessing Subject-specific Competencies in Teacher Education in Mathematics and Business and Economics – a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi); teaching project "Promotion of Action-oriented Competencies Using a Video-based Learning Tool for Simulating Authentic Teaching Situations Among Students of Business and Economics Education"; international cooperation project "Feasibility Study for Adaptation and Use of a German-language Instrument for Assessing the Subject-related Didactic Competencies of Economics Teachers in the USA"
Contact / Location:
Dr. Christiane Kuhn
Johannes Gutenberg University Mainz
Email: christiane.kuhn@uni-mainz.de
Websites: https://www.blogs.uni-mainz.de/fb03-business-education/team/research-assistants/dr-christiane-kuhn/
https://www.blogs.uni-mainz.de/fb03-business-education/dissertation-project-pedagogical-content-knowledge-of-future-teachers-in-business-and-economics-theoretical-modeling-test-development-and-validation/

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Pedagogical content knowledge (PCK)of (pre-service) teachers of business and economics (in the field of business administration)
Theory model: Structural model of the pedagogical content knowledge of teachers of business and economics (Kuhn, 2014)
Test type: Situation-based tasks in multiple-choice and open-response formats
Modality: Paper-pencil-test
Test structure:
Item pool: 17 items divided into the categories "practical requirements" (lesson planning, reacting to students' comments) and "cognitive processes" (application & analysis, creation) (see also "Further Information"); *Test* versions: 1 long version; 17 items; 1 short version: 11 items

TEST QUALITY CRITERIA

Reliability: Cronbach's α: 0.68 (N=338, 17 items); EAP reliability: 0.71
Validity:
Test content: Validation of item content by means of a standardized expert survey (N=13) (Kuhn, 2014).
Response processes: Cognitive validation of the items in cognitive interviews using the think-aloud method (N=6 students of business and economics education) (Kuhn & Brückner, 2013)
Internal test structure: As assumed, one-dimensional structure (confirmatory factor analysis (CFA), N = 338, 17 items: $\chi 2/df = 1.34$ with $\chi 2 = 159.60$ and df = 119; RMSEA = 0.03; CFI = 0.94; TLI = 0.93; WRMR = 0.90); conformity with the Rasch-model for ordinal data (partial credit model, N = 338): WMNSQ with values [0.91; 1.14]; validation of the response categories of the MC items (discriminatory power of the response categories according to the category sequence) (Kuhn et al. 2016)
Relationships with other competence indicators: As assumed, business and economics PCK is significantly, but moderately correlated with business administration (BAKT-K) and economic knowledge (WBT-K) (Kuhn et al. 2014). The comparison with a contrast group (trainee teachers and teachers with subjects other than business and economics) confirms that the test does not cover general pedagogical or general didactic knowledge (Kuhn, 2014); other expected correlations with external variables, e.g. school leaving grade, school internships completed, years of in-classroom teaching experience (Kuhn, 2014)
Consequences of testing: As assumed, test performance increases significantly between bachelor students, master students, and trainees in their practical training phase; there are no significant differences between trainees and in-service teachers (Kuhn, 2014).
Test fairness: The test assesses the same characteristic in the different groups: (a) female vs. male and (b) students vs. trainees and/or teachers (Kuhn, 2014).

PRACTICAL USE

GENERAL INFORMATION

Duration: 45 minutes (long version); 30 minutes (short version)
Testing materials: Test booklets, pens
Special features: Test administrator manual includes organizational information to ensure standardized test administration; a coding manual is provided to ensure objective scoring; also contains documentation of the didactic tasks and theoretical background.
Practical example: See previous use in above-mentioned research and teaching projects.

DOMAIN

Tested for: Business and economics education
Suitable for: Economics education in the general education sector
Not suitable for: —

TARGET GROUP

Tested for: (Pre-service) teachers of business and economics (bachelor's and master's students of business and economics education, trainees and teachers of business and economics at vocational study seminars and schools)
Suitable for: (Pre-service) teachers of economics in the US (an adapted version of the business and economics tasks is available and has already been used in the US for validation purposes); adaptations for use in other countries are conceivable
Not suitable for: Beginning students

GENERAL TEST PURPOSE

Tested for: Assessing the level of competence in the various groups
Suitable for: Assessing competence development
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Tests for assessing domain-specific content knowledge (CK) and video-based tests for assessing action-related and reflective competence in the field of business and economics (see the ARKOM test); tests for assessing PCK, CK, AC, RC in mathematics with (pre-service) teachers of mathematics for cross-domain comparative analyses
Suitable for: Further tests for assessing subject-related or generic competencies
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at group level; comparisons of competence levels between the individual phases (students, trainee teachers, in-service teachers)
Suitable for: International level; development over time
Not suitable for: —

APPLICATIONS

Tested for: See above-mentioned research and teaching projects.
Suitable for: Assessing PCK development in all phases of teacher training; comparisons with other school subjects
Not suitable for: Entry diagnostics

PERMITTED CONSEQUENCES

Tested for: Assessment of test-takers' level of PCK during university education, practical training phase, teaching service
Suitable for: Assessment of the development of PCK over the different phases; statistical prediction of the state of knowledge at a later point in time
Not suitable for: Entry diagnostics

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

In the future, the test will also be used in the project "Promotion of Subject-specific Competencies of Pre-service Teachers in Mathematics and Business & Economics Using Video-based ELMaWi Tools (ELMaWi-Transfer)"

ADDITIONAL INFORMATION

Test Structure in Detail (short version):

Pedagogical content knowledge in economics*		
Practical Requirements	Cognitive Processes	
	Application & Analysis	Creation
Lesson planning	5 items (MC)	6 items (open-response)
Reacting to student comments	3 items (MC)	3 items (open-response)

* For the core facets "sales", "procurement", "macroeconomics" related to students' competencies to be fostered (e.g. decision-making, managing conflicts)

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Kuhn, C.; Alonzo, A. C. & Zlatkin-Troitschanskaia, O. (2016). Evaluating the pedagogical content knowledge of pre- and in-service teachers of business and economics to ensure quality of classroom practice in vocational education and training. *Empirical Research in Vocational Education and Training, 8*(5). doi: 10.1186/s40461-016-0031-2
Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung.* Empirische Berufsbildungs- und Hochschulforschung, vol. 2. Landau: Verlag Empirische Pädagogik.
Kuhn, C. & Brückner, S. (2013). *Analyse des fachdidaktischen Wissens von (angehenden) Lehrkräften in der kaufmännisch-verwaltenden Bildung mit der Methode des lauten Denkens. bwp@ Berufs- und Wirtschaftspädagogik – online,* 24, pp. 1–20. Accessed at http://www.bwpat.de/ausgabe24/kuhn_brueckner_bwpat24.pdf

SECONDARY LITERATURE

Kuhn, C. & Happ, R. (2017). Effekte von curricularen Veränderungen im Zuge der Bologna Reform auf das fachdidaktische Wissen von Studierenden der Wirtschaftspädagogik. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 113(4), pp. 542–568.
Kuhn, C.; Happ, R.; Zlatkin-Troitschanskaia, O.; Beck, K.; Förster, M. & Preuße, D. (2014). Kompetenzentwicklung angehender Lehrkräfte im kaufmännisch-verwaltenden Bereich – Erfassung und Zusammenhänge von Fachwissen und fachdidaktischem Wissen. In: Winther, E. & Prenzel, M. (eds.), Perspektiven der empirischen Berufsbildungsforschung: Kompetenz und Professionalisierung [Sonderheft]. *Zeitschrift für Erziehungswissenschaft*, *17*(1), pp. 149–167. doi: 10.1007/s11618-013-0456-3

Business Start-Up Test

SUMMARY

Name: Business Start-up test

Domain: Entrepreneurship education

Assessed competencies: Expertise in setting up a business

Target group: Students of all domains

Test type: Multiple-choice-test

Modality: Paper-pencil-test

Duration: 45 minutes

Test structure: 40 items from 10 content areas: start-up marketing, start-up-related personnel management, start-up-relevant legal forms, industrial property rights, innovation and technology management, accounting and start-up-related controlling, start-up financing, strategic management and business model, business planning, start-up-related project management. Versions: two partial versions (A and B).

General test purpose: Knowledge assessment

Application scenarios: Teaching-learning tool, course comparisons, university comparisons

Not suitable for: —

Note for practical use: Instructions, a scoring booklet with solutions and further materials for administering the test are provided. A total score is generated from all tasks solved correctly. Separate scoring of individual content areas or knowledge levels is possible.

Applied in projects: Higher Entrepreneurship Education Diagnostics (HEED)

Contact / Location:
Prof. Dr. Michael Schefczyk
Technical University of Dresden
Email: info@gruenderlehrstuhl.de
Website: https://tu-dresden.de/bu/wirtschaft/lei/die-professur/beschaeftigte/schefczyk_profil

GENERAL INFORMATION

Name: Business start-up test
Applied in projects: Higher Entrepreneurship Education Diagnostics (HEED)
Contact / Location:
Prof. Dr. Michael Schefczyk
Technical University of Dresden
Email: info@gruenderlehrstuhl.de
Website: https://tu-dresden.de/bu/wirtschaft/lei/die-professur/beschaeftigte/schefczyk_profil

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Expertise in setting up a business
Theoretical model: HEED competence model with 2 levels (1st level: knowledge, i.e. remembering and understanding; 2nd level: knowledge transformation, i.e. applying, analyzing and evaluating)
Test type: Multiple-choice test
Modality: Paper-pencil test
Test structure:
Item pool: 40 items from 10 content areas: start-up marketing, start-up-related personnel management, start-up-relevant legal forms, industrial property rights, innovation and technology management, accounting and start-up-related controlling, start-up financing, strategic management and business model, business planning, start-up-related project management. The content areas were adopted from the HEED competence model for entrepreneurial competence
Versions: Two parallel versions (A and B)

PRACTICAL USE

GENERAL INFORMATION

Duration: 45 minutes
Testing materials: Test sheet, answer sheet, reference supplement
Special features: Instructions and a scoring booklet are provided for the administration and scoring of the test. All correct response options must be selected. During scoring, all correctly solved tasks are counted. Evaluation of individual content areas or competence levels is possible. The following further options are available for test scoring: 1. addition of all correctly solved N1 tasks (sum score competence level 1); 2. addition of all correctly solved N2-tasks (sum score competence level 2); 3. addition of all correctly solved tasks of a certain content area (sum score of content areas 1-10)

Practical example: The test was administered to different student groups (N=203 participants altogether) in 2 parallel test versions (version A and B). With this sample, a first validation was carried out and the items were reviewed and revised extensively. The revised test needs to be tested with a new sample.

DOMAIN

Tested for: Business start-up
Suitable for: Entrepreneurship education
Not suitable for: —

TARGET GROUP

Tested for: Students of different study domains at two universities, at different qualification levels, technology- and knowledge-based start-ups
Suitable for: Students of all disciplines
Not suitable for: The university sector focuses on technology-oriented and knowledge-based business start-ups. Accordingly, individual domain-specific skills, e.g. industrial property rights, are less relevant for low-tech start-ups.

GENERAL TEST PURPOSE

Tested for: Assessing knowledge levels
Suitable for: Assessing knowledge levels, repeated testing
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Socio-demographic questions
Suitable for: Further measures for competence assessment
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at the course level.
Suitable for: Assessing individuals over time, university level
Not suitable for: —

APPLICATIONS

Tested for: Course comparisons
Suitable for: Teaching-learning tool, university comparisons
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Counseling participants
Suitable for: Revised items needs further validation, hence, the test has not been cleared for any consequences yet
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

Following initial validation studies, the items were revised and have yet to be tested again with a new sample before reliable and valid statements are possible. The revised test can be requested from the test developers.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Weinberger, E. & Schefczyk, M. (2015). *HEED – Higher Entrepreneurship Education Diagnostics. Teilprojekt: TIP – The Instructor Perspective. Entwicklung eines Kompetenzmodells zur unternehmerischen Kompetenz.* Final report. Dresden: Technische Universität Dresden.

SECONDARY LITERATURE

Braukmann, U. & Schneider, D. (2007). Didaktische Zielklassen der Entwicklung unternehmerischer Kompetenz. In R. Bader, G. Keiser & T. Unger (eds.), *Entwicklung unternehmerischer Kompetenz in der Berufsbildung. Hintergründe, Ziele und Prozesse berufspädagogischen Handelns*, pp. 157–180. Bielefeld: W. Bertelsmann Verlag.
Haase, H. & Lautenschläger, A. (2010). The 'Teachability Dilemma' of entrepreneurship. *International Entrepreneurship and Management Journal*, 7(2), pp. 145–162. doi: 10.1007/s11365-010-0150-3

Classroom Management Teacher Expertise (CME)

SUMMARY

Name: Classroom Management Expertise (CME)

Domain: Various subjects

Assessed competencies: Classroom Management Expertise (CME)

Target group: Bachelor's and master's students, pre-service teachers in their practical phase, in-service teachers

Test type: Performance test; multiple-choice test, open-response items

Modality: Online test with videos

Duration: 20 minutes (for in-service teachers)

Test structure: 24 test questions; 4 video clips of 1-2 minutes each with 19 open-response and 5 closed items. Content dimensions: Design of transitions, organization of time processes, regulation of student behavior, provision of feedback. Cognitive processes: Accuracy of perception and holistic perception, interpretation

General test purpose: Assessing competence levels

Application scenarios: Comparisons at course level and university level

Not suitable for: So far, no concrete consequences beyond basic research have been examined.

Note for practical use: Use as paper-pencil-test or online test. Special features when coding the answers to the open-response test questions (20% must be coded twice). So far, only one total score has been returned. The feedback of results for partial scales according to content dimensions or cognitive processes is not possible due to the short test length and low number of items.

Applied in projects: Classroom Management Expertise (CME); Longitudinal Study of Pedagogical Competences of Student Teachers and Trainee Teachers (LEK-R); Teacher Education and Development Study (TEDS teaching, TEDS validation

Contact / Location:
Prof. Dr. Johannes König
University of Cologne
Email: johannes.koenig@uni-koeln.de
Website: https://www.hf.uni-koeln.de/35829

GENERAL INFORMATION

Name: Classroom Management Expertise (CME)
Applied in projects: Classroom Management Expertise (CME); Longitudinal Study of Pedagogical Competences of Student Teachers and Trainee Teachers (LEK-R); Teacher Education and Development Study (TEDS teaching, TEDS validation
Contact / Location:
Prof. Dr. Johannes König
University of Cologne
Email: johannes.koenig@uni-koeln.de
Website: https://www.hf.uni-koeln.de/35829

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Classroom Management Expertise (CME)
Theoretical model: "Competence as continuum" model (Blömeke, Gustafsson & Shavelson, 2015); facets "perception" and "interpretation" (unconsidered facet: "decision making")
Test type: Performance test; multiple-choice test, open-response
Modality: Online test with videos
Test structure:
Item pool: 24 items; 4 video clips of 1-2 minutes each with 19 open and 5 closed items. Content dimensions: design of transitions, organization of time processes, regulation of student behavior, provision of feedback. Cognitive processes: accuracy of perception and holistic perception, interpretation

TEST QUALITY CRITERIA

Reliability: Reliability of the overall score for teachers: EAP .699, WLE .706, Cronbach's α .700 (King, 2015); pre-service teachers: EAP .73 (König & Kramer, 2016); student teachers: Cronbach's α .74 (Kramer et al. 2017); Low testlet effects due to video vignettes (König, 2015; Casale et al. 2017)
Validity:
Test content: —
Response processes: Statistically significantly higher correlation with the pedagogical knowledge subscale "creating" than with the subscales "remembering", "understanding and analyzing" (König, 2015)
Internal test structure: —
Relationships with other competence indicators: Medium-high correlation with pedagogical knowledge (TEDS-M Test, .47***) (König, 2015); statistically significantly higher correlation with the pedagogical knowledge subscale of "class leadership and motivation" than with the subscales "structuring", "dealing with heterogeneity", "performance assessment" (König, 2015)

Consequences of testing: In-service teachers (average 18 years of teaching experience) perform significantly better than pre-service teachers (König & Kramer, 2016); CME scores for pre-service teachers predict student assessment of teaching quality (omnipresence: β = .47*, regular clarity: β = .36*) (König & Kramer, 2016); CME scores for working mathematics teachers predict significant performance growth from 7th grade to 8th grade (König et al. 2020); CME scores of teachers correlate negatively with the burnout scales (König, 2015) (-.23 for "reduced performance" and -.28 for "depersonalization", both $p \leq .01$) (König & Rothland, 2016). Training in seminars on class leadership (video- and transcript-based) promotes growth in CME more than seminars on other topics in the field of teaching (Kramer et al. 2017).
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 20 minutes (for working teachers)
Testing materials: Paper, pencil, test sheets, computer with projector or computer, laptop and internet connection
Special features: Use as paper-pencil-test or online test. Special features when coding the answers to the open-response test questions (20% must be coded twice). So far, only one total score has been returned. The feedback of results to partial scales according to content dimensions or cognitive processes is not possible due to the short test length and low item number.
Practical example: The total sample comprises 744 prospective and working teachers. The test was used in university teacher training seminars to examine the extent to which the acquisition of situation-specific skills for class leadership and the acquisition of pedagogical knowledge by student teachers can be promoted more effectively through instruction videos than through teaching transcripts (Kramer et al. 2017). For this purpose, training was implemented in the form of a video- and a transcript-supported seminar at the University of Cologne, each of which was controlled by means of seminars of the same module without specific media support. All groups (N=222 student teachers) took part in a pre- and post-measurement. Students in the video- and transcript-supported groups showed statistically significant increases in both test results with practical significance compared to the control group. There were no differences in competence growth between the experimental groups. Furthermore, the test was used with experienced teachers within the framework of the CME study (König, 2015); a positive correlation to pedagogical knowledge as well as negative correlations to burnout scales were demonstrated (N=119).

DOMAIN

Tested for: Various subjects
Suitable for: —
Not suitable for: —

TARGET GROUP

Tested for: Bachelor's and master's students, pre-service teachers in their practical phase, in-service teachers
Suitable for: —
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing competence levels; pre-post measurements of competence changes among students (over 1 semester)
Suitable for: Evaluating training measures
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Tests on pedagogical, technical and didactic knowledge in mathematics (GPK, MCK, MPCK); situation-specific competence facets (M_PID and P_PID); error detection test (M_Speed); tests on beliefs of teachers; observational instruments for the assessment of teaching quality by external raters, partly also video-based evaluation; self-assessment questionnaire of teachers and assessment of teaching quality by students; burnout scales; self-efficacy
Suitable for: Other measures of competence
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for the course level, university level, state and national level, international level (German-speaking countries).
Suitable for: —
Not suitable for: Individual diagnostics, international level (non-German-speaking countries)

APPLICATION SCENARIOS

Tested for: Comparisons at course or university level
Suitable for: So far, no concrete applications beyond basic research have been tested.
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: So far, no concrete consequences beyond basic research have been examined.
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The CME test is currently being used in a validation study in Germany, Austria and Switzerland to expand the possible field of application (König et al. 2018). In addition, an extension of the CME test is currently being developed that specifically measures the previously unconsidered facet of "decision-making". New findings on both aspects will soon be available.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Casale, G.; Strauss, S.; Hennemann, T. & König, J. (2016). How can class leadership expertise be measured? Review of a video-based survey tool for teachers using generalizability theory. *Empirical Special Education*, 8 (2), pp. 119–139.

König, J. (2015). Measuring Classroom Management Expertise (CME) of Teachers: A Video-Based Assessment Approach and Statistical Results. *Cogent Education*, 2(1), 991178.

König, J.; Kramer, C.; Affolter, B.; Bach, A.; Biedermann, H.; Brühwiler, C.; Greiner, U.; Hollenstein, L. & Katstaller, M. (2018). *CME testing in different training contexts: results from measurement invariance analyses. [CME Assessment in different educational contexts: Findings from measurement invariance analyses.]* Presentation on the GEBF annual conference, Basel, Switzerland.

SECONDARY LITERATURE

König, J. & Kramer, C. (2016). Teacher professional knowledge and classroom management: On the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM – The International Journal on Mathematics Education, 48* (1), pp. 139–151.

König, J. & Rothland, M. (2016). Klassenführungswissen als Ressource der Burnout-Prävention? Zum Nutzen von pädagogisch-psychologischem Wissen im Lehrerberuf. *Unterrichtswissenschaft*, 44(4), pp. 425–441.

Kramer, C.; König, J.; Kaiser, G.; Ligtvoet, R. & Blömeke, S. (2017). Der Einsatz von Unterrichtsvideos in der universitären Ausbildung: Zur Wirksamkeit video- und transkriptgestützter Seminare zur Klassenführung auf pädagogisches Wissen und situationsspezifische Fähigkeiten angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 20, Supplement 1, pp. 137–164.

Competence for Teaching Computer Science

SUMMARY

Name: Measurement of competence for teaching computer science

Domain: Computer science didactics

Assessed competencies: Teaching skills, beliefs, motivational orientations

Target group: Students of computer science education and in-service computer science teachers

Test type: Multiple-choice-test and questionnaire

Modality: Paper-pencil test, online test

Duration: 45 minutes

Test structure: Part 1: didactic contents and teaching phases (36 items); part 2: professional beliefs (34 items) and motivational orientations (13 items)

General test purpose: Assessment of the level of competence in computer science didactics, the characteristics of beliefs and motivational orientations

Application scenarios: Teacher training in computer science, at school level

Not suitable for: Assessment of content knowledge

Note for practical use: The test is available in German and English. The test needs further development and validation. Administration and analyses must be coordinated with the project team.

Applied in projects: Competencies for Teaching Computer Science (KUI)

Contact / Location:
Prof. Dr. Niclas Schaper
University of Paderborn
Email: niclas.schaper@upb.de

Prof. Dr. Peter Hubwieser
University of Munich
Email: peter.hubwieser@tum.de

Dr. Marc-Pascal Berges
University of Munich
Email: marc.berges@tum.de

Dr. Elena Bender
University of Paderborn
Email: elena.bender@upb.de

GENERAL INFORMATION

Name: Measurement of competence for teaching computer science
Applied in projects: Competencies for Teaching Computer Science (KUI)
Contact / Location:
Prof. Dr. Niclas Schaper
University of Paderborn
Email: niclas.schaper@upb.de

Prof. Dr. Peter Hubwieser
University of Munich
Email: peter.hubwieser@tum.de

Dr. Marc-Pascal Berges
University of Munich
Email: marc.berges@tum.de

Dr. Elena Bender
University of Paderborn
Email: elena.bender@upb.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Didactic competence, beliefs, motivational orientations
Theoretical model: Competence structure model for teaching in information technology (Bender et al. 2015) based on the model of professional action competence (Baumert & Kunter, 2011)
Test type: Multiple-choice test and questionnaire
Modality: Paper-pencil test, online test
Test structure:
2 test parts: part 1 for assessing didactic competencies: 6 scales with 36 items for assessing different facets of didactic content (learning content; methods and social forms; student cognition; objectives of teaching units) and teaching phases (planning and design of learning situations; reacting to students in teaching situations; evaluation of teaching processes). Part 2 for assessing professional beliefs: 2 scales for assessing professional beliefs (34 items) and motivational orientations (13 items)

TEST QUALITY CRITERIA

Reliability: Part 1: insufficient reliability; part 2: satisfactory to sufficient reliability of the individual scales
Validity:
Test content: The test items represent the theoretical competence structure model
Response processes: Results from pre-tests on comprehensibility and relevance led to revisions with a subsequent improvement of results. The newly revised items have yet to be cognitively validated.
Internal test structure: —

Relationships with other competence indicators: It is not yet possible to make valid statements on the relations with other competence indicators
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION
Duration: 45 minutes
Testing materials: Test items, questionnaire
Special features: The test is available in German and English. The test needs further development and validation. Administration and analyses must be coordinated with the project team.
Practical example: Used and tested with a sample of 155 computer science teachers (N = 20 teacher education students and N = 12 pre-service trainee teachers in their practical phase, N = 123 in-service teachers) throughout Germany. In addition, the competence structure model was used in courses on the didactics of computer science for upper secondary schools and vocational training in Bavaria and Austria.

DOMAIN
Tested for: Computer science didactics
Suitable for: —
Not suitable for: Assessment of content knowledge

TARGET GROUP
Tested for: Pre-service and in-service computer science teachers
Suitable for: —
Not suitable for: —

GENERAL TEST PURPOSE
Tested for: Assessment of competence level in computer science didactics, the characteristics of beliefs and motivational orientations
Suitable for: —
Not suitable for: Assessing competence development

USE WITH OTHER TESTS AND QUESTIONNAIRES
Tested for: —
Suitable for: Other measures of competence facets
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY
Tested for: Results are statistically meaningful for individual diagnostics, state and national level. International level: The test was used in a pilot study in Michigan (USA)
Suitable for: —
Not suitable for: —

APPLICATION SCENARIOS
Tested for: Assessing the level of competence in computer science didactics in federal states. Differentiation of competence levels according to teaching experience
Suitable for: —
Not suitable for: Assessment of content knowledge

PERMITTED CONSEQUENCES
Tested for: Statements about the structure of beliefs and motivational orientations and their interrelations
Suitable for: —
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE
The scales for assessing professional beliefs (34 items) and for assessing motivational orientations (13 items) are available (Bender, Schaper & Seifert, 2018).

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS
Bender, E.; Hubwieser, P.; Schaper, N.; Margaritis, M.; Berges, M.; Ohrndorf, L.; Magenheim, J. & Schubert, S. (2015). Towards a competency model for teaching computer science. *Peabody Journal of Education*, 90 (4), pp. 519–532.
Bender, E.; Schaper, N. & Seifert, A. (2018). Professionelle Überzeugungen und motivationale Orientierungen von Informatiklehrkräften. *Journal for Educational Research Online*, 10 (1), pp. 70–99.

SECONDARY LITERATURE
ACM/IEEE (2013). = The Joint Task Force on Computing Curricula, Association for Computing Machinery, IEEE Computer Society. (2013). Computer Science Curricula 2013. Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. Accessed on 10.03.2014. Available at: https://www.acm.org/education/CS2013-final-report.pdf.
Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In: Kunter, M.; Baumert, J.; Blum, W. Klusmann, U.; Krauss, S. & Neubrand, M. (eds.), *Professionelle Kompetenz von Lehrkräften – Ergebnisse des Forschungsprogramms COACTIV*, pp. 29–54. Münster: Waxmann.

Content Knowledge and Pedagogical Content Knowledge in Accounting

SUMMARY

Name: Test for assessing domain-specific content knowledge (CK) and pedagogical content knowledge (PCK) in accounting

Domain: Accounting, business education

Assessed competencies: Content knowledge and pedagogical content knowledge in accounting

Target group: Students of business education

Test type: Multiple-choice and open-response format

Modality: Paper-pencil test

Duration: 75 minutes; part on CK and PCK 40 minutes

Test structure: A total of 187 items and 3 scenarios: questionnaire for self-assessment of CK and PCK, test for CK and PCK in accounting, test for assessing didactic decisions, questionnaire for teaching-learning process-related attitudes, test for assessing study-related learning behavior, questionnaire for school and vocational careers

Item pool: 49 items; CK: learning content areas: Tasks and legal bases, double-entry system, procurement and sales processes. PCK: didactic content areas: knowledge about student cognition, knowledge about making content accessible, knowledge about the potential of tasks. Test versions: The item pool is divided into a multi-matrix design: 7 item clusters with 7 items each. Each test booklet contains 4 clusters = 28 items.

General test purpose: Cross-sectional assessment of knowledge level and development; evaluation of a training concept

Applications scenarios: Teaching-learning tool

Not suitable for: —

Note for practical use: The test is available in a multi-matrix design and must be scored using the provided coding and requires IRT analysis.

Applied in projects: Modeling and assessing of domain-specific and didactic competencies in business studies (KoMeWP)

Contact / Location:
Prof. Dr. Jürgen Seifried
University of Mannheim
Email: seifried@bwl.uni-mannheim.de
Website: https://www.bwl.uni-mannheim.de/seifried/

GENERAL INFORMATION

Name: Test for assessing domain-specific content knowledge (CK) and pedagogical content knowledge (PCK) in accounting
Applied in projects: Modeling and assessing of domain-specific and didactic competencies in business studies (KoMeWP)
Contact / Location:
Prof. Dr. Jürgen Seifried
University of Mannheim
Email: seifried@bwl.uni-mannheim.de
Website: https://www.bwl.uni-mannheim.de/seifried/

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Domain-specific content knowledge (CK) and pedagogical content knowledge (PCK) in accounting
Theoretical model: Structural model of the professional competence of teachers according to Weinert (2001); operationalization of professional knowledge according to Shulman (1986)
Test type: 49 items; 23 multiple-choice questions; 13 open-response questions
Modality: Paper-pencil test
Test structure:
Item pool: 187 items and 3 scenarios: questionnaire for self-assessment of CK and PCK (time: 5 minutes; 33 items), test for CK and PCK in accounting (time: 40 minutes; 28 items), test for assessing didactic decisions (time: 7 minutes; 3 scenarios), questionnaire on teaching-learning process-related views (time: 8 minutes; 51 items), test to assess study-related learning behavior (time: 10 minutes; 75 items), questionnaire on school and career development (time: 5 minutes).
CK: learning content areas: tasks and legal bases; double-entry system; procurement and sales processes; PCK: knowledge about student cognition; knowledge about making content accessible; knowledge about the potential of tasks.
Test versions: The item pool is divided according to a multi-matrix design: 7 item clusters with 7 items each (Youden Square design). Each test contains 4 clusters = 28 items.

TEST QUALITY CRITERIA

Reliability: EAP/PV-Reliabilities for CK: .79, for PCK: .73; MNSQ values (infit) between 0.93 ≤ MNSQ ≤ 1.08; test quality (item difficulty, selectivity, internal consistencies, MNSQ values, etc.) Berger et al. (2013, 2015); Fritsch et al. (2015); Schnick-Vollmer et al. (2015)

Validity:

Test content: Document analysis and curriculum analyses (accounting textbooks; (framework) curricula); expert surveys (Berger et al. 2013). The test covers central learning content of the domain.

Response processes: The think-aloud-study clarifies didactic argumentation patterns.

Internal test structure: A two-dimensional model (CK vs. PCK, Bouley et al. 2015; Fritsch et al. 2015) best represents the data structure. A distinction between different didactic sub-dimensions does not lead to an improvement of the model fit.

Relationships with other competence indicators: Relations compliant with theory were found in a German and an Austrian sample (on construct validity, Bouley et al. 2015; on criterion validity, Fritsch et al. 2015). Discriminant and convergent validity were tested with other measures of cognitive ability (school grades, study grades); mean value differences between tested groups are in line with expectations (comparison Germany – Austria).

Consequences of testing: —

Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 75 minutes; part on CK and PCK 40 minutes

Testing materials: Test sheets

Special features: The test is available in a multi-matrix design and must be scored using the provided coding and needs IRT analysis

Practical example: Surveys with students of business education in Germany (bachelor N = 590 and master N = 552 at 24 university locations) and Austria (N = 249 students at 4 universities). In addition to other tests, some of the items developed in KoMeWP (24 and 17 items, respectively) from 2014-2016 were used to assess the increase in knowledge of pre-service teachers through interventions (intervention objective: support in dealing with accounting errors in a way that promotes learning). The target group was initially 226 pre-service teachers for vocational schools (business education) and later 31 or 37 students of business education. The interventions showed effects in line with expectations; subject-related didactic competencies in particular were improved.

DOMAIN

Tested for: Accounting, business education

Suitable for: Upper secondary school economics

Not suitable for: —

TARGET GROUP

Tested for: Students of business education in Germany and Austria; pre-service trainee teachers for vocational schools

Suitable for: Students in teacher education for general upper secondary schools for the subject of economics

Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Cross-sectional assessment of the level and development of competence

Suitable for: Assessing competence development longitudinally

Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Grades

Suitable for: Other measures of competence assessment

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for the course level, university level, Germany and Austria.

