

Jähnicg, Christine Caroline

Assessing Business Knowledge of Students in German Higher Education

Faßhauer, Uwe [Hrsg.]; Fürstenau, Bärbel [Hrsg.]; Wuttke, Eveline [Hrsg.]: *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2013*. Opladen [u.a.] : Verlag Barbara Budrich 2013, S. 47-59. - (Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE))



Empfohlene Zitierung/ Suggested Citation:

Jähnicg, Christine Caroline: Assessing Business Knowledge of Students in German Higher Education - In: Faßhauer, Uwe [Hrsg.]; Fürstenau, Bärbel [Hrsg.]; Wuttke, Eveline [Hrsg.]: *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2013*. Opladen [u.a.] : Verlag Barbara Budrich 2013, S. 47-59 - URN: urn:nbn:de:0111-opus-80634 - DOI: 10.3224/978384740127

<http://nbn-resolving.org/urn:nbn:de:0111-opus-80634>
<http://dx.doi.org/10.3224/978384740127>

in Kooperation mit / in cooperation with:



<https://www.budrich.de>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:
<http://creativecommons.org/licenses/by-nc-nd/3.0/de/deed> - Sie dürfen das Werk bzw. den Inhalt unter folgenden Bedingungen vervielfältigen, verbreiten und öffentlich zugänglich machen: Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen. Dieses Werk bzw. dieser Inhalt darf nicht für kommerzielle Zwecke verwendet werden und es darf nicht bearbeitet, abgewandelt oder in anderer Weise verändert werden.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-Licence:
<http://creativecommons.org/licenses/by-nc-nd/3.0/de/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work in the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work or its contents. You are not allowed to alter, transform, or change this work in any other way.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

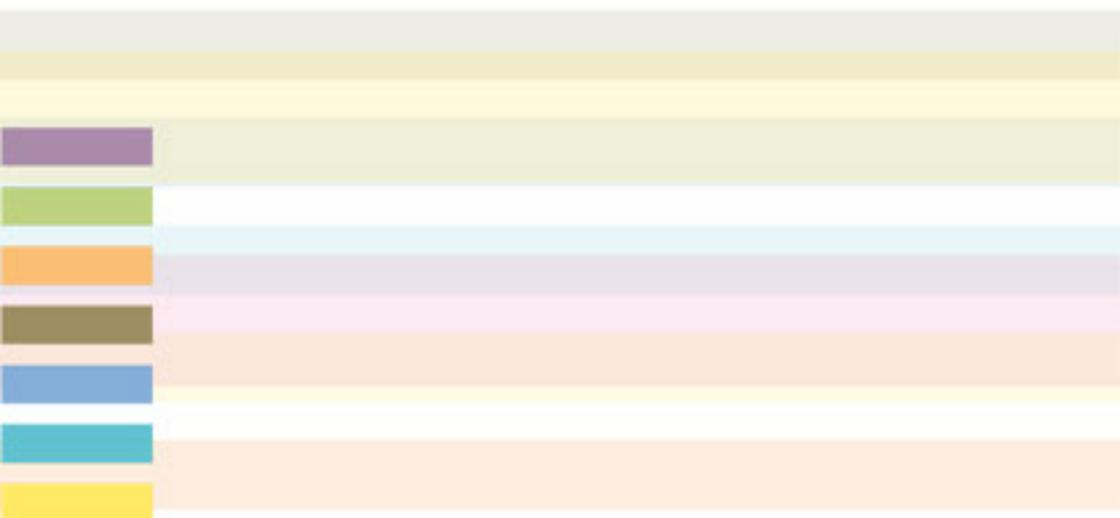
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2013

Uwe Faßhauer, Bärbel Fürstenau,
Eveline Wuttke (Hrsg.)



DGfE Deutsche Gesellschaft
für Erziehungswissenschaft

Jahrbuch der berufs- und wirtschaftspädagogischen
Forschung 2013

Schriftenreihe der Sektion
Berufs- und Wirtschaftspädagogik
der Deutschen Gesellschaft
für Erziehungswissenschaft (DGfE)

Uwe Faßhauer
Bärbel Fürstenau
Eveline Wuttke (Hrsg.)

Jahrbuch der berufs- und
wirtschaftspädagogischen
Forschung 2013

Verlag Barbara Budrich
Opladen • Berlin • Toronto 2013

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Dieses Werk ist im Verlag Barbara Budrich erschienen und steht unter folgender Creative Commons Lizenz: <http://creativecommons.org/licenses/by-nc-nd/3.0/de/> Verbreitung, Speicherung und Vervielfältigung erlaubt, kommerzielle Nutzung und Veränderung nur mit Genehmigung des Verlags Barbara Budrich.



Dieses Buch steht im OpenAccess Bereich der Verlagsseite zum kostenlosen Download bereit (<http://dx.doi.org/10.3224/978384740127>)
Eine kostenpflichtige Druckversion (Printing on Demand) kann über den Verlag bezogen werden. Die Seitenzahlen in der Druck- und Onlineversion sind identisch.

ISBN 978-3-8474-0127-8
DOI 10.3224/978384740127

Umschlaggestaltung: bettina lehfeldt graphic design, Kleinmachnow
Typografisches Lektorat: Ulrike Weingärtner, Gründau
Verlag Barbara Budrich, <http://www.budrich-verlag.de>

Vorwort	9
---------------	---

Teil I: Lehr-Lernforschung in der beruflichen Bildung

Jan Hendrik Stork

Der Einfluss unterschiedlicher Unterrichtsmodelle auf den Erwerb von mathematischen und kaufmännischen Kompetenzen im beruflichen Unterricht	11
--	----

Mandy Hommel

„Darf ich um Ihre Aufmerksamkeit bitten?!“ Wege aus dem Aufmerksamkeitstief	23
---	----

Stephan Schumann, Franz Eberle, Maren Oepke

Ökonomisches Wissen und Können am Ende der Sekundarstufe II: Effekte der Bildungsgang-, Klassen- und Geschlechtszugehörigkeit	35
---	----

Christine Caroline Jähnig

Assessing Business Knowledge of Students in German Higher Education	47
---	----

Anja Mindnich, Stefanie Berger, Sabine Fritsch

Modellierung des fachlichen und fachdidaktischen Wissens von Lehrkräften im Rechnungswesen – Überlegungen zur Konstruktion eines Testinstruments	61
--	----

Roland Happ, Susanne Schmidt, Olga Zlatkin-Troitschanskaia

Der Stand des wirtschaftswissenschaftlichen Fachwissens von Bachelorabsolventen der Universität und der Fachhochschule	73
--	----

Teil II: Didaktik und Methodik beruflichen Lernens

Nicole Kimmelmann, Wilhelmine Berg

Wie viel Deutsch darf's sein? Das Projekt „Deutsch am Arbeitsplatz“	87
---	----

Claudia Stolp, Jens Siemon

Wirkung auf Lernerfolg und Motivation durch Debriefing in Unternehmensplanspielen	99
---	----

<i>Georg Tafner</i> Supranationalität begreifbar machen. Performative Pädagogik im Planspiel.....	113
<i>Frank Arens</i> Praxisbegleitung in der Pflegeausbildung – ein blinder Fleck der Berufsbildungsforschung?	127
<i>Axel Grimm</i> Lehrerhandeln im computerunterstützten Berufsschulunterricht – Handlungsmuster von Berufsschullehrkräften in elektro- und metalltechnischen Lehr-Lernarrangements.....	139

Teil III: Hochschuldidaktik und Lehrerbildung

<i>Gabriela Kugler</i> Grundstrukturen professionellen Lehrerwissens: Wissen zur Lesekompetenzförderung von Berufsschullehrkräften	153
<i>Juliana Schlicht</i> Forschendes Lernen im Studium: Ein Ansatz zur Verknüpfung von Forschungs-, Lehr- und Lernprozessen.....	165
<i>Romy Wolff</i> Interaktionsprozesse zur Entscheidungsfindung in virtuellen und face-to-face Gruppen	177
<i>Ulrike Weyland, Karin Reiber</i> Lehrer/-innen-Bildung für die berufliche Fachrichtung Pflege in hochschuldidaktischer Perspektive	189

Teil IV: Berufs- und wirtschaftspädagogische Perspektiven

<i>Wiebke Petersen</i> Prinzipien der Regulierung des Feldes der Berufsbildung – ein Ansatz zum internationalen Vergleich von „ValNIL“	203
--	-----