Suitable for: —

Not suitable for: Individual diagnostics, state and national level, international level

APPLICATION SCENARIOS

Tested for: Teaching-learning tool, evaluation of a training concept

Suitable for: —

Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Improvement of teaching

Suitable for: —

Not suitable for: Individual diagnostics

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test is being both further developed and used by the Seifried & Wuttke working group. The test can be requested from the research group.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Berger, S.; Bouley, F.; Fritsch, S.; Krille, C.; Seifried, J. & Wuttke, E. (2015). Fachwissen und fachdidaktisches Wissen im wirtschaftspädagogischen Studium. Entwicklung eines Testinstruments und erste empirische Befunde. In: Koch-Priewe, B.; Köker, A.; Seifried, J. & Wuttke, E. (eds.), *Kompetenzerwerb an Hochschulen. Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte.* pp. 105–125. Bad Heilbrunn: Klinkhardt.

Fritsch, S.; Berger, S.; Seifried, J.; Bouley, F.; Wuttke, E.; Schnick-Vollmer, K. & Schmitz, B. (2015). The impact of university teacher training on prospective teachers' CK and PCK – A comparison between Austria and Germany. In: Zlatkin-Troitschanskaia, O. & Shavelson, R. (eds.), *Special Issue on Assessment of Domain-specific Professional Competencies, Empirical Research in Vocational Education and Training (ERVET), 7*(4). DOI 10.1186/s40461-015-0014-8. Download at http://www.ervet-journal.com/content/7/1/4

Bouley F.; Wuttke E.; Schnick-Vollmer K.; Schmitz B.; Berger S.; Fritsch S. & Seifried J. (2015). Professional Competence of Prospective Teachers in Business and Economics Education -Evaluation of a competence model using structural equation modelling. *Peabody Journal of Education 90*(4), pp. 491–502.

SECONDARY LITERATURE

Bouley, F.; Berger, S.; Fritsch, S.; Wuttke, E.; Seifried, J.; Schnick-Vollmer, K. & Schmitz, B. (2015). Der Einfluss von universitären und außeruniversitären Lerngelegenheiten auf das Fachwissen und fachdidaktische Wissen von angehenden Lehrkräften an kaufmännisch-berufsbildenden Schulen. *Zeitschrift für Pädagogik, supplement 61,* pp. 100–115. Weinheim und Basel: Beltz Juventa.

Krille, C.; Salge, B.; Wuttke, E. & Seifried, J. (2017). Evaluation of a training programme to improve the Professional Error Competence of prospective teachers. In: Wuttke, E. & Seifried, J. (eds.), *Professional Error Competence of Preservice Teachers: Evaluation and Support,* pp. 75–98. Cham: Springer.

Schnick-Vollmer, K.; Berger, S.; Bouley, F.; Fritsch, S.; Schmitz, B.; Seifried, J. & Wuttke, E. (2015). Modeling the competencies of prospective business and economics teachers. Professional knowledge in accounting. *Zeitschrift für Psychologie*, 223(1), pp. 24–30.

Content Knowledge, Pedagogical Content Knowledge, and Pedagogical Knowledge (MCK, MPCK and GPK) of Early Education Teachers in Mathematics: KomMa-Tests

SUMMARY

Name: KomMa-test: Test for the assessment of professional competence of (pre-service) early education mathematics teachers

Domain: Pre-service early education teachers (aspiring to or with technical college degree)

Assessed competencies: Mathematical content knowledge (MCK), mathematical pedagogical content knowledge (MPCK), general pedagogical knowledge (GPK)

Target group: Start and end of training (technical college and university of applied sciences), all federal states in Germany; young professionals (max. 5 years after completion of training)

Test type: Performance test; Multiple-choice items and open-response items

Modality: Paper-pencil test (recommended), also online test

Duration: 60 minutes

Test structure: 70 items; 3 partial tests: KomMa MCK test (24 items), KomMa MPCK test (28 items), KomMa GPK test (18 items)

General test purpose: Assessing the level of knowledge, development of professional knowledge in the field of mathematics

Application scenarios: For research purposes only

Not suitable for: Individual diagnostics

Note for practical use: Only for research purposes, no individual diagnostics

Applied in projects: Structure, level and development of kindergarten teachers' professional competencies in mathematics (KomMa): https://www.projekte.hu-berlin.de/de/pro-komma/projekt-komma; Pro-KomMa: https://www.projekte.hu-berlin.de/de/pro-komma

Contact / Location:
Dr. Lars Jenßen
Humboldt University of Berlin
Email: lars.jenssen@hu-berlin.de

GENERAL INFORMATION

Name: KomMa-Test: test for the assessment of professional competence of (pre-service) early education mathematics teachers
Applied in projects: Structure, level and development of kindergarten teachers' professional competencies in mathematics (KomMa): https://www.projekte.hu-berlin.de/de/pro-komma/projekt-komma; Pro-KomMa: https://www.projekte.hu-berlin.de/de/pro-komma
Contact / Location:
Dr. Lars Jenßen
Humboldt University of Berlin
Email: lars.jenssen@hu-berlin.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Mathematical content knowledge (MCK), mathematical pedagogical content knowledge (MPCK), general pedagogical knowledge (GPK)
Theoretical model: Differentiation of knowledge facets of professional knowledge into (mathematical) content knowledge, (mathematical) pedagogical content knowledge and general pedagogical knowledge according to Shulman (1986) and Ball & Bass (2002); competence as a continuum (professional knowledge as disposition of professional competence) according to Blömeke, Gustafsson & Shavelson (2015)
Type of test: Performance test; Multiple-choice items and open-response items
Modality: Paper-pencil test (recommended), also online test
Test structure:
Item pool: 70 items in the 3 subtests: KomMa-MCK-Test: 24 items for the contents "Numbers, quantities and operations", "Form, space and change", "Sizes, measurements and relations" and "Data, combinatorics and chance" with the processes "Modeling", "Problem solving", "Arguing", "Communicating", "Presenting" and "Recognizing patterns and structures"; KomMa-MPCK-Test: 28 items for the dimensions "Development and diagnosis of mathematical abilities in children" (17 items) and "Design of mathematical learning environments" (11 items); KomMa GPK test: 18 items for the contents "Educational theoretical basics" (5 items), "Psychological basics" (6 items) and "Didactic methodological basics" (7 items)

TEST QUALITY CRITERIA

Reliability: Reliability coefficients for 2PL models for the sample of prospective early education teachers (N=1851): KomMa-MCK test: Rel=0.88; KomMa-MPCK test: Rel=0.87; KomMa-GPK test: Rel=0.68

Validity:

Test content: An expert panel (N=12) supports the validity of the test content (Jenßen, Dunekacke & Blömeke, 2015).

Response processes: Analysis of the test processes using the think-aloud method with 18 prospective early education teachers (Jenßen, Dunekacke & Blömeke, 2015).

Internal test structure: Several studies with different populations each show that MCK, MPCK and GPK can be empirically differentiated in accordance with theory (Blömeke et al. 2015).

Relationships with other competence indicators: Several studies with different populations confirm the theoretically assumed relations to affective-motivational dispositions, general-cognitive abilities, action-related constructs and performance (Jenßen, Dunekacke, Eid & Blömeke, 2015; Dunekacke, Jenßen, Eilerts & Blömeke, 2016; Blömeke et al. 2015, 2017).

Consequences of testing: —

Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 60 minutes

Testing materials: Test sheets and pen or PC, laptop, tablet

Special features: Not suitable for individual diagnostics

Practical example: The paper-pencil test can be used, for example, in seminar groups or class groups. For larger samples and the use of other methods, e.g. intelligence tests or questionnaires for learning opportunities, the application in multi-matrix design is recommended. For this, information on the distribution of the items for such applications can be requested from the project coordinators. The test content explicitly refers to training content taught at technical colleges and universities of applied sciences in all federal states of Germany. So far, the tests have been used by about 1800 pre-service early education teachers and about 250 early education teachers in the career entry phase. The test is also used by the Elementary Department of the German Centre for Teacher Training (DZLM) for the evaluation of further training courses.

DOMAIN

Tested for: Early education mathematics

Suitable for: —

Not suitable for: —

TARGET GROUP

Tested for: Pre-service early education teachers (aspiring or with technical college degree) at the beginning and end of teacher training (technical college and college), all federal states in Germany

Suitable for: Career starters (max. 5 years after completion of training)

Not suitable for: Early education teachers with longer professional experience

GENERAL TEST PURPOSE

Tested for: Assessing the level of knowledge, development of professional knowledge in the field of mathematics

Suitable for: Evaluation of training courses

Not suitable for: Individual diagnostics

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Questionnaires for the assessment of mathematics-related beliefs are recommended, video tests (e.g. tests for mathematics-related situational perception and action planning by Dunekacke et al., 2016), intelligence tests

Suitable for: Further measures of competence

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for the institutional training level (technical college vs. college and beginning vs. end).

Suitable for: —

Not suitable for: Individual diagnostics, state and national level, international comparisons

APPLICATIONS

Tested for: Research purposes

Suitable for: —

Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —

Suitable for: —

Not suitable for: Individual diagnostics

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

If you are interested in the test, please contact Dr. Lars Jenßen: lars.jenssen@hu-berlin.de. If you would like to use it, please submit a short description of what you would like to use the test for and always include a reference to the basic project publications (Blömeke et al. 2015). In individual cases, the project coordinators reserve the right to permit use only after completion of the qualification work in the Pro-KomMa project.

ADDITIONAL INFORMATION

It should be noted that the field of early mathematical education is constantly being explored and that pedagogical concepts can change in practice. The test items are also subject to these changes in practice.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Blömeke, S.; Jenßen, L.; Dunekacke, S.; Suhl, U.; Grassmann, M. & Wedekind, H. (2015). Leistungstests zur Messung der professionellen Kompetenz frühpädagogischer Fachkräfte. *Zeitschrift für Pädagogische Psychologie, 29*(3-4), pp. 177–191.

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

SECONDARY LITERATURE

Jenßen, L.; Dunekacke, S.; Eid, M. & Blömeke, S. (2015). The relationship of mathematical competence and mathematics anxiety – An application of latent state-trait theory. *Zeitschrift für Psychologie, 223*(1), pp. 31–39.

Blömeke, S.; Jenßen, L.; Grassmann, M.; Dunekacke, S. & Wedekind, H. (2017). Process mediates structure: Relation of preschool teacher education and preschool teachers' knowledge. *Journal of Educational Psychology, 109*, pp. 338–354.

Dunekacke, S.; Jenßen, L.; Eilerts, K. & Blömeke, S. (2016). Epistemological beliefs of prospective preschool teachers and their relation to knowledge, perception, and planning abilities in the field of mathematics: A process model. *ZDM, 48*(1-2), pp. 125–137.

Dialogical Explanatory Test (DET) for Physics Teachers

SUMMARY

Name: Dialogical Explanation Test (DET)

Domain: Teaching physics (mechanics) and other scientific domains

Assessed competencies: Performance-based explanatory skills in physics

Target group: Pre-service and in-service physics teachers

Test type: Performance test; video analysis of explanations

Modality: Performance test

Duration: 10 minutes per test plus 10 minutes of preparation time

Test structure: Three topics for explanation; trained school students and students of teacher education (subject teaching) ask partially standardized questions

General test purpose: Assessing the relation between performance and knowledge, evaluation of interventions

Application scenarios: Development of action-oriented competencies in the teaching profession

Not suitable for: Primary and secondary education, admissions to the teaching profession, as substitute for examinations

Note for practical use: A training for students who act as addressees for the explanations is required.

Applied in projects: Professional Knowledge of Physics Student Teachers (Profile-P);
https://www.uni-due.de/didaktik_der_physik/forschungfischer.php

Contact / Location:
PD Dr. Christoph Kulgemeyer
University of Bremen
Email: kulgemeyer@physik.uni-bremen.de

GENERAL INFORMATION

Name: Dialogical Explanation Test (DET)
Applied in projects: Professional Knowledge of Physics Student Teachers (Profile-P);
https://www.uni-due.de/didaktik_der_physik/forschungfischer.php
Contact / Location:
PD Dr. Christoph Kulgemeyer
University of Bremen
Email: kulgemeyer@physik.uni-bremen.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Performance-based explanatory skills in physics
Theoretical model: Model of dialogical explanation with the intention of mediation (Kulgemeyer & Schecker, 2012; Wittwer & Renkl, 2008)
Test type: Performance test; video analysis of explanations
Modality: Performance test
Test structure:
Item pool: 3 topics to be explained to one person; trained students ask partially standardized questions. Evaluation with the help of criterion-based video analysis with regard to the appropriateness of the explanation and its intended audience (12 scoring categories)

TEST QUALITY CRITERIA

Reliability: Cronbach's α=. 772 (total score), interrater agreement: PÜ=0.91
Validity:
Test content: Categories from literature on explanation quality; topics from mechanics correspond to topics from middle school physics.
Response processes: Intended processes studied through stimulated recall interviews.
Internal test structure: One-dimensional Rasch model describes the data in accordance with the theory.
Relationships with other competence indicators: A nomological network is theoretically formulated and tested: Positive medium correlations to content knowledge, to didactic knowledge, to self-efficacy beliefs in explaining physics and to constructivist epistemological beliefs.
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 10 minutes per test
Testing materials: Material set with illustrations, scenario descriptions, paper, pen, video camera
Special features: 10 minutes of preparation time per respondent to prepare for the explanation. 10 minutes of explanation time with trained school student. Scoring of the videos according to coding manual, 50 minutes scoring time per video, scoring by sum of occurring categories
Practical example: The test was used at five German universities to evaluate the influence of university-imparted professional knowledge on the quality of action in a selected teaching situation, i.e. explaining. To this end, prospective physics teachers filled out the tests on content knowledge and didactic knowledge, which are curricularly valid for the content of the teacher training course. They also took part in the explanatory tests. Interrelations were analyzed using path models; the didactic knowledge of the subject mediates the influence of content knowledge on the quality of explanations. The school students who acted as addressees for the explanations were carefully trained. Video feedback was used and explanations were practiced in the training according to the guidelines, and also reflected and further refined based on videos.

DOMAIN

Tested for: Physics education (mechanics)
Suitable for: Other scientific domains
Not suitable for: —

TARGET GROUP

Tested for: Students of physics education
Suitable for: Pre-service teachers in their practical phase, in-service teachers
Not suitable for: School students

GENERAL TEST PURPOSE

Tested for: Analyses of relations between teaching performance and knowledge
Suitable for: Evaluating instructional interventions
Not suitable for: Admission to the teaching profession, aptitude test for lateral entrants

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Knowledge test on content knowledge and didactic knowledge, scales on attitudes
Suitable for: For other uses, please consult the test developers.
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for group comparisons.
Suitable for: Individual diagnostics
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Developing action-oriented competencies in the teaching profession
Suitable for: For other uses, please consult the test developers.
Not suitable for: Examination substitute

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: The admissibility of possible consequences must first be determined in the validation process.
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

At www.explainingscience.de you can find a digital version of the test with a simulated dialogue created by Kulgemeyer and Bartels. Interested parties can contact PD Dr. Christoph Kulgemeyer, kulgemeyer@physik.uni-bremen.de, for further information.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Kulgemeyer, C. & Tomczyszyn, E. (2015). Physik erklären – Messung der Erklärensfähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften, 21*(1), pp. 111–126.
Kulgemeyer, C. & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, pp. 1–26. DOI: 10.1002/tea.21457.
Bartels, H. & Kulgemeyer, C. (2018). Explaining physics: an online test for self-assessment and instructor training. *European Journal of Physics*, pp. 1–7. DOI: 10.1088/1361-6404/aaeb5e.

SECONDARY LITERATURE

Kulgemeyer, C. & Schecker, H. (2012). Physikalische Kommunikationskompetenz – Empirische Validierung eines normativen Modells. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, pp. 29 – 54.
Wittwer, J. & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43(1), pp. 49 – 64.

Didactic Adaptivity of Teachers

SUMMARY

Name: Didactic Adaptivity of Teachers

Domain: All subjects

Assessed competencies: Didactic adaptivity as an aspect of the planning competence of pre-service teachers (= subject-independent adaptation of tasks from the planned lesson to the cognitive prerequisites of a learning group)

Target group: Students and pre-service teachers in their practical training phase

Test type: Performance test; content analysis of written lesson plans of pre-service teachers

Modality: Reconstruction and quantification of planning decisions from written lesson plans

Duration: Depends on the scope of the written lesson planning, e.g. 15-20 minutes for a 15-page comprehensive planning

Test structure: 11 items; 4 areas: descriptive level: description of the situation (learning group), description of the task(s); analytical level: linking the requirements of the learning group and the task(s), linking the didactic adaptivity with further planning elements, with 2 to 4 items each. 3 optional items for didactic adaptivity for learners with special learning needs

General test purpose: Assessing a partial aspect of planning competence at one or more points in time

Application scenarios: Analyzing competence development, influences of learning opportunities, different processes in subgroups, e.g. for competence-oriented optimization of teaching

Not suitable for: —

Note for practical use: The reliability of the scoring depends on the quality of the written lesson plan (e.g. completeness of statements). The scoring is carried out by trained raters and then analyzed using IRT methods.

Applied in projects: Planning Competence of Teachers (PlanvoLL); https://www.hf.uni-koeln.de/36932; The Role of Professional Knowledge of Pre-Service German Teachers in their Lesson Planning: Validation and Methodological Innovation (PlanvoLL-D); https://www.hf.uni-koeln.de/37700

Contact / Location:
Prof. Dr. Johannes König
University of Cologne
Website: https://www.hf.uni-koeln.de/33049
Email: johannes.koenig@uni-koeln.de

GENERAL INFORMATION

Name: Didactic Adaptivity of Teachers
Applied in projects: Planning Competence of Teachers (PlanvoLL); https://www.hf.uni-koeln.de/36932; The Role of Professional Knowledge of Pre-Service German Teachers in their Lesson Planning: Validation and Methodological Innovation (PlanvoLL-D); https://www.hf.uni-koeln.de/37700
Contact / Location:
Prof. Dr. Johannes König
University of Cologne
Website: https://www.hf.uni-koeln.de/33049
Email: johannes.koenig@uni-koeln.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Didactic adaptivity as an aspect of the planning competence of pre-service teachers; subject-independent adaptation of tasks for the planned lesson to the cognitive prerequisites of a learning group
Theoretical model: Didactic adaptivity as situation-specific ability of perception, interpretation and decision making (König et al. 2017)
Test type: Performance test; written lesson plans of pre-service teachers are analyzed with regard to content
Modality: Reconstruction and quantification of planning decisions from written lesson plans
Test structure:
Item pool: 11 items; 4 areas: descriptive level: description of the situation (learning group), description of the task(s); analytical level: linkage of requirements of the learning group and task(s), linkage of didactic adaptivity with further planning elements), with 2 to 4 items each. 3 optional items for didactic adaptivity for learners with special learning need

TEST QUALITY CRITERIA

Reliability: Interrater reliability (Cohen's kappa for two trained raters) on average 0.83 (0.67 to 0.93)
Validity:
Test content: Increase in didactic adaptivity between the first and last teaching sample during preparatory teaching service and training ($d > 0.8$).
Response processes: An expert review of the categories supported a theory-driven construction (Buchholtz & König, 2015).
Internal test structure: The IRT scaling showed an EAP reliability of 0.703 (theta variance 1.631). The items are sufficiently distributed across the ability spectrum. The discrimination values were > 0.3. The weighted variance squares were between 0.80 and 1.20.

Relationships with other competence indicators: Validity was tested with the school leaving grade as an indicator of basic cognitive ability, the First State Examination (teaching qualification) as an indicator of academic achievement and pedagogical beliefs about teaching and learning. Convergent validity with pedagogical knowledge on how to deal with heterogeneity.
Consequences of testing: The predictor 'planning competence' is statistically significant in a regression with the scale of internal differentiation as a dependent variable: Pre-service trainee teachers with higher skill levels in planning competence were rated by their students as making a distinction in their teaching approach according to students' abilities. This finding is also evident in other samples and in the self-assessments of pre-service trainee teachers.
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: Depends on the scope of the written lesson plan
Testing materials: Written lesson plan, coding guide
Special features: The reliability of the scoring depends on the quality of the written lesson plans (e.g. completeness of the statements). The scoring is carried out by trained raters and then statistically analyzed using IRT methods.
Practical example: König, Blömeke & Dohmen (2015) score two different written plans by way of example; the scoring process can be transferred to other written plans.

DOMAIN

Tested for: All subjects
Suitable for: —
Not suitable for: —

TARGET GROUP

Tested for: Pre-service teachers (students)
Suitable for: Pre-service teachers in their practical training
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing a partial aspect of planning competence at one or more points in time
Suitable for: —
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Use with pedagogical knowledge, content knowledge, pedagogical content knowledge and tests; school leaving grade, grade of the first and second state examination; pedagogical beliefs about teaching and learning; aspects of the self-assessed quality of teaching
Suitable for: Further measures of competence assessment
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for cohorts over time; training course and training institutions; federal state level
Suitable for: Individual diagnostic purposes, national level, international level (after examining cultural and linguistic comparability)
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Analyzing competence development, influences of learning opportunities, different processes in subgroups, e.g. for competence-oriented optimization of teaching
Suitable for: —
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: —
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The survey instrument has been published (König, Buchholtz & Dohmen, 2015) so that it can be used by interested scientists. In the PlanvoLL-D project, the test was also further developed for heterogeneous and inclusive learning groups (joint teaching) and supplemented by three optional items. Moreover, the adaptivity and planning competence for the subject of German will be expanded for pedagogical content knowledge. At present, the dimension 'structuring' is being developed as an aspect of planning competence and added to analyses of existing data sets from PlanvoLL and PlanvoLL-D. In the medium term, the development of a standardized vignette-based test is planned to assess didactic adaptivity as an aspect of planning competence.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Buchholtz, C. & König, J. (2015). Erfassung von Planungskompetenz im Praxissemester. *Journal für LehrerInnenbildung*, 15(1), pp. 39–45.
König, J.; Buchholtz, C. & Dohmen, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18 (2), pp. 375–404. https://link.springer.com/article/10.1007%2Fs11618-015-0625-7

SECONDARY LITERATURE

König, J.; Bremerich-Vos, A.; Buchholtz, C.I Lammerding, S.; Strauß, S.; Fladung, I. & Schleiffer, C. (2017). Die Bedeutung des Professionswissens von Referendarinnen und Referendaren mit Fach Deutsch für ihre Planungskompetenz (PlanvoLL-D). In: Wernke, S. & Zierer, K. (Eds.), *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich?! Status Quo und Perspektiven aus Sicht der empirischen Forschung*, pp. 121–133. Bad Heilbrunn: Klinkhardt.

Economic Competence at the Beginning of Studies: WiWiSET-Test

SUMMARY

Name: WiWiSET test

Domain: Economics

Assessed competencies: Economic thinking and understanding; previous study-related knowledge in economics

Target group: Bachelor students at the beginning of their studies in business and economics; applicants and students interested in business and economics

Test type: Multiple-choice test

Modality: Paper-pencil test; also available as an online test

Duration: 25 minutes (short version)

Test setup: Long version: 45 items on basic economic principles; short version: 25 items

General test purpose: Assessing economic knowledge and understanding at the beginning of studies at one point or over time (e.g. before and after preparatory courses), assessing factors for predicting study success or termination of studies in business and economics

Application scenarios: Entry diagnostics; support diagnostics in the initial study phase and study orientation phase; assessment of prior study-related knowledge in economics, planning and organization of the initial study phase and study orientation phase (e.g. recommendations for preparatory courses, tutorials etc.), study-related decisions, self-assessment

Not suitable for: Suitability diagnostics, assessment of study performance

Note for practical use: Individual or group testing possible. The test should be administered under supervision. There are two versions of the questionnaire with reversed item order (in German, English and Japanese); the results are reported back via an online system and retrieved by the test participants using an anonymous code they created. Language skills should be controlled with a proficiency test if possible.

Applied in projects: Validation of an Entrance Examination in the Study Domain of Business and Economics – A National and International Comparative Study of Universities and Universities of Applied Sciences (WiWiSET); www.wiwi-kompetenz.de/wiwiset-2016-2019/

Contact / Location:
Prof. Dr. Hans Anand Pant
Humboldt University of Berlin
Email: hansanand.pant@hu-berlin.de

GENERAL INFORMATION

Name: WiWiSET test
Applied in projects: Validation of an Entrance Examination in the Study Domain of Business and Economics – A National and International Comparative Study of Universities and Universities of Applied Sciences (WiWiSET); www.wiwi-kompetenz.de/wiwiset-2016-2019/
Contact / Location:
Prof. Dr. Hans Anand Pant
Humboldt University of Berlin
Email: hansanand.pant@hu-berlin.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Economic thinking and understanding; prior study-related knowledge in economics
Theoretical model: For the theoretical framework model for the conceptualization of economic knowledge, see Zlatkin-Troitschanskaia et al. (2014) and Walstad et al. (2013), which differentiates between three cognitive levels and the three content dimensions (fundamentals of economics, micro- and macroeconomics).
Test type: Multiple-choice test
Modality: Paper-pencil test, also online test
Test setup:
Item pool: Long version: 45 items on basic economic principles; short version: 25 items in 3 dimensions: basic economic principles (15 items), basic microeconomics principles (5 items), bacis macroeconomics principles (5 items). The 10 items on basic micro- and macroeconomics were adapted from the fourth version of the US-American Test of Understanding College Economics (TUCE IV). The 45 items on basic economic principles were adapted from the fourth version of the US-American Test of Economic Literacy (TEL IV).
Versions: Two versions with reversed item order. The test is available in German, English and Japanese.

TEST QUALITY CRITERIA

Reliability: Psychometric requirements for difficulty, selectivity and reliability are met. Internal consistency in a representative survey with N = 3,977 beginning students: α = .75 (25 items).

Validity:

Test content: Textbook analyses, curriculum analyses of 96 economic degree programs in Germany and online rating with 78 experts confirming the validity of the content and of curricular relevance of the test.

Response processes: The theoretical cognitive assumptions on the three dimensions were confirmed in cognitive interviews with 20 economics students using the think-aloud method.

Internal test structure: The testing of measurement models for the internal structure shows theoretically compliant results; the confirmatory factor analyses (CFAs) confirm the one-dimensionality of the construct (χ^2 (778) =1198.295; RMSEA = .019 (90% CI .017-.021); SRMR = .029; CFI = .894, AIC = 72684.160).

Relationships with other competence indicators: Expected relations with other measures of cognitive abilities, e.g. school grades, fluid intelligence (assessed using BEFKI [Berlin test for the determination of fluid and crystalline intelligence]).

Consequences of testing: The WiWiSET test score from the first survey at the beginning of studies explains a significant part of the variance in the test performance in the second survey (after one academic year), also when controlling for fluid intelligence, school leaving grade, migration background, gender and age (F(6,773) = 132.78, p<.001, Adj R^2=.5037). The theoretical assumptions on the predictive validity of the WiWiSET test were confirmed.

Test fairness: DIF analyses and measurement invariance analyses point to a gender bias and a language bias. Language skills should be controlled during test application (if possible with a language proficiency test).

PRACTICAL USE

GENERAL INFORMATION

Duration: 25 minutes (short version)

Testing materials: Test booklets (or computer, laptop for online version)

Special features: Individual or group testing is possible. The test should be performed under supervision. The questionnaire is used in two versions with reversed item order; the results are reported back via an online system and retrieved by the test participants using an anonymous, self-generated code. The test is available in German, English and Japanese.

Practical example: The test was used on site at two points in time under controlled conditions at universities (start of winter semester 2016/17 at 42 universities and universities of applied sciences with N = 3,977 first-year students, and start of winter semester 2017/18 at 22 universities and universities of applied sciences with N = 2,832 students at the end of the 2nd/beginning of the 3rd semester). The test participants were able to view their results via an online feedback system.

DOMAIN

Tested for: Business and economics

Suitable for: All subjects with economic contents

Not suitable for: —

TARGET GROUP

Tested for: Bachelor students at the beginning of their studies in business and economics

Suitable for: Prospective students, higher education applicants

Not suitable for: Students in higher semesters

GENERAL TEST PURPOSE

Tested for: Assessing basic economics knowledge and understanding at the beginning of studies; assessing factors for the prognosis of study success or study dropout (after the first study year)

Suitable for: Assessing economic knowledge over time (e.g. before and after preparatory courses)

Not suitable for: Suitability diagnostics, assessment of study performance

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Intelligence (BEFKI 11+), school leaving grade

Suitable for: All other tests to assess students' cognitive and non-cognitive competencies

Not suitable for: No limitation

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for individual and group level (also over time); institutional comparative level (university of applied science vs. university); university level; state and national level; international level.

Suitable for: —

Not suitable for: —

APPLICATION SCENARIOS

Tested for: Initial diagnostics; assessment of prior study-related knowledge in economics; prediction of study success, learning process diagnostics in the initial study phase

Suitable for: Support diagnostics as well as planning and organization of the initial study phase and study orientation phase (e.g.. recommendations for preparatory courses, tutorials etc.), study and career counseling; course selection decisions, self-assessment

Not suitable for: Suitability diagnostics, assessment of study performance

PERMITTED CONSEQUENCES

Tested for: Individual and group feedback on the level of knowledge and knowledge development

Suitable for: Prediction of study success or failure, program choice recommendation, course recommendation (e.g. preliminary courses)

Not suitable for: Selection decisions, admission, assessment of academic performance

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test can be requested from the project team.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Kühling-Thees, C. Happ, R. Zlatkin-Troitschanskaia, O. & Pant, H.A. (2020). The Impact of Entry Preconditions on Student Dropout and Subject Change in Business and Economics. In O. Zlatkin-Troitschanskaia,H. A. Pant, M. Toepper & C. Lautenbach (Eds.), *Student Learning in German Higher Education Innovative Measurement Approaches and Research Results* (pp. 349–369). Wiesbaden: Springer.

Pant, H. A. Zlatkin-Troitschanskaia, O.; Schipolowski, S. & Förster, M. (2016). WiWiSET – Validierung eines Studieneingangstests in der Fachdomäne Wirtschaftswissenschaften: eine national und international vergleichende Studie an Universitäten und Fachhochschulen. In: Zlatkin-Troitschanskaia, O.; Pant, H. A.; Lautenbach, C. & M. Toepper (Eds.), *Kompetenzmodelle und Instrumente der Kompetenzerfassung im Hochschulsektor – Validierungen und methodische Innovationen (KoKoHs).* KoKoHs Working Papers, 10, pp. 70–72. Berlin & Mainz: Humboldt Universität & Johannes Gutenberg-Universität.

Zlatkin-Troitschanskaia, O. & Kühling-Thees, C. (2019). Study-related Domain-specific and Generic Competencies of Economics Students: Insights from a German-Japanese Study. *Journal of Asia Pacific Studies (Waseda University, Japan),* 35, pp. 285 – 297.

SECONDARY LITERATURE

Schipolowski, S. Wilhelm, O. & Schroeders, U. (2017). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz ab der 11. Jahrgangsstufe (BEFKI 11+)* [Berlin test of fluid and crystallized intelligence for grades 11 and above]. Göttingen: Hogrefe.

Walstad, W. B.; Rebeck, K. & Butters, R. B. (2013). The Test of Economic Literacy: Development and results. *Journal of Economic Education*, 44(3), pp. 298–309.

Walstad, W. B.; Rebeck, K. & Butters, R. B. (2013). The Test of Economic Literacy: Development and results. *Journal of Economic Education*, 44(3), pp. 298–309.

Zlatkin-Troitschanskaia, O.; Förster, M.; Brückner, S. & Happ, R. (2014). Insights from a German Assessment of Business and Economics Competence. In: Coates, H. (Ed.), *Higher Education Learning Outcomes Assessment: International Perspectives*, pp. 175–197. Frankfurt am Main: Lang. http://dx.doi.org/10.3726/978-3-653-04632-8

Facets-of-Competence (FOC) Evaluation Tool for Medical and Nursing Evaluators

SUMMARY

Name: FOC (Facets-of-Competence) evaluation tool for evaluators with a medical and nursing background

Domain: Medicine

Assessed competencies: Responsibility; teamwork and collegiality; knowledge of personal limits and possibilities; structure and work planning; dealing with mistakes; scientifically based working methods; verbal communication with colleagues, supervisors

Target group: Medical students from the 8th semester onwards

Test type: Evaluation instrument (external evaluation of medical students by doctors and nursing staff; self-evaluation)

Modality: Paper-pencil test (can also be completed online)

Duration: 5-10 minutes

Test structure: 14 items: one item for each of the 7 competence facets per case (5-point scale: 1: insufficient, 3: acceptable, 5: very good; "no judgment possible") plus one item for each of these 7 facets regarding the certainty of judgment for the final evaluation (5-point scale: 1: uncertain, 5: certain). For each item there is a blank space to enter keywords as a basis for the evaluation. Additionally, every case scenario is rated with 7 items (one per case).