<i>Stefan Wolf</i> Berufsbildung und Migration – kritische Anmerkungen über vernachlässigte Perspektiven	217
<i>Markus Linten, Christian Woll</i> Berufsbildungsdiskurs 2010 und 2011: Vergleichende Resonanzanalyse referierter und nicht-referierter Zeitschriftenbeiträge zur Berufsbildungsforschung und -praxis	233
<i>Martin Kröll</i> Das Zusammenspiel von Selbst- und Fremdorganisation in lernenden Organisationen	245
<i>Gregor Thurnherr, Samuel Schönenberger & Christian Brühwiler</i> Hilfreiche Unterstützung in der Berufsorientierung aus Sicht von Jugendlichen	259

Assessing Business Knowledge of Students in German Higher Education

Christine Caroline Jähnig

1. Introduction and Motivation

In Germany, and many other European countries, universities have begun to focus on quality assurance and quality management (Teichler 2006). This development has triggered a debate concerning students' acquisition of skills and knowledge and the adequate assessment of such learning outcomes in higher education (Zlatkin-Troitschanskaia/Kuhn 2010). Different stakeholders frequently claim that there is a gap between academic training and professional requirements (e.g. Gruber/Mandl/Renkl 2000; Reiss 2012). One possible scientific approach toward these recent criticisms is to develop and implement standardized assessments of learning outcomes in higher education. The availability of standardized test instruments potentially enables research into the constitution of students' knowledge and the supposed gaps. More importantly, criteria for successful and unsuccessful knowledge development can be identified and used as valid sources of information for all stakeholders of higher education (e.g. students, university employees and policy makers) (Black/Duhon 2003). In this paper, domain-specific knowledge is viewed as one key component of professional competence. However, for the genesis of competent action, a combination of different types of knowledge and knowledge organization is necessary, (Gruber 1999) as well as volitional, motivational and social dispositions of the individual (Weinert 2011).

Despite this positive vision, research on standardized assessment in higher education is a relatively young field in Germany. Accordingly, the availability of suitable assessment tools is limited. Surprisingly, this is also true for highly popular degrees like business studies. This lack of instruments in Germany is contrasted by recent international projects like the international¹ feasibility study on Assessment of Higher Education Learning Outcomes (AHELO) (OECD 2011) that aims at assessing what students in higher education know and can do upon graduation with a focus on engineering, economics and generic skills. To address the above described shortages, educational research must rise to the challenge of developing appropriate assessment

¹ Germany does not participate in the study.

concepts and instruments. These must fulfill the requirements of competence-oriented assessment while also complying with psychometric standards for empirical educational research. At present, only two assessment instruments for business and economics in higher education in German language are accessible. One test, with a focus on assessing aspects of economic knowledge, is the “Wirtschaftskundliche Bildungstest (WBT)” (Beck/Krumm/Dubs 1993). The WBT is an adaptation of the “American Test of Economic Literacy (TEL)” originally published by Soper and Walstad (1987). It was not explicitly developed for students in higher education and is often used in the context of Vocational Education and Training (VET), thus its transferability to higher levels of education is questionable. However, recent publications indicate adequate validity of the instrument when applied in higher education (Förster/Happ/Zlatkin-Troitschanskaia 2012). The second instrument, the Business Administration Knowledge Test (BAKT) (Bothe/Wilhem/Beck 2005), was developed to measure declarative knowledge of business students. Declarative knowledge refers to memorized factual knowledge (Anderson/Krathwohl 2001) that is not guiding action closely (Renkl 1996). Following the demand that undergraduate students should be able to “apply their knowledge and understanding in a manner that indicates a professional approach to their work or vocations” upon completing their degree (Bologna Working Group on Qualifications Frameworks 2005, p. 194), this paper argues the necessity of supplementing the existing set of tests with action-oriented items, to allow a more complete view of students’ knowledge and its application to work related situations.