General test purpose: Assessment of the state or development of competence

Application scenarios: Summative: intermediate examination, final examination; formative: before or at the end of the practical year

Not suitable for: —

Note for practical use: Key examples to simplify rating and improve standardization. Scorer training with the evaluation tool is required.

Applied in projects: First Version: The Utrecht Hamburg Trainee Responsibility for Unfamiliar Situations Test (UHTRUST); Further developed version: Medical competences (Hamburg-Oldenburg-München) ($ÄKHOM_{t1}$), adaptation for $ÄKHOM_{t2}$

Contact / Location:
Prof. Dr. med. Sigrid Harendza, MME
University Medical Center Hamburg-Eppendorf (UKE), III.
Email: harendza@uke.de
Website: https://www.uke.de/kliniken-institute/kliniken/iii.-medizinische-klinik-und-poliklinik/forschung/ag-harendza.html

GENERAL INFORMATION

Name: FOC (Facets-of-Competence) evaluation tool for medical and nursing evaluators
Applied in projects: The Utrecht Hamburg Trainee Responsibility for Unfamiliar Situations Test (UHTRUST); Medical competences (Hamburg-Oldenburg-München) ($ÄKHOM_{t1}$), adaptation for $ÄKHOM_{t2}$
Contact / Location:
Prof. Dr. med. Sigrid Harendza, MME
University Medical Center Hamburg-Eppendorf (UKE), III.
Email: harendza@uke.de
Website: https://www.uke.de/kliniken-institute/kliniken/iii.-medizinische-klinik-und-poliklinik/forschung/ag-harendza.html

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Responsibility; teamwork and collegiality; knowledge of personal limits and possibilities; structure and work planning; dealing with mistakes; scientifically based working methods; verbal communication with colleagues; supervision
Theoretical model: Concept of competence according to Epstein et al. (2002) and Van der Blijet et al. (2002); competence model derived from the CanMEDs model, competence development according to Chen et al. (2015)
Test type: Evaluation instrument (external evaluation of medical students by doctors and nursing staff; self-evaluation)
Modality: Paper-pencil test (can also be completed online)
Test structure:
Itempool: 14 items: one item for each of the seven competence facets per case (5-point scale: 1: insufficient, 3: acceptable, 5: very good; "no judgment possible") plus one item for each of these 7 items regarding the certainty of judgment for the final evaluation (5-point scale: 1: uncertain, 5: certain). For each item there is a blank space for entering keywords as a basis for the evaluation.
Test versions: Key examples to simplify rating and improve standardization

TEST QUALITY CRITERIA

Reliability: Internal consistency for senior physicians .90; assistant physicians .80; nurses .78

Validity:

Test content: The competencies to be assessed (with FOC) were aligned with the professional reality of physicians at the participating hospitals. All observable behaviors are derived from practice and tested under realistic conditions.

Response processes: The FOC questionnaire covers the stages of the competence model (see above).

Internal test structure: —

Relationships with other competence indicators: —

Consequences of testing: Currently only formative uses, with feedback for students

Test fairness: A positive ethics vote was cast; gender fairness is given.

PRACTICAL USE

GENERAL INFORMATION

Duration: 5-10 minutes

Testing materials: Observable, simulated conversation or interaction and, if applicable, further observations of a simulated working day

Special features: Scorer training with the evaluation tool is required; there are quantitative and qualitative components of the evaluation (ratings with associated fields of justification and assessment of judgment certainty).

Practical example: The instrument was used for the evaluation of competence facets in the performance of a competence-based examination at the end of medical studies (project ÄKHOM). The evaluation took place under realistic conditions (simulation situation of first day working in a hospital) and was reported back to the 70 and 90 participants respectively with the results of other competence assessments (see below).

DOMAIN

Tested for: Medicine

Suitable for: Forensic medicine, teaching profession

Not suitable for: —

TARGET GROUP

Tested for: Medical students from the 10th semester onwards

Suitable for: —

Not suitable for: Medical students before the 8th semester

GENERAL TEST PURPOSE

Tested for: Assessing students' state of competence at any time from the 8th semester onwards

Suitable for: Repeated use to assess competence development

Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Knowledge test, empathy assessment, entrustable professional activities (EPA) assessment sheet

Suitable for: —

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: The results are statistically meaningful for individuals at a certain point in time, cohorts of students, different university locations internationally: The pilot questionnaire was tested in Germany and the Netherlands (project UHTRUST).

Suitable for: Use over time

Not suitable for: —

APPLICATION SCENARIOS

Tested for: Summative: intermediate examination, final examination; formative: before or at the end of the practical year

Suitable for: Continuing medical education

Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Career entry competence

Suitable for: —

Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The FOC evaluation tool is available on request from the project team. The method can also be used formatively and longitudinally. The instrument is currently used to assess development over time.

ADDITIONAL INFORMATION

The FOC assessment tool must be used in an observable situation. Scorer training is required.

Do you have any further questions or do you need help, e.g. with the practical application of the instrument or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Prediger, S.; Fürstenberg, S.; Berberat, P.O.; Kadmon, M. & Harendza, S. (2019). Interprofessional assessment of medical students' competencies with an instrument suitable for physicians and nurses. In: *BMC Med Educ*.

SECONDARY LITERATURE

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

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

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

General Pedagogical Knowledge (GPK)

SUMMARY

Name: General Pedagogical Knowledge (GPK)

Domain: Pedagogical knowledge in different subjects

Assessed competencies: Pedagogical knowledge of pre-service and in-service teachers

Target group: Students at the beginning of teacher education, during teacher education; pre-service teachers in their practical training phase; in-service teachers

Test type: Short answers and multiple-choice items

Modality: Paper-pencil test; online use also possible

Duration: 20 minutes (short version)

Test structure: 18 items (published short version for teacher education). Test design matrix with four dimensions of content: pedagogical knowledge about dealing with heterogeneity, structuring lessons, class leadership and motivation, performance assessment and three dimensions of cognitive demands (remembering, understanding and analyzing, creating)

General test purpose: Assessing the level and development of pedagogical knowledge during and after teacher education

Application scenarios: University comparisons and evaluation, correlations with external criteria

Not suitable for: —

Note for practical use: Uniform administration under observation during the paper-pencil test or the online test is required. The answers to the open-response test questions are coded according to instructions. The answers are interpreted by means of an IRT scale. The results can be reported in the form of an overall score, partial results of individual scales (using the long version) or as competence levels (Klemenz & König, 2019).

Applied in projects: CME (https://www.hf.uni-koeln.de/35829), DIDAKTUM (https://www.hf.uni-koeln.de/34071), EMW (https://www.hf.uni-koeln.de/35034), LEK (https://www.hf.uni-koeln.de/33207), LEK-R (https://www.hf.uni-koeln.de/35966), LtP (https://www.hf.uni-koeln.de/37356), PKE (https://www.hf.uni-koeln.de/36292), PlanvoLL-D (https://www.hf.uni-koeln.de/37700), SKILL, TEDS-M (https://www.teds-unterricht.uni-hamburg.de/de/weitere-teds-studien/teds-m.html), TEDS-Follow Up (https://www.teds-unterricht.uni-hamburg.de/weitere-teds-studien/teds-fu.html), TEDS-Unterricht (https://www.teds-unterricht.uni-hamburg.de/), TEDS-Validierung (https://www.hf.uni-koeln.de/38129), ZuS-QS (http://zus.uni-koeln.de/qualitaetssicherung.html)

Contact / Location:
Prof. Dr. Johannes König
University of Cologne
Email: johannes.koenig@uni-koeln.de

GENERAL INFORMATION

Name: General Pedagogical Knowledge (GPK)
Applied in projects: CME (https://www.hf.uni-koeln.de/35829), DIDAKTUM (https://www.hf.uni-koeln.de/34071), EMW (https://www.hf.uni-koeln.de/35034), LEK (https://www.hf.uni-koeln.de/33207), LEK-R (https://www.hf.uni-koeln.de/35966), LtP (https://www.hf.uni-koeln.de/37356), PKE (https://www.hf.uni-koeln.de/36292), PlanvoLL-D (https://www.hf.uni-koeln.de/37700), SKILL, TEDS-M (https://www.teds-unterricht.uni-hamburg.de/de/weitere-teds-studien/teds-m.html), TEDS-Follow Up (https://www.teds-unterricht.uni-hamburg.de/weitere-teds-studien/teds-fu.html), TEDS-Unterricht (https://www.teds-unterricht.uni-hamburg.de/), TEDS-Validierung (https://www.hf.uni-koeln.de/38129), ZuS-QS (http://zus.uni-koeln.de/qualitaetssicherung.html)
Contact / Location:
Prof. Dr. Johannes König
University of Cologne
Email: johannes.koenig@uni-koeln.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Pedagogical knowledge of pre-service and in-service teachers
Theoretical model: Pedagogical knowledge as a cognitive facet of the professional competence of pre-service and in-service teachers
Test type: Open and closed item formats: short answers and multiple-choice items
Modality: Paper-pencil-test; online use also possible
Test structure:
Item pool: 18 items; test design matrix with 4 content dimensions: pedagogical knowledge about dealing with heterogeneity, structuring lessons, class leadership and motivation, performance assessment and 3 dimensions of cognitive demands (remembering, understanding and analyzing, creating)
Test versions: Published short version for the first phase of teacher training (König & Blömeke, 2010). Upon request, a short version for the second phase of teacher training or for in-service teachers is available as well.

TEST QUALITY CRITERIA

Reliability: EAP-Reliability .76 or .88 (overall score of the short version for teacher education)

Validity:

Test content: Findings are compliant with theory when comparing pre-service teachers at different educational stages (at university, in training).

Response processes: —

Internal test structure: The theoretically assumed differentiations of the content and cognitive requirement dimensions were statistically confirmed.

Relationships with other competence indicators: Pedagogical knowledge correlates more weakly with content knowledge than with pedagogical content knowledge in mathematics, German and English (König et al, 2018); pedagogical knowledge correlates weakly with the school leaving grade; expected correlations with other facets of professional competence (pedagogical content knowledge, content knowledge) and correlations with video-based assessments; expected correlations with learning opportunities and motivation (see overview in König 2014).

Consequences of testing: Prognostic validity is shown for in-service teachers through positive correlations between the pre-service teachers' test results and their students' ratings of the teaching quality.

Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 20 minutes for the short version

Testing materials: Test booklets or computers or laptops

Special features: Consistent administration under observation required for paper-pencil test or online test; answers to open-response test questions are coded according to instructions. Answer coding (including double coding of 20% of all questionnaires) must be guided and carried out by trained scorers. The answers will be recoded and interpreted by means of an IRT-scale (if used in rotation design). The results can be reported as an overall score, partial results of individual scales (using the long version) or as competence levels (Klemenz & König, 2019).

Practical example: The published short version (König & Blömeke, 2010) is used for practical purposes, e.g. in the quality campaign 'teacher training for the evaluation of a newly developed teaching-learning module' (König et al. 2018). In various projects (ZuS-QS, LEK, EMW), detailed analyses were carried out on the relation between the pedagogical knowledge of pre-service teachers during their training and specific characteristics of their learning opportunities. In the study 'Learning to Practice – Das Praxissemester auf dem Prüfstand (LtP)', pedagogical knowledge was assessed in a pre-post-test to evaluate the state of learning of pre-service teachers during their practical semester in the state of Nordrhein-Westfalen at three locations. Practical school activities as well as mental support can support the increase in action-oriented pedagogical knowledge in particular (requirement "creation") (König et al. 2020).

DOMAIN

Tested for: Pedagogical knowledge in different subjects

Suitable for: —

Not suitable for: —

TARGET GROUP

Tested for: Students at the beginning of teacher education, during teacher education; pre-service teachers in their practical training phase; in-service teachers

Suitable for: —

Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing the level and development of pedagogical knowledge during and after teacher education

Suitable for: —

Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Instruments for subject-didactic knowledge, "pedagogical knowledge for inclusive teaching" (König et al. 2017)

Suitable for: Further measures of competence development

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at course level, university level, state and national level, international level.

Suitable for: Individual diagnostics

Not suitable for: —

APPLICATION SCENARIOS

Tested for: University comparisons, university evaluation; relations to external criteria; prediction of professional success criteria (e.g. teaching quality); formation of competence levels

Suitable for: Individual development processes

Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —

Suitable for: —

Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

In 2018 and subsequent years, the test was used in various empirical studies; a short version has been available for interested researchers since 2010; further versions (including a long version of the test) are being prepared for publication.

ADDITIONAL INFORMATION

Projects in which the test has been used:
CME – Classroom Management Expertise; DIDAKTUM – Didactic knowledge and professional motivation of student teachers; EMW – Development of occupation-specific motivation and pedagogical knowledge in teacher training; LEK – Longitudinal survey of pedagogical competences of student teachers; LEK-R – Longitudinal survey of pedagogical competences of student teachers; LtP – Learning to Practice – The practical semester under the microscope; PKE – Professional competence of English teachers: Specialized didactic knowledge of prospective English teachers – conception, measurement, validation; PlanvoLL-D – The importance of professional knowledge of prospective German teachers for their planning of teaching; SKILL – Study on competence development in teacher training;
TEDS-M – Teacher Education and Development Study; TEDS-FU – Teacher Education and Development Study Follow up; TEDS Teaching – Teacher Education and Development Study – Teaching; TEDS Validation – Teacher Education and Development Study – Validation; ZuS-QS – Future strategy teacher training (ZuS), subproject quality assurance

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerausbildung. *Zeitschrift für Erziehungswissenschaft*, 12 (3), pp. 499–527. [Open Access: https://doi.org/10.1007/s11618-009-0085-z]

König, J. & Blömeke, S. (2010). Pädagogisches Unterrichtswissen (PUW). Dokumentation der Kurzfassung des TEDS-M-Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung. Berlin: Humboldt-Universität.

König, J.; Blömeke, S.; Paine, L.; Schmidt, B. & Hsieh, F-J. (2011). General Pedagogical Knowledge of Future Middle School Teachers. On the Complex Ecology of Teacher Education in the United States, Germany, and Taiwan. *Journal of Teacher Education*, 62 (2), pp. 188–201.

SECONDARY LITERATURE

Klemenz, S. & König, J. (2019). Modellierung von Kompetenzniveaus im pädagogischen Wissen bei angehenden Lehrkräften: Zur kriterialen Beschreibung von Lernergebnissen der fächerübergreifenden Lehramtsausbildung. *Zeitschrift für Pädagogik, 65*(3), 355-377.

König, J.; Doll, J.; Buchholtz, N.; Förster, S.; Kaspar, K.; Rühl, A.-M.; Strauß, S.; Bremerich-Vos, A.; Fladung, I. & Kaiser, G. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? Struktur des professionellen Wissens bei angehenden Deutsch-, Englisch- und Mathematiklehrkräften im Studium. *Zeitschrift für Erziehungswissenschaft*, *21*(3), pp. 611–648. [Open Access: https://doi.org/10.1007/s11618-017-0765-z]

König, J. Gerhard, K. Melzer, C. Rühl, A.-M. Zenner, J. & Kaspar, K. (2017). Erfassung von pädagogischem Wissen für inklusiven Unterricht bei angehenden Lehrkräften: Testkonstruktion und Validierung. *Unterrichtswissenschaft*, 45 (4), pp. 223–242.

Lesson Planning for History Teachers

SUMMARY

Name: Test for assessing pre-service history teachers' didactic planning knowledge in the classroom

Domain: History education

Assessed competencies: Didactic knowledge for history lesson planning

Target group: Students of history education (bachelor and graduate students), pre-service teachers in their practical phase, in-service teachers

Test type: Multiple-choice test, forced-choice test

Modality: Paper-pencil-test; online test

Duration: 60 minutes

Test structure: 165 items (dichotomous) in 4 content dimensions: goals and principles, phasing and structuring, media, work orders as well as in 2 cognitive requirement dimensions or types of knowledge: factual and conceptual knowledge, procedural knowledge. In addition, study and social data (10 items), epistemological beliefs (25 items, 6-point Likert scale) and learning opportunities (60 dichotomous items)

General test purpose: Assessing the competence level and development; formative and summative assessment, evaluation, small-scale assessment, large-scale assessment; repeated testing

Application scenarios: Course comparisons; university comparisons; output: study and teaching design; teaching-learning tool (self-evaluation)

Not suitable for: Primary or secondary education; international level

Note for practical use: —

Applied in projects: Future Social Sciences Teachers' Competencies. Shaping and assessing historical-didactical competencies of prospective history teachers (SOSCIE); http://www.ruhr-uni-bochum.de/histdidaktik/Forschung/soscie.html

Contact / Location:
Prof. Dr. Nicola Brauch
University of Bochum
Email: nicola.brauch@ruhr-uni-bochum.de

GENERAL INFORMATION

Name: Test for assessing pre-service history teachers' didactic planning knowledge in the classroom
Applied in projects: Future Social Sciences Teachers' Competencies. Shaping and assessing historical-didactical competencies of prospective history teachers (SOSCIE); http://www.ruhr-uni-bochum.de/histdidaktik/Forschung/soscie.html
Contact / Location:
Prof. Dr. Nicola Brauch
University of Bochum
Email: nicola.brauch@ruhr-uni-bochum.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Didactic knowledge for history lesson planning
Theoretical model: Model of professional competence of teachers (Baumert & Kunter, 2006)
Test type: Multiple-choice test, forced-choice test
Modality: Paper-pencil-test, online test
Test structure:
Item pool: 165 items (dichotomous) in 4 content dimensions: goals and principles, phasing and structuring, media, work orders as well as in 2 cognitive requirement dimensions or types of knowledge (factual and conceptual knowledge, procedural knowledge). In addition, study and social data (10 items), epistemological beliefs (25 items, 6-point Likert scale) and learning opportunities (60 dichotomous items). The detailed structure can be found below.

TEST QUALITY CRITERIA

Reliability: EAP/PV reliabilities: 0.870 (overall construct); knowledge areas: [1]: 0.711, [2] 0.732, [3] 0.77, [4] 0.739; types of knowledge: [5] 0,777; [6] 0,862
Validity:
Test content: Systematic identification of item contents from historical didactic theory for lesson planning; adaptation of task complexity to taxonomy by Anderson & Krathwohl (2014); expert workshops.
Response processes: Viewed in CogLabs.
Internal test structure: The fit of the data (model fit after Rasch scaling) shows significantly better values for four theoretically assumed content dimensions and two types of knowledge than for the models with one dimension each.
Relationships with other competence indicators: Small to moderate significant correlations of test results with school grades and attended courses in line with expectations.

Consequences of testing: Significant correlations of the test performance in content dimensions in line with expectations with the position of the subjects throughout the study ($F(2,362) = 30,248$, $p < .001$, $\eta 2 = .143$). The correlation with B.A. grades shows moderate prognostic quality and the scoring quality of the test.
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 60 minutes
Testing materials: Pen and test booklet or online access
Special features: The psychometric data can be broken down in a statistically precise way.
Practical example: The test was piloted from May to July 2015 at five universities (N=272 students). Systematic user feedback was obtained in cognitive labs. The revised test was piloted again at the University of Bochum in the winter semester 2015/16 (N= 365 students). The study examined the level of knowledge of beginners in the bachelor's program in history in the Integrated Proseminars (IPS), beginners in the master's phase of the Master of Education ("introductory seminar") and advanced master's students towards the end of their studies ("practical seminar"). The results point to a significant increase in the knowledge of historical didactic teaching planning over the course of studies. The generally weaker test performance in procedural knowledge is conspicuous and knowledge does not increase in the same way in all content-related dimensions.

DOMAIN

Tested for: History education
Suitable for: —
Not suitable for: —

TARGET GROUP

Tested for: Students of history education (first-year students in bachelor to master degrees in integrated proseminars, introductory seminars (M.Ed.) and practical seminars (M.Ed.))
Suitable for: Pre-service teachers in their practical phase, in-service teachers
Not suitable for: Primary and secondary education

GENERAL TEST PURPOSE

Tested for: Assessing the level of competence; formative and summative assessment, low-scale assessment, evaluation
Suitable for: Assessing competence development, large-scale assessment, repeated testing
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Study and social data, epistemological beliefs, learning opportunities
Suitable for: Other measures of competence
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for the course level.
Suitable for: Individual students, pre-service teachers at university level, 2nd phase teacher training seminars, generally also suitable for the state and national level, international level
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Course comparisons; university comparisons; output: study and teaching design
Suitable for: Teaching-learning tool (self-evaluation)
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Optimization of teaching and learning. Students' significantly lower test performance in the area of procedural knowledge calls for expanding learning opportunities for the practical application of the theoretically taught contents, as well as expanding the framework to accommodate development of learning tasks.
Suitable for: —
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test is currently being used in a dissertation project on the development of knowledge for teaching planning with pre-service history teachers (and trainee lawyers); and its content is being shortened. In addition to the paper version, the test is also available as an online version for various groups (students, pre-service teachers in their practical phase). The test can be requested from the authors of the study (see contact above).

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

ADDITIONAL INFORMATION

Test structure of the instrument

Content dimensions (knowledge areas)	Cognitive requirement dimensions (Knowledge types)	
	Factual and Conceptual Knowledge (FAK) (87 items in total)	**Procedural Knowledge (PR)** (78 items in total)
Goals and principles (54 items in total)	50 items for: historical awareness, historical culture, contemporary relevance, multi-perspectivity, problem orientation, source orientation	4 items for: problem orientation in history teaching
Phasing and structuring (41 items in total)	20 items for: entry phase, development phase, structuring concepts	21 Items for: functions of different entries, assignment of series titles to structuring concepts
Media (41 items in total)	17 items for: text sources, image sources	24 items about: possibilities of using different teaching materials in a specific lesson
Prompts (29 items in total)	0 Items	29 items on: development of image sources, prompts to promote methodological or judgmental competence; requirement areas for specific prompts, adaptation of learning activities to topic formulation

SELECTED PROJECT PUBLICATIONS

Wolf, J.; Seiffert, J.; Seifert, A.; Rothland, M. & Brauch, N. (2018). Das geschichtsdidaktische Planungswissen von angehenden Geschichtslehrer/innen. Entwicklung und Pilotierung eines Testinstruments zur Messung des fachdidaktischen Planungswissens von Lehramtsstudierenden im Fach Geschichte. In: *Geschichte in Wissenschaft und Unterricht* 69, 7/8, pp. 373–392.

Wolf, J.; Seiffert, J.; Seifert, A.; Rothland, M. & Brauch, N. (2017). Unterrichtsplanung im Fach Geschichte – Modellierung und Erfassung geschichtsdidaktischen Planungswissens von Lehramtsstudierenden. In: Wernke, S. & Zierer, K. (eds.): *Die Unterrichtsplanung: Ein in Vergessenheit geratener Kompetenzbereich? Status Quo und Perspektiven aus Sicht der empirischen Forschung*, Bad Heilbrunn: Klinkhardt, pp. 62–76.

SECONDARY LITERATURE

Anderson, L. W.; Krathwohl, D. R.; Airasian, P. W.; Cruikshank, K. A.; Mayer, R. E.; Pintrich, P. R.; Raths, J. & Wittrock, M. C. (2014). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's.* Harlow, Essex: Pearson.

Media-Pedagogical Competence

SUMMARY

Name: Test for measuring media-pedagogical competence

Domain: Trainee teachers for all subjects and school types

Assessed competencies: Media-pedagogical competence in the differentiation of media didactics, media pedagogy, media-related school improvement, media-related attitudes, media-related expectations of self-efficacy, media-technical knowledge

Target group: Trainee teachers in advanced semesters

Type of test: Multiple-choice test

Modality: Paper-pencil test

Duration: 60 minutes

Test structure: 102 items: 3 subscales on media-pedagogical competencies and on media-pedagogical attitudes (media didactics, media pedagogy, media-related school development), 1 subscale on media pedagogical self-efficacy and 1 scale on media technical knowledge

General test purpose: Assessing the level and development of media-pedagogical competence (as generic interdisciplinary competence)

Application scenarios: Assessments of the acquisition of competence by pre-service teachers in the university and practical training phase; evaluation of media-pedagogical teaching offers

Not suitable for: General media literacy; subject-specific aspects of media-pedagogical literacy

Note for practical use: Developed for use by student teachers or pre-service teachers; in particular for the evaluation of measures to promote media-pedagogical skills

Applied in projects: Modelling and Measurement of Media Pedagogical Competence (M^3K)

Contact / Location:
Prof. Dr. Bardo Herzig
Paderborn University
Email: bardo.herzig@upb.de

GENERAL INFORMATION

Name: Test for measuring media-pedagogical competence
Applied in projects: Modelling and Measurement of Media Pedagogical Competence (M^3K)
Contact / Location:
Prof. Dr. Bardo Herzig
Paderborn University
Project report deposited in the catalogue of the Technical Information Library (TIB) Hanover.
Email: bardo.herzig@upb.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Media didactics, media pedagogy, media-related school development, media-related attitudes, media-related expectations of self-efficacy, media-technical knowledge
Theoretical model: Concepts for media-pedagogical competence (Blömeke, 2000, Gysbers, 2008, Tulodziecki, Herzig & Grafe, 2010); Technological Pedagogical Content Knowledge (TPACK) model (Mishra & Koehler, 2006)
Test type: Multiple-choice test
Modality: Paper-pencil test
Test structure:
102 items; 3 partial dimensions with the subscales media didactic competence (16 items), media-pedagogical competence (14 items), media-related school improvement competence (10 items), media-technical knowledge (26 items), media-related self-efficacy (18 items), media-related beliefs and attitudes (18 items)

TEST QUALITY CRITERIA

Reliability: Internal consistency (Cronbach's α) for sample with explicit learning opportunities: α =.58 media didactics, α =.59 media education, α =.55 media-related school development
Validity:
Test content: The test contents were derived from a structured theoretical model that was developed and empirically validated in expert interviews
Response processes: Considered in individual exploratory cognitive pretests
Internal test structure: Examination of the model with three dimensions (media-didactic competence, media-pedagogical competence, media-related school improvement competence) against a one-dimensional model shows a better fit
Relationships with other competence indicators: —
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 60 minutes when using all subscales
Testing materials: Copied templates of the test items
Special features: —
Practical example: The scales were tested on a sample of 912 teacher education students at different universities in Germany. The test can be used to evaluate the effectiveness of learning opportunities to acquire media-pedagogical competence; for this purpose, surveys were conducted at three universities in media-pedagogical courses. The test was also translated into American English and used in the USA (N=109, 7 universities and colleges).

DOMAIN

Tested for: Teacher training students of all subjects and school types
Suitable for: Pre-service teachers
Not suitable for: —

TARGET GROUP

Tested for: Student teachers in advanced semesters
Suitable for: —
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing the level of competence and competence development; evaluation of teaching-learning courses
Suitable for: —
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: —
Suitable for: Further measures of competence assessment
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for individual students over time, course level, university level.
Suitable for: International level: First adaptation for the USA is validated.
Not suitable for: —

APPLICATIONS

Tested for: Evaluation of the impact of media-pedagogical learning opportunities in teacher training, evaluation of media-pedagogical teaching offers
Suitable for: —
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Assessment of study and work performance; identification of starting points for improving media-pedagogical learning opportunities
Suitable for: —
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

If you are interested in the instrument, please contact Prof. Dr. Bardo Herzig: bardo.herzig@upb.de.

SELECTED PROJECT PUBLICATIONS

Herzig, B.; Martin, A.; Schaper, N. & Ossenschmidt, D. (2015). Modellierung und Messung medienpädagogischer Kompetenz – Grundlagen und erste Ergebnisse. In: Koch-Priewe, B.; Köker, A.; Seifried, J. & Wuttke, E. (eds.): *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte*, pp. 153–176. Bad Heilbrunn: Verlag Julius Klinkhardt.

Herzig, B. & Martin, A. (2018). Lehrerbildung in der digitalen Welt – konzeptionelle und empirische Aspekte. In: Knopf, J.; Ladel, S. & Weinberger, A. (eds.): *Digitalisierung und Bildung*, pp. 89–113. Wiesbaden: Springer VS Verlag.

Tiede, J.; Grafe, S. & Hobbs, R. (2015). Pedagogical Media Competences of Pre-Service Teachers in Germany and the United States of America: A Comparative Analysis of Theory and Practice. In: Zlatkin-Troitschanskaia, O. & Blömeke, S. (eds.), Modeling and Measuring Competencies in Higher Education. A Special Issue on the Assessment of Competencies in Higher Education. *Peabody Journal of Education*, 90:4, pp. 533–545.
DOI: http://dx.doi.org/10.1080/0161956X.2015.1068083

SECONDARY LITERATURE

Herzig, B.; Schaper, N.; Breiter, A.; Hartig, J.; Grafe, S.; Martin, A.; Christoph, G.; Ossenschmidt, D.; Brüggemann, M. & Langen, T. (2013). M^{3}K – Modeling and Measuring Pedagogical Media Competencies of Pre-Service Teachers. In: Blömeke, S. & Zlatkin-Troitschanskaia, O. (Eds.): *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, pp. 55–57. Berlin & Mainz: Humboldt University & Johannes Gutenberg University.

Tiede, J. & Grafe, S. (2015). Media Pedagogy in German and U.S. Teacher Education. *Comunicar*, 49(14), pp. 19–29.
DOI: http://dx.doi.org/10.3916/C49-2016-02

Blömeke, S. (2000). *Medienpädagogische Kompetenz*. München: kopäd.

Multiple Document Comprehension (MDC)

SUMMARY

Name: MDC test

Domain: All subjects

Assessed competencies: Multiple Document Comprehension (MDC)

Target group: All German students

Test type: Performance test: single-choice & complex multiple-choice

Modality: Computer-based

Duration: 60 minutes (for 2 units)

Test structure: 67 items; 5 units with 4 different requirements

General test purpose: One-time assessment of competence level; competence development through repeated testing

Application scenarios: Competence level of student groups; identifying teaching potentials; evaluation of measures for promoting MDCs

Not suitable for: —

Note for practical use: Developed for use with students; successfully used with students; individual or group tests are possible.

Applied in projects: Computer-based measurement of students' text comprehension of multiple documents (MDC) at four German universities within the scope of the MultiTex project (Process-based assessment of multiple documents comprehension): http://multitex.bildungsforschung.uni-bamberg.de

Contact / Location:
Dr. Cornelia Schoor
Otto-Friedrich University of Bamberg
Email: cornelia.schoor@uni-bamberg.de

GENERAL INFORMATION

Name: MDC-Test
Applied in projects: Computer-based measurement of students' text comprehension of multiple documents (MDC) at two German universities within the scope of the MultiTex project (Process-based assessment of multiple documents comprehension): http://multitex.bildungsforschung.uni-bamberg.de
Contact / Location:
Dr. Cornelia Schoor
Otto-Friedrich University of Bamberg
Email: cornelia.schoor@uni-bamberg.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Multiple Document Comprehension (MDC)
Theoretical model: Documents Model Framework (Britt & Rouet, 2012), Wineburg (1991), Multiple-Document Task-Based Relevance Assessment and Content Extraction (MD-TRACE, Rouet & Britt, 2011)
Test type: Performance test; single-choice & complex multiple-choice
Modality: Computer-based
Test structure:
Item pool: 67 items; 5 units with 4 different requirements: 1. comparing content across texts, 2. linking content across texts, 3. assessing and comparing sources (source-source links), 4. representing content as generated by sources and comparing source content combinations (source-content links); 2 out of 5 units are administered; different test versions are available by combining the units in different ways.

TEST QUALITY CRITERIA

Reliability: EAP reliability = .69 or WLE reliability = .67; Infit values of the items: 0. 91–1.10; Outfit values: 0. 74–1.14
Validity:
Test content: The tasks were developed using the theoretical approaches of the Documents Model Framework (Britt & Rouet, 2012) and the strategies of Wineburg (1991).
Response processes: The evaluation of the history data so far (so-called log data) is consistent with the theoretically assumed response processes.
Internal test structure: As assumed, according to dimensional analyses, the items represent a one-dimensional capability construct.
Relationships with other competence indicators: The school leaving grade correlated positively and significantly with the MDC test score ($r=-.44$, $p < .001$). Master students achieved significantly higher MDC test scores than bachelor students.
Consequences of testing: A connection between MDC and study success is assumed, but has yet to be verified.
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 60 minutes (for 2 units)
Testing materials: Computer
Special features: Each participant receives only 2 of 5 units due to time restrictions.
Applied in projects: Two studies (N=310; N= 508) were carried out in which students at two universities worked on the test to assess their MDC skills. In the second study, tests of working memory, reading literacy and reading speed were added.
Practical example: The test can be used to assess students' level of MDC competence. Based on the assessment results, potentials for fostering students' MDC can be identified and suitable teaching materials can be developed. Individual feedback for students is possible. Feedback of test users is still to be examined.

DOMAIN

Tested for: Social sciences and humanities
Suitable for: All subjects
Not suitable for: —

TARGET GROUP

Tested for: Bachelor and master students between 17 and 42 years; two federal states (Bavaria and Hesse)
Suitable for: All higher education students in Germany
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing students' text comprehension of multiple documents (MDC); one-time testing
Suitable for: Competence development during repeated testing
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Epistemic beliefs, goal orientations, mental load, mental effort, task model, reading frequency and habits, basic computer skills, working memory, final grades, interests, self-assessed prior knowledge, reading literacy and speed, attitudes towards reading
Suitable for: No restrictions
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for comparisons within the tested group of students.
Suitable for: —
Not suitable for: —

APPLICATIONS

Tested for: Level of competence of student groups for research purposes
Suitable for: Identifying potentials and evaluating measures for fostering MDC
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Developing measures for improvement
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test will be further validated and developed. A current version of the test with corresponding documentation is being prepared for publication.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Schoor, C.; Hahnel, C.; Artelt, C.; Reimann, D.; Kröhne, U. & Goldhammer, F. (2020). Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses multipler Dokumente von Studierenden [Developing and Scaling a Test of Multiple Document Comprehension in University Students]. *Diagnostica*. doi: 10.1026/0012-1924/a000231

Hahnel, C.; Kröhne, U.; Goldhammer, F.; Schoor, C.; Mahlow, N. & Artelt, C. (2019). Validating process indicators of sourcing in an assessment of multiple document comprehension. *British Journal of Educational Psychology, 89*(3), 524-537. doi: 10.1111/bjep.12278

Hahnel, C.; Schoor, C.; Kröhne, U.; Goldhammer, F.; Mahlow, N.; & Artelt, C. (2019). The role of cognitive load for university students' comprehension of multiple documents. *Zeitschrift für Pädagogische Psychologie, 33*(2), 105-118. doi: 10.1024/1010-0652/a000238

SECONDARY LITERATURE

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

Rouet, J. & Britt, M. A. (2012). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (eds.), *Text relevance and learning from text*, pp. 19–52. Charlotte, NC: Information Age.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), pp. 73–87. doi: 10.1037/0022-0663.83.1.73

My Learning Diary

SUMMARY

Name: My Learning Diary

Domain: All subjects

Assessed competencies: Competencies of self-regulated learning (SRL) (motivation, self-efficacy, planning, goal setting, reflection) and mood

Target group: University students during exam preparation, students from secondary level onwards during exam preparation

Test type: Diary

Modality: Smartphone app (online)

Duration: 5 minutes per entry unit

Test structure: 20 items; combination of closed- and open-response questions

General test purpose: Assessing SRL competencies over time within 2 weeks or over a different period of time

Application scenarios: Teaching-learning tool, intermediate exams, final exams, course comparisons, university comparisons

Not suitable for: Primary school students, one-time tests, international level, selection of admissions, prediction of study success, prediction of discontinuation of studies, university evaluation, accreditation, rankings, assessment of previous educational levels, professional aptitude, prediction of performance development

Note for practical use: A participant code and password must be generated for each user. The resulting data needs to be preprocessed before statistical analyses can be carried out. Results are analyzed using time-series-analyses.

Applied in projects: Product- and Process Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation (PRO-SRL-EVA);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/pro-srl-eva/

Contact / Location:
Prof. Dr. Bernhard Schmitz
University of Darmstadt
Email: schmitz@psychologie.tu-darmstadt.de
Website: https://www.paedpsy.psychologie.tu-darmstadt.de/team_paedpsy/aktuelles_team_paedpsy/schmitz/schmitz.de.jsp

GENERAL INFORMATION

Name: My Learning Diary
Applied in projects: Product- and Process Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation (PRO-SRL-EVA);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/pro-srl-eva/
Contact / Location:
Prof. Dr. Bernhard Schmitz
University of Darmstadt
Email: schmitz@psychologie.tu-darmstadt.de
Website: https://www.paedpsy.psychologie.tu-darmstadt.de/team_paedpsy/aktuelles_team_paedpsy/schmitz/schmitz.de.jsp

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Competencies of self-regulated learning (SRL) (motivation, self-efficacy, planning, goal setting, reflection) and mood
Theoretical model: SRL model following Schmitz and Wiese (2006)
Test type: Diary
Modality: Smartphone app (online)
Test structure:
Item pool: 20 items, closed and open-response questions

TEST QUALITY CRITERIA

Reliability: —
Validity:
Test content: —
Response processes: —
Internal test structure: According to factor analyses, the variables measured are part of self-regulated learning as an overarching factor.
Relationships with other competence indicators: Significant correlations to other SRL tests (e.g. with the SRL questionnaire, Bellhäuser, Roth & Schmitz, 2015; r=.68**, SJT to assess motivational regulation strategies; r=.54**).
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 5 minutes per unit
Testing materials: Smartphone
Special features: A participant code and password must be generated for each user. The resulting data needs to be preprocessed before statistical analyses can be carried out. Results are analyzed using time-series-analyses.
Practical example: The learning diary in its first version was tested with psychology students (N=48) in 2016. It was tested again in its revised version with further students (N=75) of the TU Darmstadt over a period of 14 days for each student in 2017. The students received individual feedback, e.g. on their motivation development over time and on the extent to which they had achieved their personal goals. In 2018, the diary was revised again, and further variables were integrated (motivational regulation competencies and planned vs. actual use of learning strategies). It is currently used in the PRO-SRL-EVA project by psychology and mathematics students at the TU Darmstadt, the University of Augsburg and the University of Vienna. Practical consequences will be weighted after the scoring has been completed.

DOMAIN

Tested for: Mathematics and psychology
Suitable for: All courses
Not suitable for: —

TARGET GROUP

Tested for: University students during exam preparation
Suitable for: Students at secondary school level during exam preparation
Not suitable for: Primary school students

GENERAL TEST PURPOSE

Tested for: Assessing SRL competencies over the course of 2 weeks
Suitable for: Other time periods
Not suitable for: One-time tests

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Situational Judgment Test to assess motivational regulatory competence, SRL-QuAK
Suitable for: All SRL variables
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for the course level, university level.
Suitable for: Individuals
Not suitable for: State and national level, international level

APPLICATION SCENARIOS

Tested for: Teaching-learning tool, intermediate exams, final exams, course comparisons, comparisons of learning behavior between different groups of students (e.g. standard time vs. long-term students)
Suitable for: University comparisons
Not suitable for: Admission selection, prediction of study success, prediction of termination of studies, university evaluation, accreditation, rankings, assessment of previous educational levels, professional aptitude, prediction of performance development

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Improving the quality of learning
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The app can possibly be made available free of charge to all interested students so that they can better monitor their own learning process and success. However, this requires regular maintenance by an IT specialist.

ADDITIONAL INFORMATION

The procedure is not language-neutral, since the items for assessing the SRL variables are in German. Use in other languages requires a translation and validation process.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Schmitz, B. & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology*, 31(1), pp. 64–96. http://dx.doi.org/10.1016/j.cedpsych.2005.02.002

Performance-Based Measuring of Students' Communication Skills

SUMMARY

Name: Performance-based Measuring of Students' Communication Skills

Domain: University graduates of all disciplines

Assessed competencies: Strategic and communication-oriented communication skills

Target group: Students of all disciplines in Germany

Test type: Standardized performance-based role plays with closed observation sheets

Modality: Role playing, paper-and-pencil observation sheet

Duration: 45 minutes

Test structure: 20 role play scenarios; 5 scenarios each for strategic and communication-oriented communication skills; each scenario in a version for corporate or educational contexts

General test purpose: Assessing communicative skills using theoretically founded observation criteria

Application scenarios: Competence-oriented examination, module or study completion, training courses

Not suitable for: —

Note for practical use: High training effort for interlocutors and observers who need to be familiar with the theoretical framework.

Applied in projects: In courses for students and lecturers (university didactics) as training; in the project with 488 students as performance-based assessment; also as university-didactic training with lecturers and in courses with students

Contact / Location:
Prof. Dr. Edith Braun
Justus-Liebig University of Giessen
Email: edith.braun@uni-giessen.de

GENERAL INFORMATION

Name: Performance-based Measurement of Students' Communication Skills
Applied in projects: Performance based assesment of communication skills of students (KomPrü);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/komprue/
Contact / Location:
Prof. Dr. Edith Braun
Justus-Liebig University of Giessen
Email: edith.braun@uni-giessen.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Strategic and communication-oriented communication skills
Theoretical model: Strategic and communication-oriented communication according to Habermas
Test type: Performance-based role plays with standardized instruction and observation sheet
Modality: Role playing, paper-and-pencil observation sheet
Test structure:
Item pool: 20 role play scenarios; 5 scenarios each for strategic and 5 for communication skills, each for the context 'company' or 'educational institution'. Each role-playing scenario consists of an instruction sheet for the test participant and a standardized observation sheet. Each test requires a trained interlocutor and a trained observer who is familiar with the theoretical construct and the scenarios (corresponds to the test setting of oral exams with examiner and test administrator). The test participants are given 5 minutes preparation time per role-play scenario and up to 10 minutes to enact the situation.
Versions: Different combinations of 4 role play scenarios each

TEST QUALITY CRITERIA

Reliability: Rho for intercorrelated factors (Chao, 2016) = .80 for strategic communication and .84 for communication oriented towards understanding
Validity:
Test content: The situations of the role play scenarios were determined by means of a graduate survey (N= 10,000). All scenarios were discussed in a panel of experts. The role play scenarios were ascribed a high degree of authenticity.
Internal test structure: A confirmatory factor analysis confirms the hypothesis of a two-dimensional model ($\chi 2 = 103.5$; $df = 53$; $p < 0.001$; CFI = 0.95; RMSEA = 0.04).

Relationships with other competence indicators: Strategic communication correlates with para- and nonverbal communication (r = 0.60). For communication oriented towards understanding, there is a correlation of r = 0.57.
Consequences of testing: The students report high authenticity of the role play scenarios regarding professional situations.
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 45 min for four role plays
Testing materials: Instruction and observation sheet
Special features: Trained personnel (interlocutor, observer), high-effort administration for high ecological validity
Practical example: The questionnaire was administered to 488 students at 10 universities throughout Germany. The students were not familiar with the role play scenarios in advance and were able to perform them well. The role play scenarios are used in teacher training at the Freie Universität Berlin, Pädagogische Hochschule Luzern, and University of Gießen. In higher education teaching, the scenarios are used as training tools.

DOMAIN

Tested for: University graduates from all subject groups, focus on teacher education and economics
Suitable for: Lecturers in higher education teaching methodology courses
Not suitable for: —

TARGET GROUP

Tested for: Students of all subject groups in Germany
Suitable for: Lecturers and university graduates, students, graduates and teachers in other countries (due to cultural specificity, retesting in the target country is necessary)
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Observation-based assessment of strategic and communication-oriented communication
Suitable for: —
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: In partial samples further scales were used for validation purposes:
Scales of nonverbal communication: Spitzberg, B. H. & Adams, T. W. (2007). *CSRS, the Conversational Skills Rating Scale: An Instructional Assessment of Interpersonal Competence*. NCA, National Communication Association.
Short version of the WiWiKom test: see WiWiKom instrument (page 21).
Scales of pedgogical teaching knowledge: König, J. & Blömeke, S. (2010). Pädagogisches Unterrichtswissen: (PUW). *Dokumentation der Kurzfassung des TEDS-M Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung*. Humboldt-Universität zu Berlin.
Suitable for: Further measures of competence assessment
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Individual diagnostics
Suitable for: —
Not suitable for: —

APPLICATIONS

Tested for: Trainings without rating
Suitable for: —
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Feedback on the level of competence
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

There is currently an exchange with the Legacy Project in Great Britain.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Braun, E.; Athanassiou, G; Pollerhof, K. & Schwabe, U. (2018). Konzeption einer kompetenzorientierten Prüfung kommunikativer Fähigkeiten. *Beiträge zur Hochschulforschung*, pp. 34–55.
Braun E., Schwabe U., Klein D. (2018). Performance-Based Tests: Using Role Plays to Assess Communication Skills. In: McGrath S., Mulder M.,Papier J., Suart R. (eds) *Handbook of Vocational Education and Training*. Cham: Springer.

Physics Teachers' Content Knowledge (CK) in Mechanics

SUMMARY

Name: Test to assess content knowledge (CK) in physics (Profile-P-CK)

Domain: Physics education – mechanics

Assessed competencies: Physics knowledge in mechanics

Target group: Students of physics and physics education, pre-service teachers in their practical phase, in-service teachers

Test type: Multiple-choice & single-select test

Modality: Paper-pencil test

Duration: 50 minutes

Test structure: 48 items; 3 subscales: school knowledge (20 items), advanced knowledge (14 items), university knowledge (14 items)

General test purpose: Assessing competence levels and development, repeated testing

Application scenarios: University evaluation, demand planning for teaching, course comparisons, study success prediction, teaching-learning tool, accreditation

Not suitable for: International level, intermediate and final examinations, rankings, assessment of previous educational levels, professional aptitude, admission to studies, study choice counseling, examination or final grade, career counseling

Note for practical use: Testing time is fixed. Those interested in testing should contact the coordination of the Profile-P+ project: christoph.vogelsang@upb.de.

Applied in projects: Professional Knowledge of Physics Student Teachers (Profile-P; Profile-P+); https://www.uni-due.de/didaktik_der_physik/forschungfischer.php

Contact / Location:
Christoph Vogelsang
Paderborn University
Email: christoph.vogelsang@upb.de

GENERAL INFORMATION

Name: Test to assess content knowledge(CK) in physics (Profile-P-CK)
Applied in projects: Professional Knowledge of Physics Student Teachers (Profile-P; Profile-P+); https://www.uni-due.de/didaktik_der_physik/forschungfischer.php
Contact / Location:
Christoph Vogelsang
Paderborn University
Email: christoph.vogelsang@upb.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Knowledge of physics in mechanics
Theoretical model: Model of professional action competence of pre-service physics teachers (Riese & Reinhold, 2012); content area mechanics for the facets school knowledge, advanced school knowledge and university knowledge
Test type: Performance test; multiple-choice test & single-select
Modality: Paper-pencil test
Test structure:
Item pool: 48 items; 3 subscales: school knowledge (20 items), advanced school knowledge (14 items), university knowledge (14 items)

TEST QUALITY CRITERIA

Reliability: EAP/PV reliability = .84, variance = 64.7 after pilot analyses
Validity:
Test content: Textbook analyses for schools and universities, analyses of the lecture contents of introductory lectures in physics, assignment of tasks to subject levels, discussion of the assignment with subject physicists and teaching methodology experts, content validation by means of consensus coding by students in higher semesters, interview study with physics teachers to determine the relevance for the profession.
Response processes: A think-aloud study conducted to examine the response processes for knowledge facets.
Internal test structure: The dimensionality was examined in an IRT model: The data fit significantly better to the three-dimensional than to the one-dimensional model.
Relationships with other competence indicators: High correlations to a test for mathematical skills, but decreasing from school knowledge to university knowledge; slightly lower correlations to overall school leaving grade and final school grades in mathematics and physics. Very low significant correlations to final school grade in German; positive correlations to didactic physical knowledge and ability to explain physics.
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 50 minutes
Testing materials: Test booklets and pencil
Special features: Testing time is fixed (speed component); a maximum of 48 points can be achieved
Practical example: The test was used in the Profile-P and Profile-P+ projects. In Profile-P it was used cross-sectionally with N = 294 physics education students in three federal states (high proportion of first-year students). The application served to evaluate the effect of the physics teacher education course and to clarify various influencing factors (e.g. acquired credit points in teaching methodology, school leaving grade, study progress) on the level of competence. The test was further developed in Profile-P+ and then used in a longitudinal survey at three points in time in the bachelor program and with master's students before and after their practical semester. The aim of the studies is to determine the development of content knowledge and to analyze connections to the development of pedagogical content knowledge in physics as well as performance in explanatory situations, lesson planning, and lesson reflection.

DOMAIN

Tested for: Teaching physics – mechanics
Suitable for: —
Not suitable for: —

TARGET GROUP

Tested for: Students of physics and teacher education
Suitable for: Pre-service teachers in their practical phase, in-service teachers
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing competence levels and development, repeated testing
Suitable for: Evaluating teaching and learning offers
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Grades, pedagogical content knowledge in physics, performance tests (explaining, planning, reflecting)
Suitable for: Further measures of competence measurement
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for the university level and course level.
Suitable for: —
Not suitable for: International level

APPLICATION SCENARIOS

Tested for: University evaluation, demand planning, teaching, conditional course comparisons
Suitable for: Study success prediction, teaching-learning tool, accreditation
Not suitable for: Intermediate and final examinations, rankings, assessment of previous educational levels, professional aptitude

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Assessing study or work performance
Not suitable for: Admission to studies, study choice counseling, examination or final grade, career counseling

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

In Profile-P+, the test was used to analyze the relation between students' content knowledge of physics and their performance in the standard situations of explaining, planning lessons and reflecting on lessons, as well as to analyze knowledge development primarily in the bachelor's program.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Riese; J.; Kulgemeyer, C.; Zander, S.; Borowski, A.; Fischer, H.; Gramzow, Y.; Reinhold, P.; Schecker, H. & Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In: Blömeke, S. & Zlatkin-Troitschanskaia, O. (Eds.): Kompetenzen von Studierenden. *Zeitschrift für Pädagogik, Supplement 61, pp.* 55–79. Weinheim: Beltz.
Enkrott, P.; Buschhüter, D. & Borowski, A. (2018). Modeling and Development of Professional Content Knowledge of Pre-Service Physics Teachers. In: Maurer, C. (ed.). *Qualitätsvoller Chemie- und Physikunterricht- normative und empirische Dimensionen. Gesellschaft für Didaktik der Chemie und Physik*, Jahrestagung in Regensburg 2017, pp. 896–899. Universität Regensburg.

Physics Teachers' Pedagogical Content Knowledge (PCK) in Mechanics

SUMMARY

Name: Test to assess the pedagogical content knowledge in physics (Profile-P-PCK)

Domain: Physics education – mechanics

Assessed competencies: Physics pedagogical content knowledge in mechanics

Target group: Teacher education students in their bachelor and master studies, pre-service teachers in their practical phase, in-service teachers

Test type: Multiple-choice test and open-response items

Modality: Paper-pencil test

Duration: 65 minutes

Test structure: 43 items; 4 dimensions of knowledge: instruction strategies, student beliefs, experiments, teaching methodology concepts

General test purpose: Assessment of pedagogical content knowledge and its development; repeated testing; individual diagnostics

Application scenarios: University evaluation, demand planning for teaching, course comparisons, prediction of study success, teaching-learning tool, university comparisons, accreditation

Not suitable for: International level, intermediate and final examinations, rankings, assessment of previous educational levels, professional aptitude, prediction of performance development, admission to studies, study choice counseling, examination or final grade, career counseling

Note for practical use: Administration time is fixed (speed component), use according to test administrator manual. When interpreting the test results, users should keep in mind the content focus on mechanics, which is only explicitly taught in the first semesters in most physics teacher education courses.

Applied in projects: Professional Knowledge of Physics Student Teachers (Profile-P, Profile-P+);
https://www.uni-due.de/didaktik_der_physik/forschungfischer.php

Contact / Location:
Dr. Christoph Vogelsang
Paderborn University
Email: christoph.vogelsang@upb.de

Once the project has been completed, the test can be accessed at www.forschungsdaten-bildung.de.

GENERAL INFORMATION

Name: Test to assess pedagogical content knowledge in physics (Profile-P-PCK)
Applied in projects: Professional Knowledge of Physics Student Teachers (Profile-P, Profile-P+);
https://www.uni-due.de/didaktik_der_physik/forschungfischer.php
Contact / Location:
Dr. Christoph Vogelsang
Paderborn University
Email: christoph.vogelsang@upb.de

Once the project has been completed, the test can be accessed at www.forschungsdaten-bildung.de.

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Physics pedagogical content knowledge in mechanics
Theoretical model: Model of professional action competence of pre-service physics teachers and model of physics pedagogical content knowledge (Gramzow, Riese & Reinhold, 2013), and associated item development model (see section "further information")
Test type: Multiple-choice test and open-response items
Modality: Paper-pencil test
Test structure:
43 items (23 open-response, 20 closed). 4 knowledge dimensions: instruction strategies (9 items), student presentations (12 items), experiments (6 items), teaching methodology concepts (16 items)

TEST QUALITY CRITERIA

Reliability: Psychometric requirements for difficulty, selectivity and reliability were met: EAP reliability of the overall score: 0.84 (pilot study); 0.79 (Profile-P main study); intercoder reliability for open-response items: k=0.75 to 0.93 (Profile-P main study).

Validity:
Test content: Item design based on relevant literature; validation by expert survey on curricular fit at four German universities.
Response processes: Tested with 15 participants using think-aloud approach; to correctly answer the items, the test participants draw on pedagogical content knowledge in accordance with their expectations.
Internal test structure: Verification of the item assignment in the model by expert survey; a 4-dimensional model has the best fit to the empirical data in accordance with the theory (pilot study).
Relationships with other competence indicators: Expectation-compliant positive correlations to study credit points in the field of teaching methodology and the final school grade in physics; further validation in Profile-P+.
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION
Duration: 65 minutes
Testing materials: Test booklets and pens
Special features: Testing time is fixed (speed component); test administration according to the test administrator manual. When interpreting the test results, keep in mind the content focus on mechanics, which is only explicitly taught in the first semesters in most physics teacher education courses.
Practical example: The test was used in the Profile-P and Profile-P+ projects. In Profile-P it was used cross-sectionally with N = 294 physics education students in three federal states (high proportion of first-year students). The application served to evaluate the effect of the physics teacher education courses and to identify various influencing factors (e.g. acquired credit points in teaching methodology, school leaving grade, study progress) on the participants' competence levels. The test was further developed in Profile-P+ and then used in a longitudinal survey at three points throughout a bachelor program and with master's students before and after their practical semester. The surveys were carried out in Germany as well as Austria.
The aim of the studies is to determine the development of pedagogical content knowledge in physics as well as performance in explanatory situations, lesson planning, and lesson reflection.

DOMAIN
Tested for: Physics education – mechanics
Suitable for: —
Not suitable for: —

TARGET GROUP
Tested for: Bachelor's and master's students of physics education
Suitable for: Pre-service teachers in their practical phase, in-service teachers
Not suitable for: —

GENERAL TEST PURPOSE
Tested for: Assessing competence level and development, repeated testing
Suitable for: Individual diagnostics, evaluation of teaching-learning offers
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES
Tested for: Grades, content knowledge in physics, knowledge of educational science, performance tests (explaining, planning, reflecting)
Suitable for: Further measures of competence
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY
Tested for: Results are statistically meaningful for the university level and course level.
Suitable for: Individuals over time, state and national level
Not suitable for: International level

APPLICATION SCENARIOS
Tested for: University evaluation, demand planning, teaching, course comparisons
Suitable for: Study success prediction, teaching-learning tool, university comparisons, accreditation
Not suitable for: Intermediate and final examinations, rankings, assessment of previous educational levels, professional aptitude, prediction of performance development

PERMITTED CONSEQUENCES
Tested for: Course evaluation
Suitable for: Assessment of study or work performance, course recommendation
Not suitable for: Admission to studies, study choice counseling, examination or final grade, professional recommendation

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

In Profile-P+, the test is used to analyze the relation between pedagogical content knowledge and the performance of student teachers in the standard situations of explaining, planning lessons and reflecting on lessons, as well as to analyze knowledge development primarily in the bachelor's program.
Enquiries regarding the use of the test for the evaluation of degree programs for the teaching profession in physics have been received.

ADDITIONAL INFORMATION

The test is published in Gramzow (2015), but in Profile-P+ the coding has been further improved.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Gramzow, Y.; Riese, J. & Reinhold, P. (2013). Modellierung fachdidaktischen Wissens angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, pp. 7–30.
Gramzow, Y.; Riese, J. & Reinhold, P. (2013). Prospective physics teachers' pedagogical content knowledge – Validating a test instrument by using a think-aloud study. *European Science Education Research Association 2013 Conference, Nicosia, Cyprus*. http://www. esera. org/media/esera2013/Yvonne_Gramzow_12Feb2014. pdf.
Gramzow, Y. (2015). *Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik*: *Modellierung und Testkonstruktion*. Logos Berlin.

SECONDARY LITERATURE

Riese, J.; Gramzow, Y. & Reinhold, P. (2017). Das fachdidaktische Wissen von Anfängern und Fortgeschrittenen im Lehramtsstudiengang Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 23, pp. 99–112. DOI 10.1007/s40573-017-0059-2.

Practical Requirements for University Graduates

SUMMARY

Name: Assessment of university graduates' perceived practical requirements and tasks

Domain: University graduates of all domains

Assessed competencies: Practical requirements and tasks for university graduates

Target group: University graduates in Germany

Test type: Questionnaire with Likert scales

Modality: Paper-pencil test and online questionnaire

Duration: 7 minutes

Test structure: 41 items; 8 requirement and activity fields: 1. *planning and organization of work processes* (8 items), 2. *supporting others* (3 items), 3. *leadership and management* (5 items), 4. *acting independently in challenging situations* (9 items), 5. *information processing* (6 items), 6. *dealing with numbers* (3 items), 7. *foreign language communication* (4 items), 8. *manual labor* (3 items)

General test purpose: Assessing the frequency of work requirements and performed tasks as perceived by graduates in everyday working life

Application scenarios: University evaluation, accreditation, curriculum development

Not suitable for: Students, individual diagnostics

Note for practical use: —

Applied in projects: Action-related Competencies in Pedagogical Fields (KomPaed); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/kompaed/

Contact / Location:
Prof. Dr. Edith Braun
University of Giessen
Email: edith.braun@uni-giessen.de

Dr. Julia-Carolin Brachem
University of Vechta
Email: julia-carolin.brachem@uni-vechta.de

Contact email: dataservice@dzhw.eu
DOI: https://dx.doi.org/10.21249/DZHW:cmp2014:1.0.0

GENERAL INFORMATION

Name: Assessing university graduates' perceived practical requirements and tasks
Applied in projects: Action-related Competencies in Pedagogical Fields (KomPaed);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/kompaed/
Contact / Location:
Prof. Dr. Edith Braun
University of Giessen
Email: edith.braun@uni-giessen.de
Dr. Julia-Carolin Brachem
University of Vechta
Email: julia-carolin.brachem@uni-vechta.de
Contact email: dataservice@dzhw.eu
DOI: https://dx.doi.org/10.21249/DZHW:cmp2014:1.0.0

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: University graduates' perceived practical requirements and tasks
Theoretical model: Job Requirements Approach, adapted to the German higher education system
Test type: Questionnaire with Likert scales
Modality: Paper-pencil test and online questionnaire
Test structure:
Item pool: 41 items; 8 requirement and activity areas: 1. planning and organization of work processes (8 items), 2. assisting others (3 items), 3. leadership and management (5 items), 4. acting independently in challenging situations (9 items), 5. information processing (6 items), 6. dealing with numbers (3 items), 7. foreign language communication (4 items), 8. manual labor (3 items)

TEST QUALITY CRITERIA

Reliability: Internal consistency among graduates from 2001, 2005 and 2009: planning and organizing of work processes (α = .75), assisting others (α = .81), leadership (α = .72), acting independently in challenging situations (α = .79), information processing (α = .79), dealing with numbers (α = .62), foreign language communication (α = .74), manual labor (α = .77)

Validity:
Test content: The item contents were developed based on literature research and employer interviews
Response processes: Cognitive interviews were conducted to verify the results.
Internal test structure: A confirmatory factor-analysis of the overall model shows results in line with expectations (RMSEA: 0.07, SRMR: 0.08, CFI: 0.74)
Relationships with other competence indicators: —
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION
Duration: 7 minutes
Testing materials: —
Special features: —
Practical example: The questionnaire was used as part of the Germany-wide graduate survey conducted by the German Centre for Higher Education Research and Science Studies (DZHW) (cross-sectional online survey of graduates from 2001, 2005 and 2009 in 2013/2014). The questionnaire enables conclusions about the requirements and tasks that university graduates expect in the labor market. The questionnaire was also used in an in-depth online survey by the DZHW Graduate Panel in 2005 and in a more advanced form as part of the National Educational Panel Study (NEPS) "From Higher Education to the Labor Market" (DZHW 2017a, 2017b; Leibniz-Institute for Educational Trajectories 2018).

DOMAIN
Tested for: For university graduates of all domains (linguistics and cultural studies: N = 2,102; sports: N = 53; law, economics and social sciences: N = 2,826; mathematics, natural sciences: N = 1,926; human medicine, health sciences: N = 570; veterinary medicine: N = 141; agricultural, forestry and food sciences: N = 464; engineering: N = 1,891; arts, art studies: N = 281) and professional fields (military: N = 6; agriculture, forestry, animal husbandry and horticulture: N = 191; raw material extraction, production and manufacturing: N = 1,070; construction, architecture, surveying and building technology: N = 479; natural sciences, geography and information technology: N = 832; transport, logistics, protection and security: N = 122; commercial services, trade in goods, distribution, hotel and tourism: N = 284; business organization, accounting, law and administration: N = 2,205)
Suitable for: —
Not suitable for: Undergraduates students

TARGET GROUP
Tested for: Graduates from German universities
Suitable for: —
Not suitable for: Undergraduates students

GENERAL TEST PURPOSE
Tested for: Assessing the frequency of work requirements and performed tasks as perceived by university graduates in everyday working life
Suitable for: Identification of relevant interdisciplinary key competencies for university graduates for the labor market
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES
Tested for: —
Suitable for: —
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY
Tested for: Group level
Suitable for: —
Not suitable for: Individual diagnostics, international level

APPLICATION SCENARIOS
Tested for: University evaluation, accreditation, curriculum development
Suitable for: —
Not suitable for: —

PERMITTED CONSEQUENCES
Tested for: Curriculum development
Suitable for: —
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE
The questionnaire is available online via the evaluation and survey software "EvaSys" https://www.evasys.de/aktuelles/detail/panthoa-fragebogen-329.html. The data on which the evaluations are based have been made available by the Research Data Center for Higher Education and Science Research as a scientific use file for scientific use (DOI: https://dx.doi.org/10.21249/DZHW:cmp2014:1.0.0).

Do you have any questions or need help, e.g. with practical use or evaluation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS
Braun, E. & Brachem, J.-C. (2018). Erfassung praxisbezogener Anforderungen und Tätigkeiten von Hochschulabsolventinnen/-absolventen (PAnTHoa). *Zeitschrift für Hochschulentwicklung*. 13(1), pp. 209–232.
Braun, E. & Brachem, J.-C. (2017). The labour market's requirement profiles for higher education graduates. In: Kyndt, E.; Donche, V.; Trigwell, K. & Lindblom-Ylänne, S. (Eds.). *Higher Education Transitions. Theory and Research*. London: Routledge, pp. 219–237.
Braun, E. & Brachem, J.-C. (2015). Requirements higher education graduates meet on the labor market. *Peabody Journal of Education*. 90(4), pp. 574–595.