The goal of this paper is to describe the development of a test that covers basic business knowledge in line with content covered by compulsory undergraduate business courses. Moreover, the test should tap into knowledge structures that go beyond declarative knowledge and capture ‘knowledge of how’ to do something. This type of knowledge is often referred to as procedural knowledge and in this context is conceptualized as knowledge about subject-specific techniques and methods, as well as knowledge of criteria when (and how) to use appropriate procedures (Anderson/Krathwohl 2001, p. 29). This demand is linked to the intended learning outcomes, stated by the AHELO project (OECD 2010) that stress the importance of subject knowledge and its application in economics (for example the ability to apply economic reasoning and methods effectively as well as the effective use of relevant data and quantitative methods (p. 25)).

This means for assessment purposes to choose a test format that differs from classical multiple choice items. Since successful mastery of a set of domain-specific situations is associated with competence in the domain (Klieme 2004), situational items yield great potential to assess knowledge constructs that are closely related to competence (Rost 2008). It is assumed that situational items with predefined answer options, function as low fidelity simulations

(Motowidlo/Dunnette/Carter 1990) and evoke the application of business knowledge in a procedural manner that is relevant for the future work fields of business students.

2. Principles of Test Construction

The following chapter outlines the central steps of test development. First, a sketch of the domain and its relation to test construction is outlined. Second, the composition of the items is described.

2.1 Describing the domain for test construction

From a cognitive perspective, learning processes are to a large extent domain-specific (Winther 2010). In consequence, the development of test instruments should be guided by the understanding of the domain. A domain in this context is understood as business specific areas of performance that can be characterized through selected situational requirements (Winther 2010). Analyzing and modeling the domain for a test of business knowledge in higher education has presented difficulties. First, there is no overarching core curriculum that allows the extraction of information about a common understanding of content and intended effects of teaching business at universities (Förster/Happ/Zlatkin-Troitschanskaia 2012). Analysis of the intended business curriculum of German universities, on the basis of course descriptions, reveals the heterogeneity of learning opportunities for students. Second, a task-oriented approach to analyze the domain is hindered by the diverse field of employment opportunities for young adults with a bachelor's degree (Henning/Henning 2009). These conditions leave room for interpretation when trying to extract domain-specific requirements of learning and working.

For this paper, two perspectives were used to identify domain-specific demands. Work requirements were extracted from job advertisements for young professionals with a business degree. The second perspective was taken from the intended curriculum of three major universities. Overlapping course content was identified in the compulsory courses in management, production, marketing and accounting in order to extract domain-specific learning requirements. Only courses on undergraduate level were included into the analysis because compulsory courses on bachelor-level yield the largest amount of content overlap. Thus, the developed test focusses on assessing basic business knowledge related to undergraduate courses at German universities. This pragmatic approach was followed to ensure curricular validity of item content as well as to construct items that resemble real-world problems.

2.2 Conceptual framework and item construction

After identifying central elements of the domain, the domain-specific requirements need to be translated into an assessment instrument.

In this paper, knowledge is broadly organized into declarative and procedural knowledge systems. This assumption follows the cognitive architecture by Anderson (1983) that separates a factual knowledge system and a procedural knowledge system. The declarative (factual) knowledge system is organized by semantic networks; the procedural knowledge system is organized by production rules. Production rules in the sense used by Anderson are condition-action pairs, where the action or operation is activated if certain preconditions are satisfied.

The test aims at capturing mainly procedural knowledge representations of students. This requirement influences the content as well as the structure of the items. In accordance with the definition of procedural knowledge by Anderson and Krathwohl (2001), the item content focuses on domain-specific techniques and methods and on knowledge about criteria for determining when (and how) to use appropriate procedures. This content was embedded into situations that were described in the item stem. The situational framework of the items was mainly inspired by job advertisements and situational tasks in textbooks, like business cases. Textbooks and two workshops with teaching staff in business education plus one workshop with master students in business education guided the development of distracters and the refinement of item wording. To evoke the activation of a production rule when answering the item, the goal of the operation was clearly stated in each item stem. Preconditions of different complexity levels were incorporated into the item stem. The response options were predefined and presented in a multiple-choice format with a single right answer. The response options were formulated as operations to be carried out as reactions to the described problem. An example item is presented in Fig. 1. The item refers to the course content ‘decision problems in controlling’ using scoring models in particular. Being able to deliver decision memos is a frequent request in job advertisements for young business professionals. The situation was adapted from a textbook by Horváth/Gleich/Voggenreiter (2007).

Fig. 1: Freely translated example item from the test with low complexity.

The company you work for has different options for outsourcing. Your **goal** is to support the decision process by judging the outsourcing options with regard to the following criteria: quality, core competencies, capital accumulation, flexibility, independency, customer orientation and process orientation. How do you go about it?

- I write a summary for each outsourcing option and each criterion.*
- I calculate values for each outsourcing option and rank them by calculated value.*
- I rank the outsourcing options by my own expert rating.*
- I score each criterion for each outsourcing option and calculate a total score.*
- I rank the outsourcing options by my own expert rating.*
- I score each criterion for each outsourcing option and calculate a total score.*

The item has been categorized as a low complexity item because no preconditions were stated. The item would have been of higher complexity if, for example, the item stem had informed the student that the criteria are prioritized. Although business problems in real life are not delivered with predefined answer options, such a format was chosen to ensure scoring objectivity. To account for the ambiguity of real-life situations, at least one of the response options contained partially correct elements but was not scored as a correct response.

To develop items of continuously distributed difficulty, dimensions of cognitive processes should vary across the items. According to an adaptation of Anderson and Krathwohl's (2001) taxonomy by Hofmeister (2009) it was tried to develop items which stimulate processes of remembering, applying or reflecting. But at the end of item construction no clear assignment of items to categories was possible. Hence, complexity, defined by the number of preconditions, was used as an alternative cognitive process dimension to characterize the items.

3. Resulting Test Characteristics

A first version of the test with 45 situational items was piloted at the University of Göttingen. 154 subjects (82 female, 65 male) were recruited in busi-

ness studies (n = 89) or business education (n = 56)² lectures. The majority of the students were in the 4th and 5th semester of their bachelor's degree. The test was filled out in the presence of a test administrator. The students worked for approximately one hour (time was limited to max. 70 minutes) on the items and the additional questions concerning their personal background, their academic career, their work experience and their study interest with regard to business. The major research questions of this study were: (1) Can the items be organized into a one-dimensional construct? (2) To what extent does the test fulfill quality criteria (reliability and validity)?

After item evaluation by classical and probabilistic item analysis (Rost 2004), 24 items were selected for further analysis.

3.1 Dimensionality

The test was constructed under the assumption of one-dimensionality. This assumption was guided by previous empirical research in the field from Bothe (2003). Bothe (2003) suggested a one-dimensional structure for the 29 items of the BAKT. Supporting this view, Winther (2010) has also found procedural items to be organized one-dimensionally. However, there are indicators that would support the hypothesis that procedural business knowledge is organized in more than one dimension (e. g. Lehmann/Seeber, 2007). To test whether the one-dimensional model is justified it was tested against a four-dimensional model. The four dimensions were organized according to the subjects that guided test construction, namely management, production, marketing and accounting. The results of the model comparison are described in table 1.

Tab. 1: Comparative model fit between a one-dimensional and a four-dimensional model of procedural business knowledge in higher education

	Deviance (-2Log-Likelihood)	DF	Chi²-Test
1 dimension	4481.01853	25	
4 dimensions	4471.94089	36	
difference	9.0777	11	<i>p</i> = n.s.

The deviances indicate a slightly better fit for the four-dimensional model. A test for significance of this difference using a *Chi²-Test* identifies this difference as not statistically significant. Furthermore, intercorrelations between the four dimensions were high and the reliabilities of the subscales were low

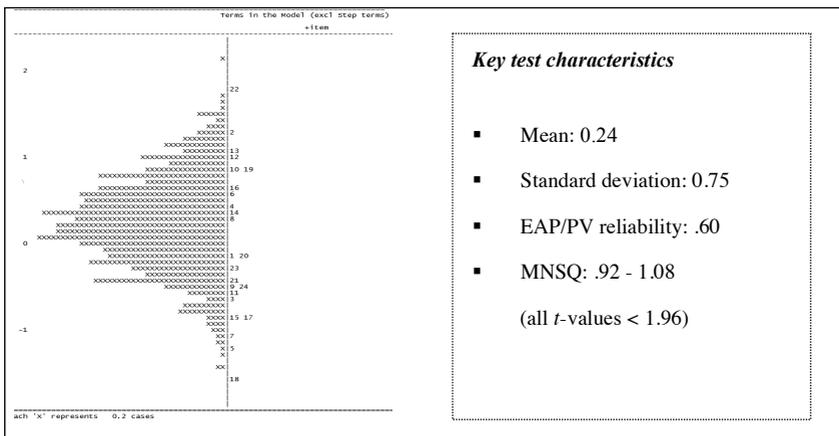
² Not all students in the sample stated their gender and their major.