SECONDARY LITERATURE

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2017a). DZHW-Absolventenpanel. https://www.dzhw.eu/forschung/projekt?pr_id=467 (accessed on 20.11.2018).

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2017b). Projekt „Tätigkeitsanforderungen Hochqualifizierter" (JobReq). https://www.dzhw.eu/forschung/projekt?pr_id=596 (accessed on 20.11.2018).

Leibniz-Institut für Bildungsverläufe e. V. (2018). Hochschulstudium und Übergang in den Beruf. https://www.neps-studie.de/studien/hochschulstudium-und-uebergang-in-den-beruf/hochschulstudium-und-uebergang-in-den-beruf (accessed on 20.11.2018).

Reflective Competence of Mathematics Teachers (Math-RC)

SUMMARY

Name: Reflective Competence of Mathematics Teachers (Math-RC)

Domain: Teacher training mathematics

Assessed competencies: Reflective competence (RC) of mathematics teachers

Target group: (Pre-service) mathematics teachers at secondary level I/II (student teachers in mathematics (bachelor, master), trainee teachers and in-service teachers of mathematics); analogue tests were also developed by the authors for use in primary and elementary education

Type of test: Performance test; video-based tasks, open response format

Modality: Computer- and video-based

Duration: 30 minutes

Test structure: 9 items; a scale for planning and reflecting mathematics teaching in the subject areas of secondary algebra (6 items) and analysis (3 items). The test covers problems concerning teaching materials (4 items), planning of first steps (2 items) and reflection on student work (3 items)

General test purpose: Assessing the level and development of competence (with other measures of professional teaching competence: tests of content knowledge (CK), pedagogical content knowledge (PCK); see also the Math-AC test)

Applications scenarios: Comparisons of the level of competence between training phases in the teacher education (bachelor, master, practical training phase, school service); comparisons between pre-service and in-service teachers in the subjects of mathematics and/or business and economics

Not suitable for: Non-German-speaking international level; entry diagnostics; first-year students

Note for practical use: Organizational and technical information in the test administrator manual to ensure standardized test administration; a coding manual is provided for objective scoring; scorer training is required.

Appliedin projects: Assessing Subject-specific Competencies in Teacher Education in Mathematics and Business and Economics – a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi); https://www.eng.elmawi.de/

Contact / Location:
Dr. Colin Jeschke
IPN – Leibniz Institute for Science and Mathematics Education Kiel
Email: jeschke@leibniz-ipn.de

GENERAL INFORMATION

Name: Reflective Competence of Mathematics Teachers (Math-RC)
Subject domain: Teacher training mathematics
Applied in projects: Assessing Subject-specific Competencies in Teacher Education in Mathematics and Business and Economics - a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi); https://www.eng.elmawi.de/
Contact / Location:
Dr. Colin Jeschke
IPN – Leibniz Institute for Science and Mathematics Education Kiel
Email: jeschke@leibniz-ipn.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Reflective competence of mathematics teachers
Theoretical model: According to the competence structure model with two facets AC and RC (see the Math-AC test) based on Lindmeier (2011) and Kuhn (2014)
Test type: Performance test; video-based tasks with open-response format
Modality: Computer- and video-based; audio recordings and written responses to assess reflexive competence (RK)
Test structure:
Item pool: 9 items; a scale for planning and reflecting on mathematics teaching in the subject areas of secondary algebra (6 items) and analysis (3 items). The test deals with problems concerning teaching materials (4 items), planning of first steps (2 items) and reflection on student work (3 items). Each item contains a short description of the situation and an illustration or a short video clip (30-60 seconds), to which the participants can respond (open-response format). The results are scored using a proven manual according to qualitative content and formal criteria (from 0 to 2 points).

TEST QUALITY CRITERIA

Reliability: Cronbach's alpha for RC =.68 (9 tasks)
Validity:
Test content: —
Response processes: —
Internal test structure: The one-dimensional structure of AC is confirmed by CFA (Jeschke et al. 2019).
Relationships with other competence indicators: As assumed, weak yet significant relationships to other competence measures, e.g. intelligence, motivation, beliefs, grades (Jeschke et al. 2019).
Consequences of testing: Theoretical differences in test performance between different status groups (bachelor, master, practical training phase, teaching service) (Jeschke et al. 2019).
Test fairness: Due to the language intensity of the test, a language bias has to be assumed.

PRACTICAL USE

GENERAL INFORMATION

Duration: 30 minutes
Testing materials: Computer or laptop with test program
Special features: Supervision by a trained test administrator is recommended. The answers should be scored by (preferably two) trained persons using the coding manual provided.
Practical example: The test can be used at any location that allows the use of suitable computers or laptops, headsets with microphone and specially developed software. It includes an introduction to the use of the software and a detailed description of each item. For each item, a picture (e.g. of teaching materials) or a video (e.g. of a problematic lesson goal) is shown, to which the participants respond. The data evaluation is based on a coding manual.

DOMAIN

Tested for: Mathematical didactics
Suitable for: Other teaching domains (e.g. business and economics education)
Not suitable for: —

TARGET GROUP

Tested for: Pre-service and in-service teachers of mathematics at secondary level II (student teachers in mathematics [bachelor, master], trainee teachers and in-service teachers of mathematics)
Suitable for: Pre-service and in-service teachers of mathematics at lower secondary level I, possibly as a short scale (corresponding tests were also developed by the authors for the primary and elementary level)
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessment of the level of competence in the two facets AC and RC (see the Math-AC test) and comparisons of the level of competence between the training phases of teacher education (bachelor, master, practical training phase, school service); comparisons between pre-service and in-service teachers in the subjects of mathematics and/or business and economics
Suitable for: Assessing competence development in all training phases; comparisons with other school subjects
Not suitable for: Entry diagnostics

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Video-based instruments for assessing AC and RC in business and economics for domain comparisons; tests for assessing PCK and CK in business and economics and mathematics; tests for assessing generic competencies (e.g. intelligence, situational awareness, ambiguity tolerance) for analyses of domain specificity (Jeschke et al. 2019)
Suitable for: Further tests for assessing subject-related or generic competencies
Not suitable for: Further tests demanding much attention, time or effort

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at the individual level, group or course level, university level, international level (German language area).
Suitable for: State and national level
Not suitable for: International level (non-German-speaking countries)

APPLICATIONS

Tested for: Comparisons of the level of competence between training phases in the teacher education (bachelor, master, practical training phase, teaching service); comparisons between business and economics and mathematics (for student teachers, trainees, in-service teachers)
Suitable for: Teaching-learning tool for estimating individual learning needs; assessment of competence development in all phases of education and further training; comparisons with other school subjects
Not suitable for: Entry diagnostics

PERMITTED CONSEQUENCES

Tested for: Assessment of the level of competence in teacher training; information on the further development of course offers on the tested competencies
Suitable for: Assessment of competence development over the various phases; statistical prediction of the level of competence at a later point in time
Not suitable for: Entry diagnostics

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test can be expanded for further content dimensions (e.g. in the form of additional tasks). The aim is to further develop the survey software to keep it up to date in the long term. The test for the assessment of Math-RC can be requested from the authors at any time. Test development is documented in detail in Lindmeier (2011).
In the future, the test will also be used in the project "Promotion of Subject-specific Competencies of Pre-service Teachers in Mathematics and Business & Economics Using Video-based ELMaWi Tools (ELMaWi-Transfer)".

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Jeschke, C.; Lindmeier, A. & Heinze, A. (2018). Wie fachspezifisch sind „fachspezifische Kompetenzen" von Lehrkräften? In: Kortenkamp, U. & Kuzle, A. (Eds.): *Beiträge zum Mathematikunterricht 2017 WTM*, pp. 1119–1122.
Jeschke, C., Lindmeier, A. & Heinze, A. (in review). Vom Wissen zum Handeln unter Druck: Wie wird fachspezifisches Wissen von Mathematiklehrkräften für Anforderungssituationen verfügbar? *Journal für Mathematik-Didaktik.*
Jeschke, C.; Kuhn, C.; Lindmeier, A.; Zlatkin-Troitschanskaia, O.; Saas, H. & Heinze, A. (2019). Performance assessment to investigate the domain-specificity of instructional skills among pre-service and in-service teachers of mathematics and economics. *British Journal of Educational Psychology*, 1-13. https://doi.org/10.1111/bjep.12277

SECONDARY LITERATURE

Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung.* Empirische Berufsbildungs- und Hochschulforschung, vol. 2. Landau: Verlag Empirische Pädagogik.
Lindmeier, A. (2011). *Modelling and measuring knowledge and competencies of teachers: A threefold domain-specific structure model for mathematics.* Münster: Waxmann.

Research Competence in Educational Sciences

SUMMARY

Name: Assessing research competence in educational science degree programs

Domain: Educational science, in particular teacher training, pedagogy, early education

Assessed competencies: Components of research literacy: information literacy, statistical literacy, critical thinking

Target group: Students of educational science

Test type: Multiple-choice test

Modality: Paper-pencil test

Duration: 30 minutes per component; 30, 60 or 90 minutes depending on test composition

Test structure: Item pools were developed for each of the three components, from which flexible, application-specific test versions can be generated.

General test purpose: Particularly suitable for repeated use in evaluation studies as parallel test versions can be flexibly created.

Application scenarios: Prediction of study success, embedded assessment, university evaluation, course comparisons, monitoring of learning success

Not suitable for: Testing at the individual level (since no adaptive test was developed)

Note for practical use: The reliability of the scales depends on the item composition; the suitability depends on the fit for the intended test purpose. Reliable statements for different practical applications are only possible to a limited extent. The answers are scored and interpreted using IRT analyses. Example values for different degree programs are available for evaluation.

Applied in projects: Learning the Science of Education (LeScEd)

Contact / Location:
Dr. Jana Groß-Ophoff
University of Education Freiburg
Email: jana.grossophoff@ph-freiburg.de

GENERAL INFORMATION

Name: Assessing research competence in educational science degree programs
Applied in projects: Learning the Science of Education (LeScEd)
Contact / Location:
Dr. Jana Groß-Ophoff
University of Education Freiburg
Email: jana.grossophoff@ph-freiburg.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Components of research literacy: information literacy, statistical literacy, critical thinking
Theoretical model: Research literacy according to the models of Davidson (2013), Davies (1999), Lambert (2009) and Pedaste et al. (2015)
Test type: Multiple-choice test
Modality: Paper-pencil-test
Test structure:
Item pool: Item pools were developed for each of the three components.
Information literacy (23 items): Ability to identify information needs, formulate adequate research questions and conduct targeted and reflective research.
Statistical competence (68 items): ability to systematically analyze and evaluate statistical information from tables, graphs or presentations of results.
Research-related critical thinking (91 items): ability to interpret evidence or to derive and apply appropriate conclusions or to critically reflect on interpretations of results, as typically found in the discussion of research articles.
Test versions: flexible, application-specific test versions can be generated.

TEST QUALITY CRITERIA

Reliability: The reliability of the scales depends on the item composition. When the test is used, the reliability is insufficient if the test results are very good or very bad.
Validity:
Test content: Tested in expert surveys
Response processes: Tested in cognitive pre-tests
Validity – internal test structure: Three dimensions of the construct confirmed in IRT analysis
Relationships with other competence indicators: Relations to motivation were examined: For instance, learning orientation as a positive predictor for self-assessed research competence, and work avoidance as a negative predictor for performance in the competence test
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 30 minutes per component; 30, 60, or 90 minutes depending on the test composition
Testing materials: Paper printouts, pens
Special features: The results are scored and analyzed using IRT analyses. Comparative values for different degree programs are available for evaluation
Practical example: In educational science degree programs, courses and modules that are used to acquire research-methodological competencies can be evaluated by means of pre-post surveys. In addition, entry requirements for research method degree programs can be determined by comparison (e.g. different degree programs or cohorts) or used to plan curricular priorities.

DOMAIN

Tested for: Educational science degree programs, in particular teacher education, educational sciences, early childhood education, health education
Suitable for: —
Not suitable for: —

TARGET GROUP

Tested for: Students of educational science
Suitable for: —
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Particularly suitable for repeated use in evaluation studies, as parallel test versions can be created flexibly
Suitable for: Predicting study success, embedded assessment, university evaluation, course comparisons, monitoring of learning success
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: In particular, use alongside motivation and interest scales is recommended
Suitable for: Other measures of competence assessment
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at course level, university level.
Suitable for: —
Not suitable for: Individual diagnostics, international level

APPLICATION SCENARIOS

Tested for: Predicting study success, embedded assessment, university evaluation, course comparisons, monitoring of learning success
Suitable for: —
Not suitable for: Individual diagnostics, international level

PERMITTED CONSEQUENCES

Tested for: Research competencies can be specifically assessed for planning teaching curricula as well as for formative and summative purposes in evaluations
Suitable for: State and national level
Not suitable for: Individual diagnostics, international level

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The reliability of the test for very good and very bad test results is currently insufficient. The items and item pools are validated. Depending on their composition, however, the reliability and fit of the compiled scales must be tested for the respective test purpose. Please consult the test developers for further details. The test is available on request.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Groß Ophoff, J.; Schladitz, S.; Leuders, J.; Leuders, T. & Wirtz, M. (2015). Assessing the development of Educational Research Literacy. The effect of courses on research methods in studies of Educational Science. *Peabody Journal of Education* 90(4), pp. 560–573.
Schladitz, S.; Groß Ophoff, J. & Wirtz, M. (2015). Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz. In: Blömeke, S. & Zlatkin-Troitschanskaia, O. (eds.), Kompetenzen von Studierenden. *Zeitschrift für Pädagogik*, Supplement 61, pp. 167–184. Weinheim: Beltz. 65
Winter-Hölzl, A.; Wäschle, K.; Wittwer, J.; Watermann, R. & Nückles, M. (2015). Entwicklung und Validierung eines Tests zur Erfassung des Genrewissens Studierender und Promovierender der Bildungswissenschaften. In: Blömeke, S. & Zlatkin-Troitschanskaia, O. (Eds.), Kompetenzen von Studierenden. *Zeitschrift für Pädagogik*, Supplement 61, pp. 185–202. Weinheim: Beltz.

SECONDARY LITERATURE

Davidson, K. (2013). Teachers' reported utilization of reading disabilities research.
Alberta Journal of Educational Research, 59(3), pp. 487–502.
Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies, 47*(2), pp. 108–121.
Lambert, C. (2009). Pedagogies of participation in higher education: A case for research-based learning. *Pedagogy, Culture & Society, 17*(3), pp. 295–309.
Pedaste, M.; Mäeots, M.; Siiman, L. A.; De Jong, T.; Van Riesen, S. A.; Kamp, E. T. & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, pp. 47–61.

Research Competence in Psychology

SUMMARY

Name: Test of Research Competence in Psychology

Domain: Social science subjects; all subjects working with empirical methods

Assessed competencies: Competencies in research methodology: elementary statistics, experiment and variance analysis, correlation and regression, factor analysis, psychometrics

Target group: Bachelor and master students

Test type: Scenario-based test with the answer formats multiple-choice, true-false answers, open-response items

Modality: Paper-pencil test

Duration: 120 minutes

Test setup: Four scenarios for assessing competence in the areas of elementary statistics, experiment and variance analysis, correlation and regression, factor analysis, psychometrics

General test purpose: Assessing students' competence level in psychological research methodologies

Application scenarios: Teaching-learning tool, course comparison (if the curriculum corresponds to the test contents)

Not suitable for: —

Note for practical use: The results are scored according to criteria and can be reported back individually, e.g. in the form of grades.

Applied in projects: Modelling and Measurement of Scientific Competence in Social Science Subjects (WiKom-SoWi)

Contact / Location:
Dr. Eric Klopp and Prof. Dr. Robin Stark
Saarland University
Email: e.klopp@mx.uni-saarland.de
r.stark@mx.uni-saarland.de

Dr. Heike Dietrich and Prof. Dr. Birgit Spinath
Heidelberg University
Email: Heike.Dietrich@psychologie.uni-heidelberg.de
Birgit.Spinath@psychologie.uni-heidelberg.de

GENERAL INFORMATION

Name: Test of Research Competence in Psychology
Applied in projects: Modelling and Measurement of Scientific Competencies in Social Science Subjects (WiKom-Sowi)
Contact / Location:
Dr. Eric Klopp and Prof. Dr. Robin Stark
Saarland University
Email: e.klopp@mx.uni-saarland.de
r.stark@mx.uni-saarland.de

Dr. Heike Dietrich and Prof. Dr. Birgit Spinath
Heidelberg University
Email: Heike.Dietrich@psychologie.uni-heidelberg.de
Birgit.Spinath@psychologie.uni-heidelberg.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Competencies in research methodology: elementary statistics, experiment and variance analysis, correlation and regression, factor analysis, psychometry
Theoretical model: Model for social science competencies according to Dietrich et al. (2015)
Modality: Paper-pencil test
Test setup:
Item pool: 120 items for declarative knowledge and for application-oriented competence assessment in four scenarios to ensure the most authentic setting possible. The scenarios present findings as they are typically found in research practice. Elementary statistics: 20 items for dealing with basic statistical concepts (mean value, dispersion, distribution, etc.); Experiment and variance analysis: 21 items for the evaluation of psychological experiments and the application of variance-analytical methods; Correlation and regression: 22 items for applying correlation and regression analyses; Factor analysis: 25 items for the application and interpretation of factor analyses; Psychometrics: 31 items for the application of psychometric procedures

TEST QUALITY CRITERIA

Reliability: Criterion-related reliability (Subkoviak, 1976) for elementary statistics (.81), experiment and variance analysis (.82), correlation and regression (.78), factor analysis (.87) and psychometrics (.84)
Validity: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 120 minutes
Testing materials: Test sheets and pens
Special features: The test is designed to be criterion-oriented and allows for individual feedback on competence levels, e.g. in the form of grades.
Practical example: It is possible to conduct the test in group settings. Use, e.g. as an evaluation instrument in method training.

DOMAIN

Tested for: Psychology
Suitable for: Other subjects working with empirical research
Not suitable for: —

TARGET GROUP

Tested for: Bachelor and master students
Suitable for: —
Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessment of competence level in psychological research methods
Suitable for: Assessing competence development in psychological research methods
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Intelligence, performance motivation, interest
Suitable for: Other measures of competence measurement
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for individuals over time, groups in different study sections, groups with different instruction methods.
Suitable for: —
Not suitable for: University level, state and national level, international level

APPLICATION SCENARIOS

Tested for: Teaching-learning tool, course comparison (if the curriculum corresponds to the test contents)
Suitable for: —
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Individual and group feedback
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test has so far only been used with a small sample; the test quality can only be determined after further application. The test is available on request from the test developers.

Do you have any further questions or need help, e.g. with practical use or evaluation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Dietrich, D.; Zhang, Y.; Klopp, E.; Brünken, R.; Krause, U.-M.; Spinath, F. M.; Stark, R. & Spinath, B. (2015). Scientific competencies in the social sciences. *Psychology Learning and Teaching*, 14(2), pp. 115–130.

SECONDARY LITERATURE

Subkoviak, M. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement*, 13, pp. 256–276.

Scientific Competence in Political Sciences

SUMMARY

Name: Test of Scientific Competence in Political Sciences

Domain: Political sciences

Assessed competencies: Scientific competence in political sciences

Target group: Students of political sciences

Test type: Performance test; scenario-based test with different response formats such as multiple-choice items, true-false items, allocation items, open-response items

Modality: Paper-pencil test

Duration: 90 minutes

Test setup: Four scenarios for the subject areas research methods (4 tasks with 3 multiple-choice items, 3 true-false items, 28 allocation items and 8 open-response items), international relations (3 tasks with 8 multiple-choice items, 19 allocation items and 2 open-response items), political theory (5 tasks with 6 multiple-choice items, 9 true-false items, 28 allocation items and 12 open-response items), and comparative government theory (4 tasks with 6 multiple-choice items, 38 allocation items and 3 open-response items)

General test purpose: Assessing political science students' competence level and development, in particular their ability to apply different cognitive operations to different content areas.

Application scenarios: Assessment of the competence level of students of political science; can also be used for the formative and summative evaluation of selected learning processes.

Not suitable for: —

Note for practical use: Currently, the only available data is from a small pilot project with 15 students. Example items and tests are available on request from the test developers.

Applied in projects: Modelling and Measurement of Scientific Competence in Social Science Subjects (WiKom-SoWi)

Contact / Location:
Prof. Dr. Birgit Spinath and Dr. Heike Dietrich
Heidelberg University
Website: https://www.psychologie.uni-heidelberg.de/ae/paeps/index.html

Prof. Dr. Robin Stark and Dr. Eric Klopp
Saarland University
Website: https://www.uni-saarland.de/lehrstuhl/stark.html

GENERAL INFORMATION

Name: Test of Scientific Competence in Political Sciences
Applied in projects: Modelling and Measurement of Scientific Competence in Social Science (WiKom-SoWi)
Contact / Location:
Prof. Dr. Birgit Spinath and Dr. Heike Dietrich
Heidelberg University
Website: https://www.psychologie.uni-heidelberg.de/ae/paeps/index.html

Prof. Dr. Robin Stark and Dr. Eric Klopp, Saarland University
Website: https://www.uni-saarland.de/lehrstuhl/stark.html

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Scientific competence in political sciences
Theoretical model: Model of scientific competence in social science subjects
Test type: Performance test; scenario-based test with different response formats such as multiple-choice items, true-false items, allocation items, open-response items
Modality: Paper-pencil test
Test setup: 173 items in 4 scenarios for the content areas research methods, international relations, political theory and comparative government theory

TEST QUALITY CRITERIA

No statements on the psychometric properties of the test are possible, since so far only one pilot with N = 15 students of political science has been conducted. Example items and tests are available on request and after coordination of the intended use.

PRACTICAL USE

GENERAL INFORMATION

Duration: 90 minutes
Testing materials: Test sheets and pens
Special features: It is possible to use the individual scenarios for the respective topic areas of interest.
Practical example: A group setting is possible

DOMAIN

Tested for: Political science
Suitable for: Other social science subjects
Not suitable for: Other study subjects

TARGET GROUP

Tested for: Political sciences
Suitable for: Other social science subjects
Not suitable for: Other study subjects

GENERAL TEST PURPOSE

Tested for: Assessing political science students' competence level and development, in particular their ability to apply different cognitive operations to different content areas.
Suitable for: —
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: —
Suitable for: Intelligence measures, study grades, other measures of competence
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: —
Suitable for: Individual and group level, course level
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Assessing political science students' competence level
Suitable for: Assessing political science students' competence development
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Research purposes only
Suitable for: Individual and group feedback
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test has so far only been used with a small sample; the test quality can only be determined after further use. The test is available on request from the test developers.

Do you have any further questions or need help, e.g. with practical use or evaluation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Dietrich, H.; Zhang, Y.; Klopp, E.; Brünken, R.; Krause, U.-M.; Spinath, F. M.; Stark, R. & Spinath, B. (2015). Scientific Competencies in the Social Sciences. *Psychology Learning and Teaching*, 14, pp. 115–130.

Scientific Competence in Sociology

SUMMARY

Name: Test of Scientific Competence in Sociology

Domain: Sociology

Assessed competencies: Scientific competence in sociology

Target group: Sociology students

Test type: Scenario-based test with different response formats

Modality: Paper-pencil-test

Duration: 90 minutes

Test setup: Five scenarios for the subject areas: Sociological theory (4 tasks each with 1 multiple-choice item, 10 true-false items, 27 allocation items and 1 open-response item); research methods (3 tasks each with 2 multiple-choice items, 9 true-false items, 21 allocation items and 6 open-response items), media use (4 tasks with 2 multiple-choice items, 20 allocation items and 6 open-response items), change of the working environment (4 tasks with 2 multiple-choice items, 6 true-false items, 27 allocation items and 1 open-response item), and educational opportunities (4 tasks with 1 multiple-choice item, 3 true-false items, 43 allocation items and 3 open-response items)

General test purpose: Assessing sociology students' competence level and development, in particular their ability to apply different cognitive operations to different content areas

Application scenarios: During and after sociology studies; use of the individual scenarios for the respective topic of interest is possible.

Not suitable for: —

Note for practical use: No assertions on the psychometric properties of the test are possible, as so far there has only been a pilot test with N = 11 sociology students.

Applied in projects: Modelling and Measurement of Scientific Competence in Social Science Subjects (WiKom-SoWi); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/wikom-sowi/

Contact / Location:
Prof. Dr. Birgit Spinath and Dr. Heike Dietrich
University Heidelberg
Email: Birgit.Spinath@psychologie.uni-heidelberg.de
Heike.Dietrich@psychologie.uni-heidelberg.de
Website: https://www.psychologie.uni-heidelberg.de/ae/paeps/index.html

Prof. Dr. Robin Stark and Dr. Eric Klopp
University des Saarlandes
Email: r.stark@mx.uni-saarland.de
e.klopp@mx.uni-saarland.de
Website: https://www.uni-saarland.de/lehrstuhl/stark.html

GENERAL INFORMATION

Name: Test of Scientific Competence in Sociology
Applied in projects: Modelling and Measurement of Scientific Competence in Social Science Subjects (WiKom-SoWi); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/wikom-sowi/

Contact / Location:
Prof. Dr. Birgit Spinath and Dr. Heike Dietrich
University Heidelberg
Email: Birgit.Spinath@psychologie.uni-heidelberg.de
Heike.Dietrich@psychologie.uni-heidelberg.de
Website: https://www.psychologie.uni-heidelberg.de/ae/paeps/index.html

Prof. Dr. Robin Stark and Dr. Eric Klopp
University des Saarlandes
Email: r.stark@mx.uni-saarland.de
e.klopp@mx.uni-saarland.de
Website: https://www.uni-saarland.de/lehrstuhl/stark.html

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Test of scientific competence in sociology
Theoretical model: Model of Scientific Competence in Social Science (Dietrich et al. 2015a; b)
Test type: Scenario-based test with different task formats
Modality: Paper-pencil test
Test setup: One scenario per subject area with introductory text and subsequent tasks. Five scenarios for the subject areas sociology theory (4 tasks each with 1 multiple-choice item, 10 true-false items, 27 allocation items and 1 open-response item), research methods (3 tasks each with 2 multiple-choice items, 9 true-false items, 21 allocation items and 6 open-response items), media use (4 tasks with 2 multiple-choice items, 20 allocation items and 6 open-response items), change of the working environment (4 tasks with 2 multiple-choice items, 6 true-false items, 27 allocation items and 1 open-response item), and educational opportunities (4 tasks with 1 multiple-choice item, 3 true-false items, 43 allocation items and 3 open-response items)

TEST QUALITY CRITERIA

No statements on the psychometric characteristics of the test are possible, since so far there has only been one pilot with N = 11 students of sociology.

PRACTICAL USE

GENERAL INFORMATION

Duration: 90 minutes
Testing materials: Test sheets and pens
Special features: It is possible to use the individual scenarios for the respective topic of interest.
Practical example: Assessment of sociology students' competence level and development during and after their studies in sociology; group settings are possible. Currently, only data from a small pilot project with 11 students is available.

DOMAIN

Tested for: Sociology
Suitable for: Other social science subjects
Not suitable for: Other subjects

TARGET GROUP

Tested for: Sociology students
Suitable for: Students of other social science subjects
Not suitable for: Students of other subjects

GENERAL TEST PURPOSE

Tested for: Assessing sociology students' competence level and development, in particular their ability to apply different cognitive operations to different content areas
Suitable for: —
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: —
Suitable for: Intelligence measures, study grades, other measures of competence assessment
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: —
Suitable for: Individual and group level, course level
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Assessing sociology students' competence level
Suitable for: Assessing sociology students' competence development
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Research purposes only
Suitable for: Individual and group feedback
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test has so far only been used on a small sample; the test quality can only be determined after further use. The test is available on request from the test developers.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Dietrich, H.; Zhang, Y.; Klopp, E.; Brünken, R.; Krause, U.-M.; Spinath, F. M.; Stark, R. & Spinath, B. (2015a). Es gibt nichts Praktischeres als solide wissenschaftliche Kompetenzen – Erwerb und Gebrauch wissenschaftlicher Kompetenzen in der Soziologie. *Zeitschrift für Sozialmanagement*, 13, pp. 49–62.
Dietrich, H.; Zhang, Y.; Klopp, E.; Brünken, R.; Krause, U.-M.; Spinath, F. M.; Stark, R. & Spinath, B. (2015b). Scientific Competencies in the Social Sciences. *Psychology Learning and Teaching*, 14, pp. 115–130.

Scientific Thinking in Natural Sciences: Ko-WaDiS Test

SUMMARY

Name: Ko-WADiS test

Domain: Natural sciences; teachers of biology, chemistry, physics and other natural science subjects, e.g. geology, meteorology; other empirical-scientific subjects, e.g. psychology

Assessed competencies: Scientific thinking, scientific reasoning

Target group: Bachelor's and master's students, pre-service teachers in their practical phase, in-service teachers, high school graduates, doctoral students in science didactics

Test type: Performance test; multiple-choice test in single-best-answer format

Modality: Paper-pencil test

Duration: 35-45 minutes (no time limit)

Test structure: 63 items; 7 sub-competencies: formulating scientific questions, generating hypotheses, planning investigations, evaluating and interpreting data, determining the purpose of scientific models, testing models, changing models. Contexts from 3 subject domains: biology, chemistry, physics; Multi-matrix design (Balanced-Incomplete-Block-design) with 9 parallel test booklet versions with 21 items per test booklet (7 items per subject, 3 items per partial competence)

General test purpose: Assessing competence level and competence development; assessing the effectiveness of learning opportunities

Application scenarios: Assessing the level and development of competence, e.g. annual comparison of several student cohorts; course comparisons and evaluation of courses

Not suitable for: Individual diagnostics; disciplines that do not work empirically

Note for practical use: The method is used in multi-matrix design (Balanced-Incomplete-Block-design), which results in missing values due to the design. It should be scored using probabilistic methods.

Applied in projects: Evaluating the development of scientific literacy in teacher education (Ko-WADiS); https://didaktik.physik.fu-berlin.de/projekte/kowadis/index.html; Competence modelling and assessment: Validation of the test score interpretations of a scientific reasoning test (ValiDiS); https://www2.hu-berlin.de/biologie/biodidaktik/forschung/Validis.html

Contact / Location:
Freie Universität Berlin
Institute of Biology
Humboldt University Berlin
Website: https://www2.hu-berlin.de/biologie/biodidaktik/

GENERAL INFORMATION

Name: Ko-WADiS test
Applied in projects: Evaluating the development of scientific literacy in teacher education (Ko-WADiS); https://didaktik.physik.fu-berlin.de/projekte/kowadis/index.html; Competence modelling and assessment: Validation of the test score interpretations of a scientific reasoning test (ValiDiS); https://www2.hu-berlin.de/biologie/biodidaktik/forschung/Validis.html

Contact / Location:
Freie Universität Berlin
Institute of Biology
Humboldt University Berlin
Website: https://www2.hu-berlin.de/biologie/biodidaktik/

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Scientific thinking, scientific reasoning
Theoretical model: Scientific thinking according to Mayer (2007) and scientific modelling according to Upmeier zu Belzen & Krüger (2010)
Test type: Performance test; multiple-choice test in single-best-answer format
Modality: Paper-pencil test
Test structure:
Item pool: 63 items; 7 partial competencies: formulating scientific questions, generating hypotheses, planning investigations, evaluating and interpreting data, determining the purpose of scientific models, testing models, changing models. Contexts from three subject domains: biology, chemistry, physics
Test versions: Multi-matrix design (Balanced-Incomplete-Block-design) with 9 parallel test booklet versions with 21 items per test booklet (7 items per subject, 3 items per partial competence)

TEST QUALITY CRITERIA

Reliability: EAP/PV-Reliabilities: .56-.81; Cronbach's α per test booklet: .44-.81

Validity:

Test content: The validity of the tests in terms of content and curriculum was confirmed by construction according to standardized instructions and theoretical models, expert ratings for all items and expert ratings for selected items.