(between .49-.58). Thus, subsequent analysis was conducted under the assumption of an underlying one-dimensional model.

3.2 Item difficulty and reliability

The selected items were analyzed under the one-dimensional Rasch Model using the item response modeling software ConQuest (Wu/Adams/Wilson/Haldane 2007). In a Rasch Model items and respondents can be projected along the same dimension and become directly comparable. ConQuest provides indices and visual displays such as the Wright-Map to help to establish whether item difficulties are spread sufficiently along the continuum. If items clump together in one spot of the difficulty/ability continuum it is necessary to develop items that cover the entire dimension of the construct, including extreme positions. The Wright-Map of this study is presented in figure 2.

Fig. 2: Wright-Map of the one-dimensional procedural business knowledge model with key test characteristics to the right of the map.



The Wright-Map displays a good distribution of items across the dimension. The upper and middle parts of the dimension could be better represented in the test. Each item shows a very good local item fit as can be derived from the MNSQ values. These are close to the ideal value of one and show no significant deviation from a perfect fit. The mean is slightly above zero, which indicates that the test has the tendency to be too easy rather than too hard. A critical result is the relatively low reliability of the test. EAP/PV reliability can be interpreted like Cronbach's alpha. The value of .60 does not conform to the requested benchmark of .70 in empirical research. Consequently, the reliability of the test has to be classified as questionable (George/Mallery

2002). Possible explanations for this and means of improving the reliability will be discussed in chapter 4.

3.3 Indicators of validity

There is no one simple indicator to measure the validity of a test. Because of this, multiple alternative paths are available for establishing validity claims. This paper follows the approach of Förster et al. (2012) by formulating and testing different hypotheses that are associated with different types of validity. The first two hypotheses serve as indicators for curricular validity, defined as the congruence between item content and curriculum (Hartig/Frey/Jude 2012). The number of attended test-relevant courses and the average grade were assessed via self-report. The following hypotheses were tested and results of hypothesis testing are displayed in table 2:

H1a: Person parameters³ are positively correlated with the number of test-relevant attended courses.

H1b: Person parameters are negatively correlated with the average grade in the test-relevant attended courses.

Tab. 2: Overview of results of hypothesis testing for indicators of curricular validity.

Hypothesis	Test	Result	Indication / Interpretation
H1a accepted	Pearson correlation	$r = .26; p < .01$ N = 133	Number of courses and grades in courses correlate with test results.
H1b accepted	Spearman correlation	$r_s = -.22; p < .05$ N = 104	Indication of adequate curricular validity.

Three more hypotheses were derived and tested to assemble indicators of construct validity. The reasoning behind hypothesis 2a is that students who have accomplished business related VET prior to starting their university degree profit from their prior knowledge in the domain of business. This effect has been shown for the beginning of their studies (compare Förster et al. 2012). Consequently, VET should have a positive effect on students test results at the beginning of their studies. And there should be no such (or a much weaker) positive effect of previous VET on test performance for students with more semesters of study.

³ Person parameters were estimated using ConQuest Software and maximum likelihood estimation.

H2a: Participants who have accomplished VET in the field of business have higher person parameters; this effect diminishes over time of study.

To estimate how the test results are influenced by domain-unspecific factors of academic achievement, the person parameters were correlated with the average grade of general education. To account for effects of time passed between finishing high school and taking the test, students with previous VET will not be considered in this analysis. Including them would pander to an underestimation of the correlation between high school grade and results because their completion of general education is usually longer ago and thus less predictive. For construct validity, only a moderate correlation should occur between these two measures. A high correlation would indicate that the construct was not operationalized in a domain-specific manner.

H2b: Person parameters are negatively correlated with the average grade of accomplished VET.

H2c: Person parameters of participants that have not accomplished VET are only moderately negatively correlated with the average grade of general education.

As fairness of tests has become a quality criterion of emerging importance (AERA/APA/NCME⁴ 2002), one more hypothesis was postulated to test for indicators of test fairness. Since previous empirical studies have reported gender effects in favor of male test takers (Förster et al. 2012), the hypothesis is postulated as follows:

H3: Person parameters of male participants are significantly higher than those of female participants.

It is to be noted that acceptance of hypothesis H3 indicates problematic test fairness. The results of the statistic tests are presented in table 3.

⁴ American Educational Research Association/American Psychological Association/National Council on Measurement in Education.

Tab. 3: Overview of results of hypothesis testing for indicators of construct validity and test fairness.

Hypothesis	Test	Result	Indication /Interpretation
H2a rejected	ANOVA ⁵	All effects n.s. N = 128	VET does not have a systematic effect on test results.
H2b accepted	Spearman correlation	$r_s = -.38; p < .01; N = 47$	Students with a good grade from VET have better results in tests and this correlation is slightly stronger for VET than for highschool grades.
H2c accepted	Spearman correlation	$r_s = -.29; p < .01; N = 97$	Mixed indication of construct validity.
H3 accepted	t-Test	$t_{(147)} = -2.75; p < .01; d = -.46$	Males significantly outperform females. Negative indicator for testfairness.

4. Discussion

4.1 Summary

At the beginning of this paper it was postulated that the development of assessment instruments in higher business education is necessary. The principles of test development for a test of procedural business knowledge were then outlined. Results of a pilot study were presented with a particular focus on deriving indicators of test validity. The results yielded some encouraging and some critical aspects for future test use. The results supported the claim that curricular validity was established. The results for construct validity were less clear. Although, completing VET before starting a university degree had an effect on person parameters at a descriptive level (students with previous VET achieved .17 logits more on the test than students without previous VET), the effect was far from statistically significant when testing for it in a two-factorial ANOVA. This might be caused by unequal and small group sizes. The stronger correlation of the average grade from VET with the test results compared to the average grade from high school (for participants without VET) is interpreted as a result in favor of construct validity. The test seems to capture more domain-specific knowledge than general knowledge. Significant gender differences should prompt a DIF analysis to explore item gender interactions as possible reasons for underperformance by females. As

⁵ A two-factorial ANOVA was conducted with completion of VET (yes/no) as one factor and semesters spent studying (<3/4/5/6/>7 semesters) as second factor. Gender and number of courses were controlled.

with most research in its pilot phase, the study has certain limitations. These limitations and their consequences for future research are discussed in the following section.

4.2 Limitations and future research

The results of the pilot study have shown that test construction has successfully achieved curricular validity. Thus far, this result only holds true for the sample of students from one university. Whether curricular validity can be claimed for a more heterogeneous sample of students from different universities, remains subject to future empirical research.

A limitation of the test is its questionable reliability. Low reliability is a common feature of situational tests (Schmitt/Chan 2006). It often reflects heterogeneous task requirements (Abele et al. 2012). Nonetheless, the aim is to improve reliability before the tests next deployment. Distractor analysis should help to identify guessing. Moreover, all identifiable sources of measurement error should be eliminated. Minimizing the effects of guessing is also expected to raise the difficulty of the test and have a positive effect on the distribution of item difficulty.

As pointed out in section 2.1, domain modeling in the field of higher business education has proven to be a difficult task. The heterogeneous course content between and within universities as well as the diverse employment opportunities for young business professionals are a serious threat to the systematic identification of domain-relevant requirements of students at university level. The pragmatic approach of developing items that are guided by overlapping core course content of universities and task-descriptions of job advertisements in alignment with textbook material weakens the theoretical basis of the test. Future research efforts should be made to develop a consistent model of the domain as well as a more elaborate model of item complexity. Those models could serve as a sound basis for more elaborate dimensional model testing. Potential conflicts between ecological validity (do the items approximate real-world problems?) (Süß 1999) and curricular validity (do the items conform to course content?) should be further investigated and discussed with experts from the university as well as professionals from companies that employ university business graduates.