Response processes: The multiple-choice item response options were derived from open student responses and results of the think-aloud study with students on selected items. Eye tracking study using the cued retrospective method for an online version.

Internal test structure: The analyses of dimensionality speak in favor of a one-dimensional competence structure; however, fitness measures of multidimensional models are not significantly worse than those of the one-dimensional model.

Relationships with other competence indicators: Validity analyses with known groups were carried out for suitable samples in accordance with theory; positive correlations of the construct with other competence indicators are given for partial samples: basic cognitive ability, reasoning, complex problem solving, teachers' professional knowledge, grade of higher education entrance qualification.

Consequences of testing: Increases in competence in regular studies and in targeted support measures (interventions) are covered by the test.

Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 35-45 minutes (no time limit)

Testing materials: Pencils, questionnaires

Special features: The method is used in a multi-matrix design (Balanced-Incomplete-Block-design), which results in missing values. It should be scored using probabilistic methods.

Practical example: Multiple-choice items were administered for the first time at two universities in 2012. Tests were also administered at four other universities; smaller comparative samples (N < 100) were collected at further universities. The number of items originally developed was reduced from 141 to 123 following psychometric and qualitative analyses. In the further course of the project, the 63 items with the best psychometric properties were selected for an optimized instrument. The items with biological contexts were also translated into English, Spanish and Greek and used in cooperation in Australia, Chile, Canada, Spain and Cyprus.

DOMAIN

Tested for: Teaching biology, chemistry, physics

Suitable for: Natural sciences, teaching of other natural science subjects, e.g. geology, other empirical-scientific subjects, e.g. psychology

Not suitable for: Non-empirical domains

TARGET GROUP

Tested for: Bachelor's and master's students from beginning to end of their studies

Suitable for: High school graduates, pre-service teachers in their practical training phase, in-service teachers, doctoral students of science didactics

Not suitable for: —

GENERAL TEST PURPOSE

Tested for: Assessing competence level and competence development

Suitable for: Assessing the effectiveness of learning opportunities

Not suitable for: Individual diagnostics

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Domain-specific knowledge tests, complex problem solving (Genetics Lab), video study for experimentation, intelligence structure test I-S-T screening subscales (verbal, numeric, figural)

Suitable for: Further measures of competence

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for groups of more than 100 people, university level.

Suitable for: Can also be used at course level under the prerequisites of large groups, state and national level, international level.

Not suitable for: Individuals (a computer-based adaptive test with Ko-WADiS items will be developed for this purpose.)

APPLICATION SCENARIOS

Tested for: Assessing the competence level and competence development of student cohorts

Suitable for: Under prerequisites of large groups, course comparisons and evaluation of courses

Not suitable for: Individual diagnostics

PERMITTED CONSEQUENCES

Tested for: —

Suitable for: —

Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

A computer-based, adaptive version of the Ko-WADiS test is currently being developed in the Physics Didactics working group at Freie Universität Berlin. The aim is to make competence measurement even more economical, reliable and flexible while maintaining validity, also for individual diagnostics.
International comparative studies will be initiated using the available translations and the comparative validation studies.

ADDITIONAL INFORMATION

The test was developed with the aim to assess the competence development of entire student cohorts reliably, validly and economically. Individual diagnostics are not intended and are subject to a significantly increased measurement error. The reliability of the test and the evidence of validity support the interpretation of the test scores as a measure of competence in the field of scientific reasoning. The longitudinal section results show an almost evenly linear increase in competence over the course of study. The goal of an economic, sufficiently reliable and valid competence assessment in large scale use was achieved, and the test can be used to measure the competence development of scientific ways of thinking over an entire course of study.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Hartmann, S.; Upmeier zu Belzen, A.; Krüger, D. & Pant, H. A. (2015). Scientific reasoning in higher education: Constructing and evaluating the criterion-related validity of an assessment of preservice science teachers' competencies. *Zeitschrift für Psychologie, 223*, 47–53. doi: 10.1027/2151-2604/a000199

Stiller, J.; Hartmann, S.; Mathesius, S.; Straube, P.; Tiemann, R.; Nordmeier, V. et al. (2016). Assessing scientific reasoning: A comprehensive evaluation of item features that affect item difficulty. *Assessment and Evaluation in Higher Education, 41*, 721–732. doi: 10.1080/02602938.2016.1164830

Krell, M.; Redman, C.; Mathesius, S.; Krüger, D. & van Driel, J. (2018). Assessing Pre-Service Science Teachers' Scientific Reasoning Competencies. *Research in Science Education.* doi: 10.1007/s11165-018-9780-1

Mathesius, S.; Krell, M.; Upmeier zu Belzen, A. & Krüger, D. (2019). Überprüfung eines Tests zum wissenschaftlichen Denken unter Berücksichtigung des Validitätskriteriums relations-to-other-variables. *Zeitschrift für Pädagogik, 65*(4), 492–510. doi: 10.3262/ZP1904492

SECONDARY LITERATURE

Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In: D. Krüger & H. Vogt (eds.), *Theorien in der biologiedidaktischen Forschung*, 177–186. Berlin: Springer. doi: 10.1007/978-3-540-68166-3_16

Upmeier zu Belzen, A. & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften, 16*, pp. 41–57.

Self-Regulated Learning – Questionnaire for Action and Knowledge

SUMMARY

Name: Self-Regulated Learning – Questionnaire for Action and Knowledge (SRL-QuAK); SRL-QuAKs (short version)

Domain: All courses

Assessed competencies: Self-regulated learning (knowledge and application)

Target group: University students

Test type: Situation-specific questionnaire (7-point Likert scale)

Modality: Online test

Duration: 120 minutes (long version), 15 minutes (short version)

Test structure: Partial scales: metacognition, cognition, frustration, boredom. 366 items (long version), 96 items (short version)

General test purpose: Assessing discrepancies between SRL knowledge and application as well as the reasons for not applying favorable SRL strategies; Assessing competence level and development; repeat test after 1 semester; longitudinal survey to assess competence development

Application scenarios: Evaluation of interventions and training to promote SRL (parts); prediction of course and examination performance; prediction of study success, prediction of performance development, university comparisons

Not suitable for: Use outside of the university context; international level

Note for practical use: SRL-QuAKs can be used in various online questionnaire programs (unipark, soscisurvey, limesurvey).

Applied in projects: Product- and process-oriented modeling and assessment of self-regulated learning competencies in tertiary education (PRO-SLR); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/pro-srl/; Product- and Process Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation (PRO-SLR-EVA); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/pro-srl-eva/

Contact / Location:
Univ.-Prof. Dipl.-Psych. Dr. Barbara Schober
University of Vienna
Email: barbara.schober@univie.ac.at
Website: https://bildung-psy.univie.ac.at/forschung/aktuelle-forschungsthemen/selbstreguliertes-lernen/
Location: Research Data Centre (FDZ) at the Institute for Educational Quality Improvement (IQB)

GENERAL INFORMATION

Name: Self-Regulated Learning – Questionnaire for Action and Knowledge (SRL-QuAK); SRL-QuAKs (short version)
Applied in projects: Product- and process-oriented modeling and assessment of self-regulated learning competencies in tertiary education (PRO-SLR); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/pro-srl/; Product- and Process Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation (PRO-SLR-EVA); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/pro-srl-eva/

Contact / Location:
Univ.-Prof. Dipl.-Psych. Dr. Barbara Schober
University of Vienna
Email: barbara.schober@univie.ac.at
Website: https://bildung-psy.univie.ac.at/forschung/aktuelle-forschungsthemen/selbstreguliertes-lernen/
Location: Research Data Centre (FDZ) at the Institute for Educational Quality Improvement (IQB)

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Self-regulated learning (SRL) (knowledge and application)
Theoretical model: SRL model by Dresel et al. (2015), based on Zimmerman's (2000) cyclical model of self-regulated learning
Test type: Situation-specific questionnaire (7-point Likert scale)
Modality: Online test
Test structure:
The questionnaire consists of scales on metacognition, cognition, frustration and boredom, all of which are further divided into subscales focusing on knowledge and application. At the beginning of each scale, a short vignette is given, asking students to put themselves in a learning situation (e.g. attending a lecture, writing a scientific paper or preparing for a major exam). This is followed by a short description of the formal challenges of the situation (e.g. preparing a large amount of material with a focus on comprehension). In the next step, the declarative strategy knowledge is assessed by asking about possible strategies with regard to their suitability for this situation (e.g. "I structure the material to be learned.") and to what extent they are applied. In order to measure procedural knowledge, concrete implementation options are now presented for the individual strategies (e.g. "I use mind maps, posters or the like to visualize the material"). The participants should again indicate whether they consider the mentioned strategies to be beneficial and whether they apply them.
Item pool: 366 items.
Partial scales: subscales that focus on knowledge or application.
Test versions: 1 long version, 1 short version (96 items)

TEST QUALITY CRITERIA

Reliability: Cronbachs α: .78-.94
Validity:
Test content: Relevant learning situations were selected that are highly significant across many study programs and that place great demands on the self-regulation of learning (e.g. "writing a scientific work", "preparing for a major exam"). This was ensured by conducting separate surveys of experts in the first project phase of PRO-SRL (Dresel et al., 2015; Steuer et al., 2015). Moreover, interviews and preliminary studies with teachers and students in the first project phases (Foerst, Klug, Schober & Spiel, 2017) underlined the high relevance of differentiating between knowledge about strategies and their application in practice.
Response processes: —
Internal test structure: The presented vignettes refer to SRL-relevant challenging situations with very specific requirements. These requirements as well as appropriate strategies for these situations have been identified by students and experts in preliminary studies. The validity results from the identification of relevant SRL challenges and "suitable" approaches and strategies – backed by the literature and experts – in the sense of a criterion and as well as from measuring knowledge and using the instrument to assess students. This is based on the assumption of a certain situation specificity, which in sum makes it possible to calculate a score that reflects the identification and application of suitable strategies as a whole.
Relationships with other competence indicators: In several studies, as expected, convergent correlations with other variables were found: for example, between cognition, application and performance (r = .24) and frustration, application and performance (r = .29). Within the subscales, high correlations with other variables were consistently found (e.g. ability self-concept, interest, intrinsic motivation, academic grades, organizing, critical review).
Consequences of testing: —
Test fairness: There were no gender differences in the partial scales.

PRACTICAL USE

GENERAL INFORMATION

Duration: 120 minutes (long version), 15 minutes (short version)
Testing materials: PC, Internet access
Special features: SRL-QuAKs can be used in various online questionnaire programs (unipark, soscisurvey, limesurvey)
Practical example: The first validation study of the long version SRL-QuAK was conducted in 2016 at the University of Vienna in the fields of Economics and Psychology. Students have a high level of knowledge of SRL strategies, but rarely apply them. Based on the results, two intervention studies were conducted in a Bachelor thesis seminar (WS 16/17) and a Master thesis seminar (SS 18). An efficient short version of the questionnaire was also prepared (SRL-QuAKs) and tested in SS 2018 in a validation study with *N* = 340 students from different fields of study at Austrian universities.

DOMAIN

Tested for: Economics, psychology
Suitable for: All study programs (domain-independent)
Not suitable for: Use outside of the university context

TARGET GROUP

Tested for: Students in all semesters; validation long version, *N* = 408 (2016, University of Vienna); validation short version, *N* = 340 (2018, universities in Austria)
Suitable for: Students at universities of applied sciences, students in German-speaking countries (D-A-CH)
Not suitable for: Use outside of the university context

GENERAL TEST PURPOSE

Tested for: Identification of discrepancies between SRL knowledge and application and the reasons for not applying favorable SRL strategies; identification of competence level and development; test repetition after one semester
Suitable for: Longitudinal survey to assess competence development
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Intelligence, grades, motivational orientation
Suitable for: —
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for individuals over time (beginning and end of semester), course level.
Suitable for: University level, state and national level, international level (German-speaking countries)
Not suitable for: International level (non-German-speaking countries)

APPLICATION SCENARIOS

Tested for: Evaluation of interventions and trainings to promote SRL (parts); prediction of course and examination performance
Suitable for: Prediction of study success, prediction of performance development, university comparisons
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Assessment of academic performance
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test could be used at universities as a component of student evaluations and in degree courses with curricula including the teaching of SRL competencies. The test is also suitable for evaluating the success of SRL training studies. The test can be requested from the authors.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Dresel, M.; Schmitz, B.; Schober B.; Spiel, C.; Ziegler, A.; Engelschalk, T.; Jöstl, G.; Klug, J.; Roth, A.; Wimmer, B. & Steuer, G. (2015). Competencies for successful self-regulated learning in higher education: Structural model and indications drawn from expert interviews. *Studies in Higher Education*, 40(3), pp. 454–470. DOI:10.1080/03075079.2015.1004236

Foerst, N. M.; Klug, J. Jöstl, G.; Spiel, C. & Schober, B. (2017). Knowledge vs. Action: Discrepancies in University Students' Knowledge about and Self-Reported Use of Self-Regulated Learning Strategies. *Frontiers in Psychology*, 8, p. 1288. DOI:10.3389/fpsyg.2017.01288

Foerst, N. M.; Pfaffel, A.; Klug, J.; Spiel, C. & Schober, B. (2019). SRL to Go? – Promoting SRL via Smartphone-App. *Unterrichtswissenschaft, 47*(3), S. 337-366. DOI:10.1007/s42010-019-00046-7

Situational General Pedagogical Knowledge and Situational Didactical Knowledge of Mathematics Teachers (P-PID & M-PID)

SUMMARY

Name: Mathematics-related perceptual, interpretation, and decision-making skills (M-PID); Pedagogical perceptual, interpretation, and decision-making skills (P-PID)

Domain: Mathematics teachers and mathematics teaching students

Assessed competencies: Situational Pedagogical Knowledge (P-PID), Situational Mathematical Didactical Knowledge (M-PID)

Target group: In-service mathematics teachers at the secondary school level, students of mathematics education in their master's program

Test type: Performance test; open-response answers, evaluation scale

Modality: Online test with videos

Duration: 60 minutes

Test structure: 42 items (P-PID), 36 items (M-PID)

General test purpose: Assessing competence level and competence development, evaluating measures for further training

Application scenarios: Group comparisons for evaluation

Not suitable for: Non-expert teachers and non-teachers, individual case studies, selection processes

Note for practical use: The test contains video vignettes that can only be viewed once. Closed items are scored automatically, the answers to open-response test items are coded (20% double). Personal feedback is differentiated according to the strengths and weaknesses of the participants.

Applied in projects: Teacher Education and Development Study: TEDS-Follow Up, TEDS-Teaching, TEDS-Validation, http://www.dr-gabriele-kaiser.de/teds-validierung.html; TEDS-East-West, http://www.dr-gabriele-kaiser.de/research-east-west.html

Contact / Location:
Prof. Dr. Gabriele Kaiser
University of Hamburg
Email: Gabriele.Kaiser@uni-hamburg.de
Websites:
http://www.dr-gabriele-kaiser.de/teds-validierung.html
http://www.dr-gabriele-kaiser.de/research-east-west.html

GENERAL INFORMATION

Name: Mathematics-related perceptual, interpretation, and decision-making skills (M-PID); Pedagogical perceptual, interpretation, and decision-making skills (P-PID)
Applied in projects: Teacher Education and Development Study: TEDS-Follow Up, TEDS-Teaching, TEDS-Validation, http://www.dr-gabriele-kaiser.de/teds-validierung.html; TEDS-East-West, http://www.dr-gabriele-kaiser.de/research-east-west.html

Contact / Location:
Prof. Dr. Gabriele Kaiser
University of Hamburg
Email: Gabriele.Kaiser@uni-hamburg.de
Websites:
http://www.dr-gabriele-kaiser.de/teds-validierung.html
http://www.dr-gabriele-kaiser.de/research-east-west.html

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Situational Pedagogical Knowledge (P-PID), Situational Mathematical Didactical Knowledge (M-PID)
Theoretical model: PID-Modell (perception of classroom situations, interpretation, decision-making; Blömeke et al. 2015)
Test type: Performance test; open-response format, scoring scale
Modality: Online test with videos
Test structure:
Item pool: 42 items (P-PID) for the categories class leadership, constructive support and cognitive activation; 36 items (M-PID) for mathematical and mathematical didactic knowledge

TEST QUALITY CRITERIA

Reliability: Total score for M-PID: WLE = .67; total score for P-PID WLE=.70 (Blömeke et al. 2016 based on TEDS-FU data); low testlet effects (Blömeke et al. 2015)

Validity:

Test content: Expert ratings on the appropriateness of video clips and open-response questions

Response processes: Double expert rating for evaluation of the Likert scales (Hoth et al. 2016)

Internal test structure: Empirical separability of scales perception and interpretation / decisions (König et al. 2014; Yang et al. 2018). The correlations between the individual competence facets were examined, i.e. M-PID in connection with MCK, MPCK and P-PID in connection with GPK.

Relationships with other competence indicators: —

Consequences of testing: M-PID is related to subject-specific teaching quality and cognitive activation (Jentsch et al., under review); relationship between teachers' competencies and learners' progress, mediated by teaching quality.

Test fairness: DIF analyses for participants with different degrees of professionalization point to cultural dependency in the comparison of East Asian and German teachers (Yang et al. 2018).

PRACTICAL USE

GENERAL INFORMATION

Duration: 60 minutes

Testing materials: Computer or laptop with Internet connection

Special features: The test contains video vignettes that can only be viewed once. Closed items are scored automatically. In addition, the answers to open-response test items should be coded manually (20% double). The personal feedback is differentiated according to the strengths and weaknesses of the participants.

Practical example: In the TEDS-Validation project (2016-2019), the tests M-PID and P-PID are used within the video-based scoring of teacher competencies. Participants watch each of the three videos once, each showing a brief overview of a mathematics lesson at secondary level with selected specific scenes. The participants are then asked open and closed questions. From a mathematics-didactical perspective, the teaching videos refer to the misconceptions of the students which are to be perceived, to the interpretation of these misconceptions and the development of corresponding courses of action, e.g. as a continuation of teaching or as feedback for the students who are involved. In addition, questions about the mathematical handling of mathematical problems play a role (enactive-iconic-symbolic) as well as questions about curricular integration of the shown teaching actions. The cognitive potential of tasks is also analyzed. From a pedagogical perspective, questions of classroom management will be a focus as well as forms of lesson planning to avoid teaching disruptions, various forms of student work (cooperative versus competitive work) and their further development.

DOMAIN

Tested for: Teaching mathematics

Suitable for: —

Not suitable for: External teachers and non-teachers

TARGET GROUP

Tested for: In-service mathematics teachers at the secondary school level

Suitable for: Master students in mathematics teaching

Not suitable for: Non-target group

GENERAL TEST PURPOSE

Tested for: Assessing competence level

Suitable for: Assessing competence development, evaluation of further training measures

Not suitable for: Other testing purposes

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Tests for CME, GPK, MCK, MPCK; error detection test (M_Speed); tests for beliefs of teachers; observational instruments for the assessment of teaching quality by external raters, also video-based scoring; self-assessment form for teachers

Suitable for: Additional use of a tool to assess the quality of teaching by students

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Group level

Suitable for: University level, state and national level, international level

Not suitable for: Case-by-case studies

APPLICATION SCENARIOS

Tested for: The tests can be used internationally for comparative purposes, as has already been shown with random samples of Chinese teachers.

Suitable for: Group comparisons for the evaluation of courses oriented towards the profession; identification of the profiles of teacher groups, i.e. strength and weakness analyses of groups.

Not suitable for: Selection processes

PERMITTED CONSEQUENCES

Tested for: —

Suitable for: Reliability at the individual level not yet tested

Not suitable for: Selection processes

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The tests are currently being subjected to further validation analyses (e.g. prediction of student performance). International comparative studies with culturally comparable or culturally very different groups are being carried out. The tests were used with Chinese mathematics teacher students in master's programs. The tests are also intended for use in the training of mathematics teachers in Germany, probably in their master's program as well as in their practical training phase.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

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

Kaiser, G.; Busse, A.; Hoth, J.; König, J. & Blömeke, S. (2015). About the Complexities of Video-Based Assessments: Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *International Journal of Science and Mathematics Education*, 13(3), pp. 369–387.

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

SECONDARY LITERATURE

Blömeke, S.; König, J.; Suhl, U.; Hoth, J. & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift für Pädagogik*, 61 (3), pp. 310–327.

Hoth, J.; Schwarz, B.; Kaiser, G.; Busse, A.; König, J. & Blömeke, S. (2016). Uncovering predictors of disagreement: ensuring the quality of expert ratings. *ZDM Mathematics Education*, 48(1-2), pp. 83–95.

Jentsch, A.; Schlesinger, L.; Heinrichs, H.; Kaiser, G.; König, J.; & Blömeke, S. (under review). Unterrichtsqualität unter einer mathematikdidaktischen Perspektive – Konzeptualisierung, Messung und Validierung. *Journal für Mathematik-Didaktik*

Yang, X.; Kaiser, G.; König, J. & Blömeke, S. (2018). Professional Noticing of Mathematics Teachers: a Comparative Study between Germany and China. *International Journal of Science and Mathematics Education 17*(5), 943–963. doi.org/10.1007/s10763-018-9907-x.

Situational Judgement Test for Motivational Regulation

SUMMARY

Name: Situational Judgement Test on Strategy Knowledge for Motivational Regulation

Subject domain: All subjects

Assessed competencies: Conditional strategy knowledge for motivational regulation (knowledge about [un-]suitable strategies for motivational regulation depending on situations and requirements)

Target group: Students in German-speaking contexts

Test type: Performance test; situational judgement test

Modality: Paper-pencil or online test

Duration: 15 minutes

Test structure: The test consists of 8 (short version: 5) standardized situation descriptions on motivational problems before or during learning (learning situations: "writing a scientific paper" and "preparing for an exam") with 9 items, each on possible strategies, the suitability of which can be assessed on a 6-point Likert scale.

General test purpose: Assessing competence level and competence development (e.g. after intervention)

Application scenarios: Prediction of study success, diagnosis of learning problems, development of study performance, prediction of study termination

Not suitable for: Intermediate examinations, final examinations, accreditation, use outside of university contexts

Note for practical use: Group testing possible; scoring according to expert judgments on the difference between suitable and unsuitable strategies per situation and requirement. The test is described in detail in Steuer et al. (2019). If interested, please contact the authors.

Applied in projects: Product- and Process-Oriented Modeling and Assessment of Self-Regulated Learning Competencies in Tertiary Education (PRO-SRL) https://www.kompetenzen-im-hochschulsektor.de/pro-srl/; Product- and Process Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation (PRO-SRL-EVA) https://www.uni-augsburg.de/de/fakultaet/philsoz/fakultat/psychologie/forschung/pro-srl-eva/

Contact / Location:
Prof. Dr. Markus Dresel
University of Augsburg
Email: markus.dresel@phil.uni-augsburg.de

GENERAL INFORMATION

Name: Situational Judgement Test on Strategy Knowledge for Motivational Regulation
Applied in projects: Product- and Process-Oriented Modeling and Assessment of Self-Regulated Learning Competencies in Tertiary Education (PRO-SRL); https://www.kompetenzen-im-hochschulsektor.de/pro-srl/; Product- and Process Oriented Modeling and Assessment of Self-Regulation Competencies in Higher Education – Further Validation (PRO-SRL-EVA); https://www.uni-augsburg.de/de/fakultaet/philsoz/fakultat/psychologie/forschung/pro-srl-eva/
Contact / Location:
Prof. Dr. Markus Dresel
University of Augsburg
Email: markus.dresel@phil.uni-augsburg.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Conditional strategy knowledge for motivational regulation
Theoretical model: Heuristic model of the SRL by Dresel et al. (2015) as well as assumptions for the adaptation of specific strategies to different situations (Engelschalk, Steuer & Dresel, 2015, 2016)
Test type: Performance test; situational judgment test
Modality: Paper-pencil or online test
Test structure:
The test consists of 8 (short version: 5) standardized situation descriptions on motivational problems (low expectation of success in completing the learning task or low subjective value of the learning task) before or during learning (in two learning situations: "writing a scientific paper" and "preparation for an examination"). For each vignette, different strategies for regulating one's own motivation are presented, the suitability of which for solving the presented motivation problem is to be assessed by the test participants. The assessment of suitability is based on Likert scales from 1 (not suitable at all) to 6 (completely suitable). From the comparison with expert assessments of the suitability of the strategies, an assessment of the conditional strategic knowledge of the test participants is identified. For this purpose, the strategies which the experts clearly assessed as suitable for the given situation are compared in pairs with those strategies that they clearly assessed as unsuitable. The resulting pair differences are combined to form a score that reflects the conditional strategy knowledge of the test participants.
Item pool: 8 vignettes with 9 items, each with different strategies, which are to be judged on their suitability.
Test versions: 1 original version (8 vignettes), 1 short version (5 vignettes)

TEST QUALITY CRITERIA

Reliability: Cronbach's α ≥ .92 (Bäulke, Eckerlein & Dresel, 2018; Steuer, Engelschalk, Eckerlein & Dresel, 2019)

Validity:

Test content: Learning situations were selected as motivational problem situations ("writing a scientific paper", "preparing for an examination") which, according to separate expert interviews (Dresel et al. 2015; Steuer et al. 2015), are of great importance in many degree programs and place great demands on the self-regulation of learning. Within these two learning situations, motivational problems were constructed and presented according to a 2x2 system ("low expectation of success" vs. "low subjective value" x "problem of starting in the preactional phase" vs. "problem of continuing in the action phase"), based on standard models of motivational psychology and own research indicating that students differentiate precisely between these aspects in their motivational regulation (Engelschalk et al., 2015, 2016).

Response processes: —

Internal test structure: The vignettes reflect different motivational problem situations to ensure the validity of the test through a theoretically justified and representative sample of situations. However, the strategy knowledge measured with the method for matching strategies to situations is to be understood as an overarching factor.

Relationships with other competence indicators: In line with their hypotheses, several studies (Bäulke et al., 2018; Steuer et al., 2019) found convergent relationships with other variables: quantity and scope of motivational regulation (r =. 46), effectiveness of motivational regulation (r = .36 to r = .51), effort during studies (r = .50), procrastination (r = -.33), intention to terminate studies (r = -.23).

Consequences of testing: —

Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 15 minutes

Required material: Test booklet or online platform for test administration

Special features: Group testing possible; scoring based on pair differences

Practical example: The test was successfully used in two studies (*N*=135 and *N*=131) as an indicator for the effectiveness of a training approach to improve motivation regulation. Changes in the strategy knowledge for motivational regulation between a pre- and a post-test at the beginning and at the end of the semester were assessed. The test was able to map the changes through motivational regulation training and predict various dependent variables (e.g. effort). The participants received feedback on the results, which they could use for self-assessment and reflection on their motivational regulation behavior. Since the first development, the test has been shortened by three situations – based on feedback from students – and is also available in a short version with five vignettes. The test was also used at three universities (*N*=450) to map differences in conditional strategy knowledge between students of different subject semesters and study courses.

DOMAIN

Tested for: Mathematics, teaching, economics, psychology

Suitable for: All other courses

Not suitable for: Use outside of university contexts

TARGET GROUP

Tested for: Bachelor and master students in different semesters in Germany (Bavaria, Hesse) and Austria (Vienna)

Suitable for: All other students

Not suitable for: Use outside of university contexts

GENERAL TEST PURPOSE

Tested for: Assessing competence level and competence development (e.g. after intervention)

Suitable for: Evaluation of trainings

Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Quantity and quality of strategy use, grades, intelligence

Suitable for: General assessment of motivational regulation behavior

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for individuals over time and at course level

Suitable for: Degree course and institutional level, international level

Not suitable for: —

APPLICATION SCENARIOS

Tested for: Prediction of study success, diagnosis of learning problems

Suitable for: Prediction of study achievement development, study termination prediction

Not suitable for: Intermediate examinations, final examinations, accreditation

PERMITTED CONSEQUENCES

Tested for: Determination of the need for intervention, assessment of the motivational components of learning behavior

Suitable for: —

Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test is suitable for assessing strategy knowledge for motivational regulation and can be used for testing once or measuring changes over time, e.g. before and after an intervention. The test will continue to be used in the project "Procrastination as a risk factor for discontinuing studies: a motivational and action-regulatory perspective" (ProkRASt) and shall also be used in a subsequent transfer project. If interested in the test, please contact the authors.

The test should be used with two further instruments to ensure the quality and quantity of the strategy application (Engelschalk et al. 2017). These were also developed within the PRO-SRL and PRO-SRL-EVA projects. In combination, the three instruments represent the motivational regulation of students holistically.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Steuer, G.; Engelschalk, T.; Eckerlein, N. & Dresel, M. (2019). Assessment and relationships of conditional motivational regulation strategy knowledge as an aspect of undergraduates' self-regulated learning competencies. *Zeitschrift für Pädagogische Psychologie, 33,* 95–104. DOI:10.1024/1010-0652/a000237

Engelschalk, T.; Steuer, G. & Dresel, M. (2016). Effectiveness of motivational regulation: Dependence on specific motivational problems. *Learning and Individual Differences*, *52*, pp. 72–78. doi:10.1016/j.lindif.2016.10.011

Engelschalk, T.; Steuer, G. & Dresel, M. (2017). Quantity and quality of motivational regulation among university students. *Educational Psychology*, *37*, pp. 1154–1170. doi: 10.1080/01443410.2017.1322177

SECONDARY LITERATURE

Bäulke, L.; Eckerlein, N. & Dresel, M. (2018). Interrelations between motivational regulation, procrastination and college dropout intentions. *Unterrichtswissenschaft*. doi: 10.1007/s42010-018-0029-5

Dresel, M., Schmitz, B., Schober, B., Spiel, S., Ziegler, A., Engelschalk, T., Jöstl, G., Klug, J., Roth, A., Wimmer, B., & Steuer, G. (2015). Competencies for successful self-regulated learning in higher education: Structural model and indications drawn from expert interviews. *Studies in Higher Education, 40,* 454–470. DOI:10.1080/03075079.2015.1004236

Engelschalk, T.; Steuer, G. & Dresel, M. (2015). Wie spezifisch regulieren Studierende ihre Motivation bei unterschiedlichen Anlässen? Ergebnisse einer Interviewstudie. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47,* 14–23. doi: 10.1026/0049-8637/a000120

Sustainability Management Assessment (Ko-NaMa)

SUMMARY

Name: Ko-NaMa Assessment

Domain: Operational Sustainability Management, Business Administration

Assessed competencies: Competencies in sustainability management

Target group: Bachelor students (4th semester and higher) and master students of business administration or economics; students with basic business knowledge

Test type: Performance test (knowledge test, computer-based simulation of a business setting) and questionnaire on socio-demographic characteristics as well as interests, learning opportunities and attitudes towards sustainability.

Modality: Computer-based

Duration: 90 minutes in the short version, 120 to 150 minutes in the long version (a modular design allows for increased measuring precision by extending the test time with regard to the individual test components). Digital test booklets can be compiled for specific purposes.

Test structure: The assessment consists of 3 test modules: Module I consists of 3 knowledge tests (184 items). Module II presents a simulation of a business setting (13 operating situations with 72 items for the model company) to measure the application of different types of knowledge (declarative, procedural, schematic and strategic). Module III consists of a questionnaire to assess socio-demographic characteristics as well as interests, learning opportunities and attitudes.

General test purpose: Measuring the competence level and competence structure in the field of sustainability management; measuring competence development

Application scenarios: Competence measurement, evaluation of learning outcomes

Not suitable for: Company and business contexts; students in the first three semesters of study

Note for practical use: An open source software is used, the learning management system ILIAS was used as a test environment. All test components, including questionnaires, are set up there. The scoring of the collected responses is largely automated for the declarative components. The simulation contains a number of open-response items (e.g. for decision reasons) that have to be scored manually. A scoring manual is available.

Applied in projects: Simulation-based measurement and validation of a competence model for sustainability management (Ko-NaMa)

Contact / Location:
Prof. Dr. Susan Seeber
University of Göttingen
Website: https://www.uni-goettingen.de/de/531774.html

Storage location: Research Data Centre of the Institute for Educational Quality Improvement (IQB)

GENERAL INFORMATION

Name: Ko-NaMa Assessment
Applied in projects: Simulation-based measurement and validation of a competence model for sustainability management (Ko-NaMa)
Contact / Location:
Prof. Dr. Susan Seeber
University of Göttingen
Website: https://www.uni-goettingen.de/de/531774.html

Storage location: Research Data Centre of the Institute for Educational Quality Improvement (IQB)

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Competencies in sustainability management
Theory model: The theoretical competence model has a four-dimensional structure: (1) declarative knowledge about sustainability from a general social perspective, (2) declarative knowledge in "classical" business administration, (3) declarative knowledge about sustainability management, and (4) the ability to generate strategies and justifications for specific options in terms of sustainability in entrepreneurial action situations, which are derived from economic, ecological and social considerations.
Test type: Performance test (knowledge test, computer-based simulation of a business setting) and questionnaire for sociodemographic data, interests, learning opportunities and beliefs regarding sustainability.
Modality: Computer-based
Test structure:
The assessment consists of 3 test modules: The *first* test module consists of 3 knowledge tests with 80 items on declarative knowledge about business administration, 53 items on declarative knowledge about sustainability from a societal perspective and 51 items on declarative knowledge about sustainability management and sustainability controlling. During a 10-minute test period, the participants completed approximately 22 to 24 items per knowledge test. To measure the application of different types of knowledge (declarative, procedural, schematic and strategic) a company

simulation is presented in test module II. For this purpose, 13 operational situations with a total of 72 items were constructed for the context of a model company. Each situation comprises several items. The situations are simulated via video vignettes, audio clips and (fictional) company correspondence, and include action situations in strategic management, product development and procurement processes (supplier selection, purchase of machinery), marketing, further training, transportation and customer service, complaint. The situations were arranged in a balanced booklet design so that each study participant only considers three to four of the situations. A processing time of 45 minutes was set for this test module. Test module III consists of a questionnaire on socio-demographic characteristics, interests, learning opportunities and attitudes. The test time is approximately 15 minutes.

TEST QUALITY CRITERIA

Reliability: The psychometric quality depends on testing time. Good reliability is achieved from 20 minutes and upwards for each declarative knowledge component in test module 1. Regarding the simulation, there is only experiential data on a 45-minute testing time. Taking into account the heterogeneity of the items and the limited number of items that a participant can respond to in 45 minutes, the test shows satisfactory reliability, which allows for statements about groups to be made. A 45-minute test period is not suitable for an individual diagnosis in the field of practical competences; for this purpose, a test time of 60 to 80 minutes is recommended. In the test components of module 1 (measurement of the declarative knowledge), the following EAP/PV reliabilities are achieved: (1) Declarative knowledge about sustainability from a societal perspective: .59, (2) declarative knowledge about business administration: .51 and (3) declarative knowledge about sustainability management: .57. With a testing time of 45 minutes in test module 2 (company simulation), an EAP/PV reliability of .55 is achieved. On average, 3-4 of a total of 13 situations were processed.

Validity:

Test content: According to a university teacher survey on the curricular validity of the test items, students are familiar with the business contents. In corporate sustainability management there are differences in curricular validity, since corporate sustainability management is often unsystematically anchored and integrated in the sub-disciplines of business administration as an optional module. In addition, sustainability management is a relatively new branch of business administration. Standardization (development of textbooks, content coordination) is still in the development stage.

Response processes: —

Internal test structure: —

Relationships with other competence indicators: —

Consequences of testing: —

Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 90-150 minutes

Testing materials: Computers or tablets, pocket calculators, headphones, wireless network via mobile servers and two high-performance routers (for fewer than 30 participants)

Special features: Scoring is largely automated for the declarative components. The simulation contains both closed- and open-response items (e.g. reasoning) that have to be scored manually. A scoring manual is available. The evaluation duration depends on the number of test participants.

Practical example: The test was administered to 872 bachelor students of various economics courses.

DOMAIN

Tested for: Operational sustainability management

Suitable for: Business and economics

Not suitable for: Contexts outside of business and economics

TARGET GROUP

Tested for: Bachelor students (4th semester and higher) and master students of business administration, economics, economic education

Suitable for: Students with basic business knowledge (e.g. business studies as a minor subject)

Not suitable for: Students in the 1st-3rd semester and test takers from domains other than economics

GENERAL TEST PURPOSE

Tested for: Measuring the competence level and the competence structure as well as determining explanatory factors for competence development

Suitable for: Measuring competence development

Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Motivation, attitudes, beliefs regarding sustainability and socio-demographic data. Questionnaires on motivation, attitudes, beliefs about sustainability and socio-biographical data are included in the instrument.

Suitable for: —

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at course and university level.

Suitable for: Individual diagnostics with extension of the minimum testing time from 90 to 120 – 150 minutes (with 20 minutes per declarative knowledge test component (total 60 minutes) and 60 – 90 minutes for the simulative component, i.e. 4 or 5 scenarios); international level (requires curricular examinations); longitudinal studies

Not suitable for: —

APPLICATIONS

Tested for: Competence measurement, evaluation of learning outcomes

Suitable for: Adaptation for examination purposes possible

Not suitable for: Admission procedures at universities

PERMITTED CONSEQUENCES

Tested for: —
Suitable for: Competence test in sustainability management, if there is a good agreement between the content taught at the respective university and the test
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test will not be further developed in terms of content, nor will it be updated or technologically adapted. Regular administration of the test is planned within a module on sustainability management at the Faculty of Economic Sciences and within the Central Institute for Languages and Transferable Skills (ZESS) courses at the Georg August University Göttingen.
The instrument is available at: Research Data Centre of the Institute for Educational Quality Improvement (IQB).

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SECONDARY LITERATURE

Anke, J. M. (2019). *IT-gestützte Lern- und Assessmentmodule für nachhaltiges Wirtschaften. Empirische Erkenntnisse und Gestaltungsansätze zum Einsatz IT-gestützter Lern- und Assessmentmodule*. Göttingen: Cuvillier Verlag.
Michaelis, C.; Aichele, C.; Hartig, J.; Seeber, S.; Dierkes, S.; Schumann, M.; Anke, J. M.; Siepelmeyer, D. & Repp, A. (2020). Impact of Affective-Motivational Dispositions on Competence in Sustainability Management. In Zlatkin-Troitschanskaia, O.; Pant, H. A.; Toepper, M. & Lautenbach, C. (Eds.): *Student Learning in German Higher Education. Innovative Measurement Approaches and Research Results.* Wiesbaden: Springer.
Seeber, S.; Michaelis, C.; Repp, A.; Hartig J.; Aichele, C.; Schuman, M.; Anke, J. M.; Dierkes, S. & Siepelmeyer, D. (2019). Assessment of Competencies in Sustainability Management: Analyses to the Construct Dimensionality. *Zeitschrift für pädagogische Psychologie*. 33 (2), pp. 1–11.

Teachers' Action-related and Reflective Competence: ARKOM-Business and Economics

SUMMARY

Name: Teachers' Action-related and Reflective Competence: ARKOM-Business and Economics

Domain: Business and economics education

Assessed competencies: Economics teachers' action-related and reflective competence

Target group: (Pre-service) teachers in business and economics (students of economics education (bachelor, master), trainees and teachers of business and economics at vocational study seminars ("Referendariat") and vocational schools); (pre-service) economics teachers in the US area

Test type: Performance test, open-response format

Modality: Computer- and video-based; audio recording for assessing action-related competence (AC) with speed component; written input for assessing reflexive competence (RC)

Duration: 60 minutes

Test structure: 14 video-based tasks: AC: 7 tasks; RC: 7 tasks

General test purpose: Assessing the competence level and competence development in both facets AC and RC

Application scenarios: Comparisons of competence levels between training phases in teacher education (bachelor, master, practical training phase, teaching service); comparisons between pre-service and in-service teachers in the subjects of business and economics and/or mathematics; intervention studies with a pre-post measurement design to assess the development of competence among pre-service teachers in business and economics.

Not suitable for: Entry diagnostics; beginning students

Note for practical use: The test administrator manual includes organizational and technical information to ensure standardized test administration; a coding manual is provided for objective scoring; scorer training is required.

Applied in projects: Assessing Subject-specific Competencies in Teacher Education in Mathematics and Business and Economics – a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi), https://www.eng.elmawi.de/; teaching project "Promotion of Action-oriented Competencies Using a Video-based Learning Tool for Simulating Authentic Teaching Situations Among Students of Business and Economics Education", international cooperation project "Feasibility Study for Adaptation and Use of a German-language Instrument for Assessing the Subject-related Didactic Competencies of Economics Teachers in the USA"

Contact / Location:
Dr. Christiane Kuhn, Hannes Saas
Johannes Gutenberg University Mainz
Email: Christiane.kuhn@uni-mainz.de; hannes.saas@uni-mainz.de

GENERAL INFORMATION

Name: Teachers' Action-related and Reflective Competence: ARKOM-Business and Economics
Subject domain: Business and economics Education
Applied in projects: Assessing Subject-specific Competencies in Teacher Education in Mathematics and Business and Economics – a Quasi-experimental Validation Study with a Focus on Domain-Specificity (ELMaWi), https://www.eng.elmawi.de/; teaching project "Promotion of Action-oriented Competencies Using a Video-based Learning Tool for Simulating Authentic Teaching Situations Among Students of Business and Economics Education", international cooperation project "Feasibility Study for Adaptation and Use of a German-language Instrument for Assessing the Subject-related Didactic Competencies of Economics Teachers in the USA"
Contact / Location:
Dr. Christiane Kuhn, Hannes Saas
Johannes Gutenberg University Mainz
Email: Christiane.kuhn@uni-mainz.de; hannes.saas@uni-mainz.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Business and economics teachers' action-related and reflective competence
Theory model: Competence structure model with the two facets AC and RC based on Kuhn (2014), Lindmeier (2011); see also Zlatkin-Troitschanskaia et al. (2019), Kuhn et al. (2018; 2020).
Test type: Performance test; open-response format (Kuhn et al. 2018)
Modality: Computer- and video-based: Audio recording for assessing AC with speed component; written input for assessing RC
Test structure:
Item pool: 14 video-based tasks: AC: 7 tasks; RC: 7 tasks

TEST QUALITY CRITERIA

Reliability: Cronbach's alpha: for AC .64 (N=390, 7 tasks), for RC .61 (N=384, 7 tasks)
Validity:
Test content: Validation of item content by means of a standardized expert survey (N=41) (Kuhn et al. 2018)
Response processes: Cognitive validation of items in cognitive interviews using the think-aloud method (N=44) (Zlatkin-Troitschanskaia et al. 2019)
Internal test structure: The one-dimensional models of AC and RC are confirmed in CFA.
Relationships with other competence indicators: As assumed, correlations with CK and PCK in business and economics and mathematics as well as with AC and RC in mathematics and with generic constructs (e.g. intelligence, ambiguity tolerance) (Jeschke et al. 2019; Kuhn et al. 2020)
Consequences of testing: As assumed, differences in competence levels between the various phases (bachelor, master, practical training phase, teaching service) (Kuhn et al. 2020)
Test fairness: Due to the language difficulty, a language bias can be assumed.

PRACTICAL USE

GENERAL INFORMATION

Duration: 60 minutes
Testing materials: Computers or laptops with the test program installed, headsets with microphones
Special features: The test administrator manual includes organizational and technical information to ensure a standardized test administration. A coding manual is provided for objective scoring; scorer training is required.
Practical example: In an intervention study with pre-post measurement design, the competence development of AC and RC in business and economics was analyzed in a university course in teacher training (Saas et al. 2020).

DOMAIN

Tested for: Business and economics education; commercial-administrative domain
Suitable for: Economics education in the general education sector; further domains with business and economics content
Not suitable for: —

TARGET GROUP

Tested for: (Pre-service) teachers in business and economics (bachelor and master students of business and economics education, trainees and teachers in business and economics at vocational study seminars and schools in Germany)
Suitable for: (Pre-service) economics teachers in the U.S.: an adapted and synchronized version of the economics tasks and videos is available and has been used for validation purposes in the U.S.; adaptations for use in other countries are conceivable.
Not suitable for: Beginning students

GENERAL TEST PURPOSE

Tested for: Assessment of competence level in both facets and comparisons of competence level between the training phases of teacher education (bachelor, master, practical training phase, teaching service); comparisons between pre-service and in-service business and economics and/or mathematics teachers; intervention study with pre-post-measurement design to assess the competence development of pre-service teachers in business and economics
Suitable for: Assessment of competence development in all training phases; comparisons with other school subjects
Not suitable for: Entry diagnostics

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Video-based tests for assessing AC and RC in mathematics for domain comparisons; tests for assessing PCK and CK in business and economics and mathematics; tests for assessing generic competencies (e.g. intelligence, ambiguity tolerance) for domain-specificity analyses (Jeschke et al. 2019; Kuhn et al. 2020)
Suitable for: Further tests for assessing subject-related or generic competencies
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at the group level, course level, university level.
Suitable for: International level
Not suitable for: —

APPLICATIONS

Tested for: Comparisons of competence level between training phases in teacher education (bachelor, master, practical training phase, school service); comparisons between business and economics and mathematics (for students, trainees, in-service teachers); intervention study with pre-post measurement design to assess the competence development of students in a university course in business and economics education
Suitable for: Assessment of competence development in all training phases; comparisons with other school subjects
Not suitable for: Entry diagnostics

PERMITTED CONSEQUENCES

Tested for: Assessment of competence level in university teacher training, practical training phase, teaching service
Suitable for: Assessment of competence development over the various phases; statistical prediction of competence level at a later point in time
Not suitable for: Entry diagnostics

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test can be extended with further content dimensions (e.g. in the form of additional tasks). In addition to adaptation and synchronization for (pre-service) teachers of economics in the U.S. which has already been used in a validation study in the U.S. the test is also available for adaptations in other countries. The test for assessing the AC and RC in business and economics can be requested from the Mainz project team. In the future, the test will also be used in the project "Promotion of Subject-specific Competencies of Pre-service Teachers in Mathematics and Business & Economics Using Video-based ELMaWi Tools (ELMaWi-Transfer)" as well as in the project "Technology and Economics: Integrated Didactics (TWIND)".

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Kuhn, C.; Zlatkin-Troitschanskaia, O.; Lindmeier, A.; Jeschke, C.; Saas, H. & Heinze, A. (2020). Relationships between domain-specific knowledge, generic attributes, and instructional skills – Results from a comparative study with pre- and in-service teachers of mathematics and economics. In Zlatkin-Troitschanskaia, O.; Pant, H. A.; Toepper, M. & Lautenbach, C. (Eds.): *Student Learning in German Higher Education: Innovative Measurement Approaches and Research Results.* Wiesbaden: Springer.

Kuhn, C.; Zlatkin-Troitschanskaia, O.; Brückner, S. & Saas, H. (2018). A new video-based tool to enhance teaching economics. *International Review of Economics Education, 27*, 24–33, https://doi.org/10.1016/j.iree.2018.01.007

Saas, H.; Kuhn, C. & Zlatkin-Troitschanskaia, O. (2020). Ein videobasiertes Lehr-Lernformat als innovativer hochschuldidaktischer Ansatz in der wirtschaftspädagogischen Lehrerausbildung. In: *Zeitschrift für Erziehungswissenschaft.*

Zlatkin-Troitschanskaia, O., Kuhn, C., Brückner, S. & Leighton, J. P. (2019). Evaluating a technology-based assessment (TBA) to measure teachers' action-related and reflective skills. *International Journal of Testing (IJT), 19*(2), 148–171. https://doi.org/10.1080/15305058.2019.1586377

SECONDARY LITERATURE

Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung.* Empirische Berufsbildungs- und Hochschulforschung, vol. 2. Landau: Verlag Empirische Pädagogik.

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

Teachers' Competence in German as a Foreign Language (DaZKom-Video)

SUMMARY

Name: DaZKom-video test

Domain: Teacher training

Assessed competencies: Pre-service and in-service teachers' competence in German as a foreign language (DaZ)

Target group: Pre- and in-service teachers

Test type: Video vignettes with verbal, recorded answers (performance-related competence measurement); open-ended format

Modality: Computer-based

Duration: 45-60 minutes

Test structure: 12 video vignettes with two items each; the participants are shown videos depicting DaZ-relevant classroom situations (1-2 minutes) with a brief context description and are asked to react to them; additional survey of DaZ-relevant learning opportunities and sociodemographic data

General test purpose: Assessment tool for seminars in teacher training, advanced training and further education for teachers in the field of German as a foreign language (DaZ)

Application scenarios: Used in three pilot studies to date; potential for higher education evaluation, evaluation of teacher training and further education, demand planning in higher education teaching

Not suitable for: Application to individuals, statements at individual level (admission and study-related statements)

Note for practical use: To ensure that the assessment can be conducted in a standardized manner, the provided organizational and technical information should be considered. A coding manual with a points system (0-2 points) is provided for evaluation, which allows trained coders to code the responses.

Applied in projects: GSL-Competency-Video (DaZKom-Video); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/dazkom-video/

Contact / Location:
Prof. Dr. Barbara Koch-Priewe
Bielefeld University
Email: bkoch-priewe@uni-bielefeld.de
Website: https://www.uni-bielefeld.de/erziehungswissenschaft/ag4/projekte/dazkom-video.html

GENERAL INFORMATION

Name: DaZKom-video test
Applied in projects: GSL-Competency-Video (DaZKom-Video); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/dazkom-video/
Contact / Location:
Prof. Dr. Barbara Koch-Priewe
Email: bkoch-priewe@uni-bielefeld.de
Website: https://www.uni-bielefeld.de/erziehungswissenschaft/ag4/projekte/dazkom-video.html

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: German as a foreign language (DaZ) competence of pre-service and in-service teachers
Theory model: DaZKom model (Köker et al. 2015), competence as a continuum (Blömeke et al. 2015), expert model according to Dreyfus & Dreyfus (1989)
Test type: Performance-related competence measurement; video vignettes with verbal answers
Modality: Computer- and video-based; offline testing
Test structure:
Item pool: 12 video vignettes with two items each (24 items); additional survey of sociodemographic data and DaZ-relevant learning opportunities

TEST QUALITY CRITERIA

Reliability: IRT reliability $\alpha = 0.6$
Validity:
Test content: —
Response processes: —
Internal test structure: Verification of dimensional models (Lemmrich et al. 2019)
Relationships with other competence indicators: Relationships with other measures of competence (DaZ certificate, teaching experience, research in the field of DaZ)
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 45-60 minutes
Testing materials: Tablets, headsets; consent forms for participation and data processing
Special features: Appropriate room size for oral testing and rater training required. Prior to conducting the assessment, test administrators and coders receive training. The evaluation is based on a coding manual with a points system (0-2 points).
Practical example: In three pilot studies, the test was administered on tablets to 500 participants (trainee teachers and in-service teachers) at advanced and further training courses across Germany. The technical implementation works flawlessly and the participants gave positive feedback on the authenticity of the video scenarios.
The participants are shown 12 videos depicting authentic classroom situations and are then asked to respond to two questions (1. What do you perceive? 2. If you were the teacher, how would you react in this situation?)

DOMAIN

Tested for: Teachers of all school subjects and at all types of schools
Suitable for: —
Not suitable for: —

TARGET GROUP

Tested for: Pre- and in-service teachers
Suitable for: Professional development trainers, pre-service teachers
Not suitable for: Beginning students

GENERAL TEST PURPOSE

Tested for: Seminars for teacher training as well as further and advanced training for teachers in the field of German as a foreign language (DaZ)
Suitable for: Description of the competence development of student groups over the course of their studies
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Questionnaires on sociodemographic data, learning opportunities, beliefs, personal characteristics (Big Five)
Suitable for: —
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at group level.
Suitable for: Course level, university level, state and national level
Not suitable for: Entry diagnostics

APPLICATIONS

Tested for: Pre- and in-service teachers of all subjects and at all types of schools
Suitable for: Evaluation of university courses, evaluation of teacher training and further education, demand planning in university courses
Not suitable for: Admission, prediction of study success

PERMITTED CONSEQUENCES

Tested for: Appraising the level of DaZ competence of (pre-service) teachers
Suitable for: Adaptation of teacher training and professional development training to needs
Not suitable for: No consequences at individual level such as admission decision, exmatriculation, statements on study success, work performance

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test is currently being used in pilot and standardization studies. In the future, the test will be used for the evaluation of learning opportunities as well as for further training and education of teachers.

Do you have any further questions or do you need help, e.g. with practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Hecker, S.-L. & Nimz, K. (2020). Expertinnen-Ratings zur Deutsch-als-Zweitsprache-Kompetenz von Lehrkräften: Eine Vorstudie zur Konzipierung eines videobasierten Testinstruments. *Berlin-Brandenburger Beiträge zur Bildungsforschung.*

Lemmrich, S.; Hecker, S.-L.; Klein, S.; Ehmke, T.; Köker, A.; Koch-Priewe, B. & Ohm, U. (2019). Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften. Skalierung und dimensionale Struktur des Testinstruments. In Ehmke, T.; Kuhl, P. & Pietsch, M. (Eds.): *Lehrer. Bildung. Gestalten. Beiträge zur empirischen Forschung in der Lehrerbildung.* Weinheim & Basel: Beltz Juventa, pp. 188–202.

Nimz, K.; Hecker, S.-L. & Köker, A. (2018). Videobasierte Messung von DaZ-Kompetenz bei Lehrkräften. In Caruso, C.; Hofmann, J.; Rohde, A. & Schick, K. (Eds.): *Sprache im Unterricht. Ansätze, Konzepte, Methoden.* Trier: WVT, pp. 439–452.

Teachers' Competence in German as a Second Language (DaZKom)

SUMMARY

Name: DaZKom test

Domain: German as a second language in teacher training for all school subjects and school types

Assessed competencies: German as a second language (DaZ) competence for pre-service and in-service teachers

Target group: Pre-service and in-service teachers

Type of test: Competence assessment

Modality: Paper-pencil test; computer-based

Duration: 40 minutes for the short version; 60 minutes for the long version

Test structure: Stimuli with 2-6 items; additional survey of sociodemographic data / DaZ-relevant learning opportunities

General test purpose: Assessment tool for seminars in teacher training, advanced training and further education for teachers in the field of German as a second language (DaZ)

Application scenarios: Pre-post measurements for the evaluation of DaZ-relevant learning opportunities

Not suitable for: Selection decisions (admission to degree courses)

Note for practical use: Test administrator manual; a coding manual with a points system (0-2 points) is provided for scoring

Applied in projects: *National projects*: Sprachen-Bilden-Chancen (https://www.sprachen-bilden-chancen.de/), Professional teachers' actions to promote subject-based learning under changing social conditions (ProfaLe; https://www.profale.uni-hamburg.de/en/projekt.html), BiProfessional (https://www.uni-bielefeld.de/einrichtungen/biprofessional/), Continuing Education Initiative Nordrhein-Westfalen (NRW); *International projects*: e-Learning Communities for Academic Language Learning in Mathematics and Science, (eCALLMS; https://sehd.ucdenver.edu/ecallms/about/); International Consortium for Multilingual Excellence in Education (ICMEE; https://cehs.unl.edu/icmee/); University of Nebraska-Lincoln, USA; Diversity in Education (DivEd; http://dived.fi/en/); University of Turku, Finland

Contact / Location:
Prof. Dr. Barbara Koch-Priewe
Bielefeld University
Email: bkoch-priewe@uni-bielefeld.de

GENERAL INFORMATION

Name: DaZKom test
Applied in projects: *National projects*: Sprachen-Bilden-Chancen (https://www.sprachen-bilden-chancen.de/), Professional teachers´ actions to promote subject-based learning under changing social conditions (ProfaLe; https://www.profale.uni-hamburg.de/en/projekt.html), BiProfessional (https://www.uni-bielefeld.de/einrichtungen/biprofessional/), Continuing Education Initiative Nordrhein-Westfalen (NRW); *International projects*: E-Learning Communities for Academic Language Learning in Mathematics and Science, (eCALLMS; https://sehd.ucdenver.edu/ecallms/about/); International Consortium for Multilingual Excellence in Education (ICMEE; https://cehs.unl.edu/icmee/); University of Nebraska-Lincoln, USA; Diversity in Education (DivEd; http://dived.fi/en/); University of Turku, Finland
Contact / Location:
Prof. Dr. Barbara Koch-Priewe, Bielefeld University
Email: bkoch-priewe@uni-bielefeld.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: German as a second language (DaZ) competence of pre-service and in-service teachers
Theory model: DaZKom model (Köker et al. 2015); competence as a continuum (Blömeke et al. 2015)
Test type: Competence test
Modality: Paper-pencil test; computer-based version
Test structure: 40 or 60 minutes for the DaZKom test, 10-15 minutes for questionnaire on socio-demographics, DaZ-related learning opportunities and beliefs related to multilingualism (if applicable)

TEST QUALITY CRITERIA

Reliability: WLE-reliability = 0.78 (Ehmke & Hammer, 2018)
Validity:
Test content: In the first expert review (N=7), the experts rated on a four-point Likert scale (1 = very important to 4 = unimportant) the three theoretical dimensions of the DaZ competence model, *domain-specific terminology*, (M=1.71, SD=0.49), multilingualism (M=1.71, SD=0.76) and didactics (M=1.00, SD=0.00) as very important to important (Gültekin-Karakoç, 2018) according to their relevance for the DaZ competence of future teachers. In the second expert rating (N=6), the test items were assigned to the three dimensions (rater agreement: P = 63%, Cohen's Kappa = 0.43). In addition, the relevance (M = 3.52, SD = 0.46; P = 47%) and representativeness (M = 3.31, SD = 0.67; P = 39%) of the respective test items for the dimension were assessed (four-point Likert scale: 1 = not relevant at all / representative to 4 = very relevant / representative).
Response processes: —
Internal test structure: —
Relationships with other competence indicators: Statistically significant correlations of DaZ competence with linguistic knowledge (r = 0.25) and pedagogical teaching knowledge in the sub-dimension of

dealing with heterogeneity (r = 0.21), but none with didactical mathematics knowledge (Hammer et al. 2015; 2016). Further correlations between DaZ competence with reading speed (r = 0.34) and reading ability (r = 0.26) were found; no statistically significant correlations with cognitive abilities (reasoning) (Hammer & Ehmke, 2018).
Consequences of testing: In a standard setting procedure, three standard levels ("below minimum standard", "minimum standard", "regular standard") were determined based on N = 1383 student responses and characterized in terms of content and criticism based on the item requirements (Gültekin-Karakoç et al. 2016; Gültekin-Karakoç, 2018).
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 40 minutes for the short version; 60 minutes for the long version
Testing materials: Test booklets; consent forms for data processing
Special features: Test administrator manual; the scoring is based on a coding manual with a points system (0-2 points).
Practical example: So far, the test has been used nationally in the projects Sprachen-Bilden-Chancen (https://www.sprachen-bilden-chancen.de/), Professional teachers´ actions to promote subject-based learning under changing social conditions (ProfaLe; https://www.profale.uni-hamburg.de/en/projekt.html), BiProfessional (https://www.uni-bielefeld.de/einrichtungen/biprofessional/), Continuing Education Initiative Nordrhein-Westfalen (NRW) as well as internationally in e-Learning Communities for Academic Language Learning in Mathematics and Science, (eCALLMS; https://sehd.ucdenver.edu/ecallms/about/); International Consortium for Multilingual Excellence in Education (ICMEE; https://cehs.unl.edu/icmee/); University of Nebraska-Lincoln, USA; Diversity in Education (DivEd; http://dived.fi/en/); University of Turku, Finland.

DOMAIN

Tested for: All subjects, all school types (N= approx. 1500 pre-service teachers)
Suitable for: —
Not suitable for: Application to individuals, statements at individual level (admission and study-related statements)

TARGET GROUP

Tested for: Pre-service and in-service teachers
Suitable for: —
Not suitable for: Selection decisions (admission to degree courses)

GENERAL TEST PURPOSE

Tested for: Seminars for teacher training as well as further and advanced training for teachers in the field of German as a second language (DaZ)
Suitable for: Description of the competence development of student groups over the course of a degree program, e.g. for the purpose of evaluating courses
Not suitable for: Selection decisions (admission to degree courses)

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Questionnaires on DaZ-related learning opportunities, beliefs about multilingualism at school
Suitable for: —
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at group level.
Suitable for: Course level, university level, state and national level
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Measuring the German as a Second Language Competence of pre-service and in-service teachers
Suitable for: University evaluation, evaluation of teacher training and further education, demand planning in university courses
Not suitable for: Selection decisions (admission to degree courses)

PERMITTED CONSEQUENCES

Tested for: Appraisal of the level of DaZ competence of (pre-service) teachers
Suitable for: Adaptation of teacher training and further practical training to existing needs
Not suitable for: No consequences at individual level such as admission decision, exmatriculation, statements on study success, work performance

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

Project GSL-Competency-Video (DaZKom-Video);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/dazkom-video/
Do you have any further questions or do you need help, e.g. with practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Ehmke, T. & Hammer, S. (2018). Skalierung und dimensionale Struktur des DaZKom-Testinstruments. In: Ehmke, T.; Hammer, S.; Koch-Priewe, B.; Köker, A. & Ohm, U. (eds.): *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache*. Münster: Waxmann, pp. 129–148.
Hammer, S. & Ehmke, T. (2018). Ergebnisse einer Validierungsstudie zum DaZKom-Testinstrument. In: Ehmke, T.; Hammer, S.; Koch-Priewe, B.; Köker, A. & Ohm, U. (eds.): *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache*. Münster: Waxmann, pp. 185–200.

SECONDARY LITERATURE

Ehmke, T.; Hammer, S.; Koch-Priewe, B.; Köker, A. & Ohm, U. (2018). *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache.* Münster: Waxmann.
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 GSLTestinstruments. In: Blömeke, S. & Zlatkin-Troitschanskaia, O. (Eds.). Kompetenzen von Studierenden. *Zeitschrift für Pädagogik*, supplement 61, pp. 32–54. Weinheim: Beltz.
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: Koch- Priewe, B.; Köker, A.; Seifried, J. & Wuttke, E. (Eds.), *Welche Kompetenzen brauchen Lehramtsstudierende und angehende ErzieherInnen? Theoretische und empirische Zugänge*, pp. 177–206. Bad Heilbrunn: Klinkhardt.

Test on Teachers' Educational Knowledge (BilWiss-2.0)

SUMMARY

Name: BilWiss-2.0 test

Domain: Teacher training

Assessed competencies: Educational knowledge of (pre-service) teachers

Target group: Teachers in training (teacher training, preparatory service), in-service teachers

Test type: Performance test (multiple-choice test with single- and multiple-choice tasks)

Modality: Paper-pencil test; online version available

Duration: 45 minutes for the short version; approx. 15–20 minutes per subscale in the long version

Test structure: 119 closed multiple-choice items (short version: 65 items) in the 6 content areas (scales): classroom design; learning & development; diagnostics & evaluation; educational theory; school as an educational institution; teaching profession. In the long version 23, 24, 21, 14, 19, 18 items per scale; in the short version 15, 10, 12, 9, 9, 10 items per scale.

General test purpose: Investigation of the knowledge level and knowledge development of teacher groups

Application scenarios: Prediction of performance development, (course) comparisons, study success prediction, dropout prediction, (university) evaluations, formative assessment

Not suitable for: Individual diagnostics, approval decisions, tests, accreditations

Note for practical use: The use of the long version within a test session should be avoided due to test length; the use of individual scales of the long version or the use of the short version is recommended. The test, information on the test procedure and evaluation can be requested from the BilWiss project team in Frankfurt.

Applied in projects: Returns and development of educational knowledge—Validation of a competence test for teacher students (BilWiss-UV);
Website: http://bilwiss.paedpsych.de

Contact / Location:
Simone Emmenlauer
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 6
60629 Frankfurt am Main
Email: emmenlauer@paed.psych.uni-frankfurt.de

Research data at the Research Data Centre (FDZ) at the Institute for Educational Quality Improvement (IQB):
https://www.iqb.hu-berlin.de/fdz/studies/BilWiss

GENERAL INFORMATION

Name: BilWiss-2.0-test
Applied in projects: Returns and development of educational knowledge—Validation of a competence test for teacher students (BilWiss-UV);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/bilwiss-uv/
Contact / Location:
Website: http://bilwiss.paedpsych.de

Research data at the Research Data Centre (FDZ) at the Institute for Educational Quality Improvement (IQB):
https://www.iqb.hu-berlin.de/fdz/studies/BilWiss

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Educational knowledge of (pre-service) teachers
Theoretical model: Model of teachers' professional competence (Baumert & Kunter, 2006); Classification of teachers' professional knowledge according to Schulman (1986); Classification of educational science topics based on the BilWiss Delphi study (Kunina-Habenicht et al. 2012)
Test type: Performance test (multiple-choice test with single- and multiple-choice tasks)
Modality: Paper-pencil test; online version available

Test structure:

Item pool: 119 closed multiple-choice items in the 6 content areas (scales): lesson design (23 items), learning & development (24 items), diagnostics & evaluation (21 items), educational theory (14 items), school as educational institution (19 items), teaching profession as profession (18 items)

Test versions: Long version (119 items; use of individual scales possible); short version (65 items, each 15, 10, 12, 9, 9, 10 items per above scale)

TEST QUALITY CRITERIA

Reliability: EAP/PV reliabilities of the scales from the IRT scaling: **Long version:** lesson design (.72), learning & development (.52), diagnostics & evaluation (.59), education theory (.65), school as an educational institution (.62), teaching profession as a profession (.61) **Short version:** Structure of instruction (.71), Learning & development (.49), Diagnostics & evaluation (.57), Theory of education (.64), School as an educational institution (.60), Teaching profession as profession (.59). The EAP/PV reliability of the short test is .88. Internal consistency of the short test: Cronbach's $\alpha = .86$.

Validity:

Test content: Delphi study with experts from the first and second phases of teacher training in North Rhine-Westphalia provided educational topics as a basis for test development; large coverage of educational topics from the Delphi study in the long version; state representatives of the ministries of education and cultural affairs confirm relevance of the test contents for university teacher training in the respective federal states; high content fit of the test items with topics from the KMK standards for teacher training in the educational sciences

Response processes: In cognitive interviews, participants successfully responded to the test items using knowledge from university teacher training courses.

Internal test structure: The six content areas mentioned above were confirmed in confirmatory factor analyses.

Relationships with other competence indicators: Testing the correlations between educational knowledge and relevant outcome variables: buffering effect for emotional exhaustion in preparatory service; knowledge in the field of school organization predicts proactive engagement in the teaching profession; educational knowledge predicts the quality of the analysis of teaching

Consequences of testing: The test is not suitable for individual diagnostics, but is recommended exclusively for use in research (e.g. for group comparisons) and for educational monitoring.

Test fairness: The knowledge test is language-heavy. Non-native German speakers perform significantly worse.

PRACTICAL USE

GENERAL INFORMATION

Duration: 45 minutes in the short version

Required material: Paper and pen or computer, laptop, tablet or smartphone with internet access

Special features: The use of the long version within a test session should be avoided due to the test length; the use of individual scales of the long version or the use of the short version is recommended. The test, information on the test procedure and evaluation can be requested from the BilWiss project team in Frankfurt. Preparations are currently underway to store the test at the DIPF Leibniz Institute for Research and Information in Education. A 2-PL-IRT scaling as well as a scaling according to classical test theory are possible for the scoring.

Practical example: With a view to psychometric parameters, the tasks from previous projects were modified or newly generated and tested in 2016 in a pilot study with 555 student teachers. The fit of the test items with the educational studies was assessed in cognitive interviews with 60 student teachers in 2016. Following item optimization and selection, a field test (N=600 pre-service teachers) was conducted and the final version was developed. Since 2017, a longitudinal study has been carried out with student teachers at four universities in four federal states (at the beginning N=928) to assess the change in educational knowledge during participation in corresponding learning opportunities.

DOMAIN

Tested for: Teacher training – educational science (= subject-independent) part of teaching profession

Suitable for: —

Not suitable for: —

TARGET GROUP

Tested for: Pre-service teachers (university education, practical training phase) in four federal states (North Rhine-Westphalia, Hesse, Bavaria, Baden-Württemberg)

Suitable for: Pre- and in-service teachers (all over Germany)

Not suitable for: Participants without sufficient knowledge of German

GENERAL TEST PURPOSE

Tested for: Assessment of the knowledge level and knowledge development of groups of teachers

Suitable for: Intervention evaluations

Not suitable for: Individual diagnostics

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Other aspects of competence (beliefs, motivation, self-regulation), career success criteria, crystalline and fluid intelligence

Suitable for: Further knowledge tests (e.g. domain-specific knowledge, didactic knowledge)

Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Course and group level; also over time
Suitable for: State and national level
Not suitable for: Individual level, international comparisons

APPLICATIONS

Tested for: Prediction of performance development, (course) comparisons
Suitable for: Study success prediction, dropout prediction, (university) evaluations, formative assessment, educational monitoring
Not suitable for: Approval decisions, tests, accreditations

PERMITTED CONSEQUENCES

Tested for: Measures to improve teaching
Suitable for: Formative feedback
Not suitable for: Approval decisions, accreditations

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The test is regularly used to examine the development of educational knowledge during visits to learning opportunities and for different groups and locations. In addition, the influence of educational knowledge on school-based professional practice will be examined. The test can be requested from the project staff in Frankfurt for use and the data from the BilWiss research program can be requested from the IQB Research Data Centre for subsequent use.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Kunina-Habenicht, O.; Lohse-Bossenz, H.; Kunter, M.; Dicke, T.; Förster, D.; Gößling, J. et al. (2012). Welche bildungswissenschaftlichen Inhalte sind wichtig in der Lehrerbildung? *Zeitschrift für Erziehungswissenschaft, 15*, pp. 649–682. DOI: 10.1007/s11618-012-0324-6

Kunina-Habenicht, O.; Maurer, C.; Wolf, K.; Holzberger, D.; Schmidt, M.; Dicke, T. et al. (2020). Der BilWiss-2.0-Test: Ein revidierter Test zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften. *Diagnostica.*

Kunter, M.; Kunina-Habenicht, O.; Baumert, J.; Dicke, T.; Holzberger, D.; Lohse-Bossenz, H. et al. (2016). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung – Ergebnisse des Projekts BilWiss. In: Gräsel, C. & Trempler, K. (eds.), *Entwicklung von Professionalität pädagogischen Personals. Interdisziplinäre Betrachtungen, Befunde und Perspektiven,* pp. 37–54. Wiesbaden: Springer-Online.

Linninger, C.; Kunina-Habenicht, O.; Emmenlauer, S.; Dicke, T.; Schulze-Stocker, F.; Leutner, D.; et al. (2015). Assessing teachers' educational knowledge: Construct specification and validation using mixed methods. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47, pp.* 72–83. doi: 10.1026/0049-8637/a000126

SECONDARY LITERATURE

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. In: *Zeitschrift für Erziehungswissenschaft, 9*, pp. 469–520. doi: 10.1007/s11618-006-0165-2

Tests for Engineering Mathematics: Categorization System (CAT)

The KOM@ING Tests for Engineering Mathematics comprise four instruments: Categorization System (CAT), an analysis grid for quantifying mathematical components in engineering tasks; PARI, a semi-structured expert interview for eliciting experts' and instructors' expectations of students' competencies regarding mathematical components; Low Inference Analysis (LIA) with a think-aloud approach for qualitative analyses of students' thought processes; and Student-Expert Solutions (SEL), for combining test results into augmented sample solutions and competence-based item (re)design.

SUMMARY

Name: Categorization System (CAT)

Domain: Electrical engineering basics, mathematical components

Assessed competencies: Competencies in classifying the content of students' test responses to engineering (basic tasks in electrical engineering)

Target group: Teachers and students of engineering (basic courses)

Test type: Auxiliary content classification grid for exams

Modality: Exams

Duration: Exam (120 minutes) plus content categorization

Test structure: Written examination with open-ended and closed items, category grid to differentiate mathematical components in students' responses to electrical engineering tasks (with partial competencies such as creating formulas and using a formula to calculate a value)

General test purpose: Auxiliary instrument for the content analysis of studens' written exam responses to basic tasks in electrical engineering; quantitative evidence to support the results of other studies

Application scenarios: Content analyses of (partial) competencies from written examinations

Not suitable for: Pure mathematics

Note for practical use: Part of a set of four joint and complementary instruments (KOM@ING Expert Interview, Student-Expert Solution, Low-Inference Analysis)

Applied in projects: Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education (KOM@ING-Paderborn); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/

Contact / Location:
Dr. Jörg Kortemeyer
Clausthal University of Technology
Email: joerg.kortemeyer@tu-clausthal.de

GENERAL INFORMATION

Name: Categorization System (CAT)
Applied in projects: Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education (KoM@ING-Paderborn); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/)
Contact / Location:
Dr. Jörg Kortemeyer
Clausthal University of Technology
Email: joerg.kortemeyer@tu-clausthal.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Competencies in classifying the content students' test responses to engineering (basic tasks in electrical engineering)
Theoretical model: Quantitative, evaluative content analyses
Test type: Auxiliary content classification grid for exams
Modality: Exams
Test structure:
Written examination with open-ended and closed items, category grid to identify mathematical components in students' responses to electrical engineering tasks (with partial competencies such as creating formulas and using a formula to calculate a value)

TEST QUALITY CRITERIA

Reliability: Based on exam scoring and inter-rater reliability (over 80%)
Validity:
Test content: Use of real exam tasks
Response processes: Verified by conducting studies on task response processes
Internal test structure: —
Relationships with other competence indicators: —
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: Exam (120 minutes) plus content categorization
Testing materials: Examination tasks and student responses under examination conditions
Special features: —
Practical example: The instrument was used to analyze student answers from a written exam in 2012 taken by 90 students at the University of Paderborn for the course Fundamentals of Electrical Engineering B (GET-B). The results of the exam analysis were reported back to the teachers. Their expectations, assessed through expert interviews (PARI), often differed from students' actual test responses and mistakes. A decisive point in the analyses was the differentiation of mathematical al sub-competencies, e.g., creating a formula vs. using it to calculate a value. The students were able to apply either both or none of these two skills: those who created the formula correctly were typically also able to determine the correct result. The exams were submitted anonymously, hence no individual feedback could be given to the participants.

DOMAIN

Tested for: Mathematical requirements in electrical engineering basics
Suitable for: Mathematical requirements in general engineering tasks in any semester
Not suitable for: Purely mathematical tasks

TARGET GROUP

Tested for: Teachers and students of engineering in basic courses
Suitable for: Engineering students in different semesters
Not suitable for: Pure mathematics

GENERAL TEST PURPOSE

Tested for: Quantitative evidence to support the results of other studies on assessing competencies, quantitative feedback on mathematical requirements
Suitable for: Classifying students' task responses for standards setting
Not suitable for: Analyses of qualitative responses

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Expert interviews, Student-Expert Solution and Low-Inference Analysis
Suitable for: Tests on mathematical competencies, e.g., the LiMSt (Learning Strategies in Mathematical Subjects); other competence measures
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful at the individual level.
Suitable for: Group level, university level
Not suitable for: State and national level, international level

APPLICATION SCENARIOS

Tested for: Analysis of mathematical al sub-competencies in written examinations
Suitable for: Analysis of responses to different exam types (e.g., oral)
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Optimization of sample task solutions for students
Suitable for: Competence-oriented redesign of entry phase of engineering degree courses
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The instrument was used in the Jörg Kortemeyer's dissertation, which is available at go.upb.de/kortemeyer. Contact has already been established with other institutions wishing to use the instrument. It will be used to optimize sample task solutions for students.

ADDITIONAL INFORMATION

The instrument is described in more detail in Chapter 3 of Jörg Kortemeyer's dissertation.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Kortemeyer, J.; Biehler, R. & Schaper, N. (2014). *Hilft der sogenannte Modellierungskreislauf Lösungsprozesse bei ingenieurwissenschaftlichen Anwendungsaufgaben besser zu verstehen?* Dortmund: Universitätsbibliothek Dortmund.
Kortemeyer, J. & Biehler, R. (2017). The interface between mathematics and engineering – problem solving processes for an exercise on oscillating circuits using ordinary differential equations. In: *Proceedings of the Tenth Conference of the European Society for Research in Mathematics Education.*
Kortemeyer, J. (2018). *Mathematische Kompetenzen in ingenieurwissenschaftlichen Grundlagenveranstaltungen: normative und empirische Analysen zu exemplarischen Klausuraufgaben aus dem ersten Studienjahr in der Elektrotechnik* (Dissertation).

Tests for Engineering Mathematics: Expert Interview (PARI)

The KOM@ING Tests for Engineering Mathematics comprise four instruments: Categorization System (CAT), an analysis grid for quantifying mathematical components in engineering tasks; PARI, a semi-structured expert interview for eliciting experts' and instructors' expectations of students' competencies regarding mathematical components; Low Inference Analysis (LIA) with a think-aloud approach for qualitative analyses of students' thought processes; and Student-Expert Solutions (SEL), for combining test results into augmented sample solutions and competence-based item (re)design.

SUMMARY

Name: KOM@ING Expert Interview (PARI)

Domain: Mathematics in electrical engineering

Assessed competencies: Competence expectations and normative requirements of teachers and other experts in regard to students' competencies in mathematics in electrical engineering

Target group: Teachers and other experts for mathematics in electrical engineering (in basic courses)

Test type: Semi-structured interview

Modality: Two-person interview

Duration: 30-60 minutes

Test structure: Three stages: Responding to tasks, reflection (five questions per solution step), didactic reconstruction (three questions per solution step, two general questions)

General test purpose: Auxiliary instrument for the optimization of exam tasks: Assessing competence expectations and normative requirements of experts and teachers

Application scenarios: Support in designing tasks and sample solutions for students, eliciting expert suggestions

Not suitable for: Pure mathematics

Note for practical use: Part of a set of four joints and complementary instruments (Categorization Systems, Student-Expert Solution, Low-Inference Analysis)

Applied in projects: Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education (KOM@ING-Paderborn), https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/

Contact / Location:
Dr. Jörg Kortemeyer
Clausthal University of Technology
Email: joerg.kortemeyer@tu-clausthal.de

GENERAL INFORMATION

Name: KOM@ING Expert Interview (PARI)
Applied in projects: "Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education" (KoM@ING-Paderborn); https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/)
Contact / Location:
Dr. Jörg Kortemeyer
Clausthal University of Technology
E-Mail: joerg.kortemeyer@tu-clausthal.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Competence expectations and normative requirements of teachers and other experts in regard to students' competencies in mathematics in electrical engineering
Theoretical model: PARI methodology (PARI: Precursor Action Response Interpretation; according to Means et al.)
Test type: Semi-structured interview
Modality: Two-person interview
Test structure:
Three stages: Responding to tasks, reflection (five questions per solution step), didactic reconstruction (three questions per solution step, two general questions)

TEST QUALITY CRITERIA

Reliability: Comparison with sample solutions from other studies; high degree of matches
Validity:
Test content: Representative content according to comparisons with curricula from other universities and examination tasks
Response processes: Comparison with sample solutions from other studies
Internal test structure: —

Relationships with other competence indicators: Part of a set of four joint and complementary instruments (Categorization System, Student-Expert Solution, Low-Inference Analysis)
Consequences of testing: Consequences for the creation of tasks and sample solutions for students
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 30-60 minutes
Testing materials: Interview guide, video camera, visualizer, task text
Special features: —
Practical example: The instrument was used at the University of Paderborn in 2013 for interviews with research associates of the LEA Institute for Power Electronics and Electrical Drives. For this purpose, the employees were asked to solve tasks from the basic course *Fundamentals of Electrical Engineering B* (GET-B) using a visualizer and to verbalize their thoughts (think-aloud method), whereby the visualizer filmed the transcript from above and recorded sound. A video camera was used to recorded gestures. In the second phase, questions were asked with reference to the instrument to better understand the participants' response process. In the third phase, participants were asked to reconstruct the tasks didactically; for example, they were asked to name possible motives for setting the task and learning goals or to identify typical mistakes made by second semester students, the target group for the task. A semi-structured interview guideline (PARI-like) was developed for the interview, which was supplemented for the subsequent interviews. The participating experts drew conclusions for their teaching from the interviews. Differences in interpretation arose from the fact that the experts acted in different roles, as task administrators and coordinators of training or as lecturers in the course.

DOMAIN

Tested for: Electrical engineering tasks in basic courses
Suitable for: General engineering tasks in courses of any semester
Not suitable for: Purely mathematical tasks

TARGET GROUP

Tested for: Engineering students in basic courses
Suitable for: Engineering students in different semesters
Not suitable for: Students of pure mathematics

GENERAL TEST PURPOSE

Tested for: Assessing competence expectations and normative requirements
Suitable for: Supplementing sample solutions for students
Not suitable for: Pure mathematics without reference to practical application

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Combinations with Student-Expert Solution, Low-Inference Analysis and Categorization
Suitable for: Lecture documents and learning instructions for students
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: The results are statistically meaningful for individual interviews, parallel interviews with several experts from different institutions.
Suitable for: Comparison with analog student interviews
Not suitable for: University comparisons, state and national level, international level

APPLICATION SCENARIOS

Tested for: Support in creating and providing sample solutions
Suitable for: Writing textbooks that take into account students' cognitive resources
Not suitable for: Pure mathematics

PERMITTED CONSEQUENCES

Tested for: Optimization of sample exam task solutions for students
Suitable for: Competence-oriented redesign of the entry phase of engineering degree courses
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The instrument was used in Jörg Kortemeyer's dissertation, which is available at go.upb.de/kortemeyer. Contact has already been established with other institutions wishing to use the instrument. It will be used to optimize sample task solutions for students.

ADDITIONAL INFORMATION

The instrument is described in more detail in Chapter 3 of Jörg Kortemeyer's dissertation.
Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Kortemeyer, J.; Biehler, R. & Schaper, N. (2014). *Hilft der sogenannte Modellierungskreislauf Lösungsprozesse bei ingenieurwissenschaftlichen Anwendungsaufgaben besser zu verstehen?* Universitätsbibliothek Dortmund.
Kortemeyer, J. & Biehler, R. (2017). The interface between mathematics and engineering – problem solving processes for an exercise on oscillating circuits using ordinary differential equations. In: *Proceedings of the Tenth Conference of the European Society for Research in Mathematics Education.*
Kortemeyer, J. (2018). *Mathematische Kompetenzen in ingenieurwissenschaftlichen Grundlagenveranstaltungen: normative und empirische Analysen zu exemplarischen Klausuraufgaben aus dem ersten Studienjahr in der Elektrotechnik* (Dissertation).

SECONDARY LITERATURE

Biehler, R.; Kortemeyer, J. & Schaper, N. (2015, February). Conceptualizing and studying students' processes of solving typical problems in introductory engineering courses requiring mathematical competences. In: *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education,* pp. 2060–2066.

Tests for Engineering Mathematics: Low-Inference Analysis (LIA)

The KOM@ING Tests for Engineering Mathematics comprise four instruments: Categorization System (CAT), an analysis grid for quantifying mathematical components in engineering tasks; PARI, a semi-structured expert interview for eliciting experts' and instructors' expectations of students' competencies regarding mathematical components; Low Inference Analysis (LIA) with a think-aloud approach for qualitative analyses of students' thought processes; and Student-Expert Solutions (SEL), for combining test results into augmented sample solutions and competence-based item (re)design.

SUMMARY

Name: Low-Inference Analysis (LIA)

Domain: Electrical engineering basics

Assessed competencies: Students' thought processes while responding to basic tasks in applied mathematics

Target group: Engineering students in basic courses

Test type: Cooperative performance test, think-aloud approach and video recording during item response and retrospective interview

Modality: Paper-pencil test

Duration: 60-90 minutes

Test structure: Video recording of item response processes and interviews

General test purpose: Auxiliary instrument for the identification of students' thought processes and competencies while responding to engineering tasks

Application scenarios: Teaching-learning tool for assessing and comparing students' thought processes when working on basic tasks

Not suitable for: Pure mathematics

Note for practical use: Part of a set of four joint and complementary instruments (KOM@ING Expert Interview (PARI), Student-Expert Solution (SES), Categorization System (CAT))

Applied in projects: Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education (KOM@ING-Paderborn);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/

Contact / Location:
Dr. Jörg Kortemeyer
Clausthal University of Technology
Email: joerg.kortemeyer@tu-clausthal.de

GENERAL INFORMATION

Name: Low-Inference Analysis (LIA)
Applied in projects: Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education (KoM@ING);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/
Contact / Location:
Dr. Jörg Kortemeyer, Clausthal University of Technology
Email: joerg.kortemeyer@tu-clausthal.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Students' thought processes while responding to basic tasks in applied mathematics
Theoretical model: Qualitative content analysis
Test type: Cooperative performance test, think-aloud approach during filmed item response and retrospective interview
Modality: Paper-pencil test
Test structure:
Video recording of item response processes and interviews

TEST QUALITY CRITERIA

Reliability: Analysis of the working processes by additional electrical engineering experts; high correspondence
Validity:
Test content: Comparison with curricula of different universities
Response processes: Comparison with student responses
Internal test structure: —
Relationships with other competence indicators: —
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 60-90 minutes
Testing materials: Task text, two cameras or one camera and one visualizer
Special features: Part of a set of four joint and complementary (KOM@ING Expert Interview (PARI), Student-Expert Solution (SES), Categorization System (CAT))
Practical example: The low-inference analyses were applied to examine the response processes of pairs of students and filmed at the Universities of Paderborn and Hannover in 2013. Transcripts were created of the students' item response processes, which were examined using methods of qualitative content analysis. In the assessment, pairs of second semester students were asked to solve a problem using the think-aloud method. The participants reported that working in teams of two revealed additional competencies among the students. Changes in study regulations and curricula are not known and require much effort to identify as the acquisition of competence occurs in two courses simultaneously (mathematics for engineering and basics of electrical engineering).

DOMAIN

Tested for: Electrical engineering basics
Suitable for: General engineering tasks in any semester
Not suitable for: Purely mathematical tasks

TARGET GROUP

Tested for: Engineering students in basic courses
Suitable for: Engineering students in different semesters
Not suitable for: Students of pure mathematics

GENERAL TEST PURPOSE

Tested for: Identification of students' mental processes and competencies while responding to engineering tasks
Suitable for: Creating an enhanced sample solution
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Combination with expert interviews, student-expert solutions and categorizations
Suitable for: Tests on mathematical competencies, e.g. the LiMSt (Learning Strategies in Mathematical Subjects); other competence measurements
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for multiple implementations of the same task in groups.
Suitable for: Analysis of student responses
Not suitable for: Individual diagnostics, university level, international level

APPLICATIONS

Tested for: Teaching-learning tool to measure students' thought processes while working on basic tasks, comparisons of verbalized thought processes
Suitable for: Creation of a competence catalogue for basic courses
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Optimization of sample task solutions for students
Suitable for: Competence-oriented redesign of initial phases in engineering study courses
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The instrument was used in Jörg Kortemeyer's dissertation, which is available at go.upb.de/kortemeyer. Contact has already been established with other institutions wishing to use the instrument. It will be used to optimize sample task solutions for students.

ADDITIONAL INFORMATION

The instrument is described in more detail in Chapter 3 of Jörg Kortemeyer's dissertation.

Do you have any further questions or do you need help, e.g., with the practical application or score interpretation? If you are interested or require assistance, please contact us!

SELECTED PROJECT PUBLICATIONS

Kortemeyer, J.; Biehler, R. & Schaper, N. (2014). *Hilft der sogenannte Modellierungskreislauf Lösungsprozesse bei ingenieurwissenschaftlichen Anwendungsaufgaben besser zu verstehen?* Universitätsbibliothek Dortmund.
Kortemeyer, J. & Biehler, R. (2017). The interface between mathematics and engineering – problem solving processes for an exercise on oscillating circuits using ordinary differential equations. In: *Proceedings of the Tenth Conference of the European Society for Research in Mathematics Education.*
Kortemeyer, J. (2018). *Mathematische Kompetenzen in ingenieurwissenschaftlichen Grundlagenveranstaltungen: normative und empirische Analysen zu exemplarischen Klausuraufgaben aus dem ersten Studienjahr in der Elektrotechnik* (Dissertation).

SECONDARY LITERATURE

Biehler, R.; Kortemeyer, J. & Schaper, N. (2015, February). Conceptualizing and studying students' processes of solving typical problems in introductory engineering courses requiring mathematical competences. In: *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education,* pp. 2060–2066.

Tests for Engineering Mathematics: Student-Expert Solution (SEL)

The KOM@ING Tests for Engineering Mathematics comprise four instruments: Categorization System (CAT), an analysis grid for quantifying mathematical components in engineering tasks; PARI, a semi-structured expert interview for eliciting experts' and instructors' expectations of students' competencies regarding mathematical components; Low Inference Analysis (LIA) with a think-aloud approach for qualitative analyses of students' thought processes; and Student-Expert Solutions (SEL), for combining test results into augmented sample solutions and competence-based item (re)design.

SUMMARY

Name: Student-Expert Solution (SEL)

Domain: Electrical engineering tasks in basic courses

Assessed competencies: Normative solution competence for basic mathematical tasks in electrical engineering

Target group: Engineering students in basic courses

Test type: Auxiliary tool for creating sample solutions to tasks by experts

Modality: Paper-pencil-test

Duration: 10 hours (incl. familiarization with the solution given by the students, interview, creation of the expert solution)

Test structure: Two parts: "task solution", augmented with more detailed mathematical and electro-technical annotations; "cognitive resources" for additional annotations on task solution

General test purpose: Auxiliary tool for creating sample solutions to tasks; definition of a standard using expert interviews; creation of an analytical basis for qualitative and quantitative studies (LIAs; CAT)

Application scenarios: Solution to four tasks in Jörg Kortemeyer's dissertation as a basis for further qualitative analyses (LIAs) and quantitative analyses (Categorizations)

Not suitable for: Pure mathematics

Note for practical use: Part of a set of four joint and complementary instruments (Expert Interview, Categorization Systems, Low-Inferential Analysis).

Applied in projects: Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education (KoM@ING-Paderborn);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/

Contact / Location:
Dr. Jörg Kortemeyer, Clausthal University of Technology
Email: joerg.kortemeyer@tu-clausthal.de

GENERAL INFORMATION

Name: Student-Expert Solution (SEL)
Applied in projects: Modeling and developing competences – integrated IRT based and qualitative studies with a focus on mathematics and its usage in engineering education (KoM@ING-Paderborn);
https://www.blogs.uni-mainz.de/fb03-kokohs-eng/koming/
Contact / Location:
Dr. Jörg Kortemeyer
Clausthal University of Technology
Email: joerg.kortemeyer@tu-clausthal.de

THE INSTRUMENT

CONTENT AND STRUCTURE

Assessed competencies: Normative solution competence for basic mathematical tasks in electrical engineering
Theory model: Normative mathematical-didactic task analyses
Test type: Auxiliary tool for creating sample solutions to tasks by experts
Modality: Paper-pencil test
Test structure:
Two parts: "task solution", augmented with more detailed mathematical and electro-technical annotations; "cognitive resources" for additional annotations on task solution

TEST QUALITY CRITERIA

Reliability: —
Validity:
Test content: Comparison with curricula of different universities and analyses by further experts
Response processes: Comparison with student solutions
Internal test structure: —
Relationships with other competence indicators: —
Consequences of testing: —
Test fairness: —

PRACTICAL USE

GENERAL INFORMATION

Duration: 10 hours (incl. working out the short solution, which was then supplemented in the expert interviews by aspects such as competence expectations)
Testing materials: Sample solution for the tasks (e.g. for proofreaders), expert interview to assess the competencies
Special features: Part of a set of four joint and complementary instruments (Expert Interview (PARI), Categorization Systems (CAT), Low-Inferential Analysis (LIA)).
Practical example: The instrument was developed for the analysis of task responses from video studies and written exams to develop a basis for interpretation. Since 2013, the instrument has been used in studies at the University of Paderborn and the Leibniz University Hannover for the analysis of item responses and response processes. The instrument delivers comprehensive information on detailed requirements based on participants' response processes. These include many physical considerations related to electrical engineering as well as considerations to reduce the mathematical effort. Consequence of the results is the teachers' stronger focus on the expectations placed on students in the test as well as on implicit competence expectations, which was relevant in the context of producing student-expert solutions.

DOMAIN

Tested for: Electrical engineering basics
Suitable for: General engineering tasks in any semester
Not suitable for: Internal mathematical tasks

TARGET GROUP

Tested for: Engineering students in basic courses
Suitable for: Engineering students in different semesters
Not suitable for: Pure mathematics

GENERAL TEST PURPOSE

Tested for: Auxiliary tool for creating a sample solution; definition of a standard using expert interviews; creation of an analysis basis for qualitative and quantitative studies. The instrument describes the interface between the competence areas of mathematics and electrical engineering in more detail and takes particular account of the teachers' competence expectations for students
Suitable for: Creation of an extended sample solution for students
Not suitable for: —

USE WITH OTHER TESTS AND QUESTIONNAIRES

Tested for: Combination with PARI, LIA and CAT
Suitable for: Tests of mathematical competencies, e.g. the LiMSt (learning strategies in mathematical subjects), other measures of competence
Not suitable for: —

SUITABILITY

SCOPE OF APPLICABILITY

Tested for: Results are statistically meaningful for parallel compilation by several experts from different institutions.
Suitable for: Assessment basis for examinations
Not suitable for: —

APPLICATION SCENARIOS

Tested for: Solution to four tasks in Jörg Kortemeyer's dissertation as basis for further qualitative analyses (LIAs) and quantitative analyses (Categorizations)
Suitable for: All tasks that involve interfaces between mathematics and other subjects
Not suitable for: —

PERMITTED CONSEQUENCES

Tested for: Optimization of sample exam solutions for students
Suitable for: Competence-oriented redesign of the entry phase of engineering degree courses
Not suitable for: —

FURTHER INFORMATION

FURTHER DEVELOPMENT AND FUTURE USE

The instrument was used in Jörg Kortemeyer's dissertation, which is available at go.upb.de/kortemeyer where contact information is available for other institutions that would like to use the instruments. It is to be used for the optimization of task solutions for students. The test can be carried out according to the descriptions in the dissertation.

Do you have any further questions or do you need help, e.g. with the practical application or score interpretation? If you are interested, please contact us!

SELECTED PROJECT PUBLICATIONS

Kortemeyer, J.; Biehler, R. & Schaper, N. (2014). *Hilft der sogenannte Modellierungskreislauf Lösungsprozesse bei ingenieurwissenschaftlichen Anwendungsaufgaben besser zu verstehen?* Universitätsbibliothek Dortmund.
Kortemeyer, J. & Biehler, R. (2017). The interface between mathematics and engineering – problem solving processes for an exercise on oscillating circuits using ordinary differential equations. In: *Proceedings of the Tenth Conference of the European Society for Research in Mathematics Education.*
Kortemeyer, J. (2018). *Mathematische Kompetenzen in ingenieurwissenschaftlichen Grundlagenveranstaltungen: normative und empirische Analysen zu exemplarischen Klausuraufgaben aus dem ersten Studienjahr in der Elektrotechnik.* (Dissertation)

SECONDARY LITERATURE

Biehler, R.; Kortemeyer, J. & Schaper, N. (2015, February). Conceptualizing and studying students' processes of solving typical problems in introductory engineering courses requiring mathematical competencies. In: *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education*, pp. 2060–2066.

Imprint

Portfolio of KoKoHs Assessments
Test Instruments for Modeling and Measuring
Domain-specific and Generic Competencies of
Higher Education Students and Graduates

Publisher

Olga Zlatkin-Troitschanskaia, Hans Anand Pant, Marie-Theres Nagel, Dimitri Molerov, Corinna Lautenbach, Miriam Toepper

Composition and Layout

axeptDESIGN.de

Print

PRINZ-DRUCK Print Media GmbH & Co KG

Picture credits

U1 and U4: © connel_design/Fotolia

First print run: January 2020

KoKoHs