In conclusion, the newly developed test shows weaknesses in its theoretical basis. Still, empirical analysis revealed promising first results. The future challenge is to reestablish claims of reliability and validity in a major study with a larger and more diverse sample. Expert judgement and revision of problematic items are considered as key factors in this process. Finally, the test will have to undergo a construct validation by proving to assess a diffe-

rent knowledge construct than the test for declarative business knowledge (BAKT).

References

- Abele, S./Greiff, S./Gschwendtner, T./Wüstenberg, S./Nickolaus, R./Nitzschke, A./Funke, J. (2012): Dynamische Problemlösekompetenz: Ein bedeutsamer Prädiktor von Problemlöseleistungen in technischen Anforderungskontexten? In: Zeitschrift für Erziehungswissenschaft, vol. 15, 2, S. 363–391.
- AERA/APA/NCME (2002): Standards for educational and psychological testing. Washington, DC.
- Anderson, J. R. (1983): The architecture of cognition. Cambridge, MA.
- Anderson, L. W./Krathwohl, D. R. (2001): A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York.
- Beck, K./Krumm, V./Dubs, R. (1998): Wirtschaftskundlicher Bildungs-Test (WBT). Göttingen.
- Black, H. T./Duhon, D. L. (2003): Evaluating and Improving Student Achievement in Business Programs: The Effective Use of Standardized Assessment Tests. Journal of Education for Business, 79(2), S. 90–98.
- Bothe, T. (2003): Entwicklung und Erprobung eines deklarativen Wirtschaftswissenschaftstests. Diplomarbeit, Universität Mannheim.
- Bothe, T./Wilhelm, O./Beck, K. (2005): Business administration knowledge: Assessment of declarative business administration knowledge: Measurement development and validation. Unpublished manuscript.
- Bologna Working Group on Qualifications Frameworks (2005): A framework for qualifications of the European higher education area. Copenhagen: Ministry of Science, Technology and Innovation.
- Förster, M./Happ, R./Zlatkin-Troitschanskaia, O. (2012): Valide Erfassung des volkswirtschaftlichen Wissens von Studierenden der Wirtschaftswissenschaften und der Wirtschaftspädagogik – eine Untersuchung der diagnostischen Eignung des Wirtschaftskundlichen Bildungstests (WBT). In *bwp@* Berufs- und Wirtschaftspädagogik – online, Ausgabe 22, S. 1–21. Online: <http://www.bwpat.de> (10-09-2012).
- George D./Mallery P. (2002): SPSS for Windows Step by Step: A Simple Guide and Reference. Boston.
- Gruber, H. (1999): Erfahrung als Grundlage kompetenten Handelns. Bern: Huber.
- Gruber, H./Mandl, H./Renkl, A. (2000): Was lernen wir in Schule und Hochschule: Träges Wissen? In: H. Mandl/J. Gerstenmaier (Eds.), Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze Göttingen: Hogrefe, S. 139–152.
- Hartig, J./Frey, A./Jude, N. (2012): Validität. In H. Moosbrugger/A. Kelava (Eds.), Testtheorie und Fragebogenkonstruktion. Berlin: Springer, S. 143–171.

- Henning, C./Henning, W. (2009): Studienführer Wirtschaftswissenschaften: Betriebswirtschaft, Volkswirtschaft, internationale Betriebswirtschaft, internationale Volkswirtschaft, Wirtschaftswissenschaften/Ökonomie, Wirtschaftsinformatik, Wirtschaftspädagogik. Eibelstadt: Lexika.
- Hofmeister, W. (2009): Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell. In: *bwp@ Berufs- und Wirtschaftspädagogik – online*, Ausgabe 8, S. 1–21. Online: <http://www.bwpat.de> (10-09-2012).
- Horváth, P./Gleich, R./Voggenreiter, D. (2007): Controlling umsetzen; Fallstudien, Lösungen und Basiswissen. Stuttgart.
- Klieme, E. (2004): Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, 6, S. 10–13.
- Lehmann, R./Seeber, S. (2007) (Hrsg.): ULME III. Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen. Hamburg: Behörde für Bildung und Sport.
- Motowidlo, S. J/Dunnette, M. D/Carter, G. W. (1990): An alternative selection procedure: The low fidelity simulation. Renkl, A. (1996): Träges Wissen: Wenn Erlerntes nicht genutzt wird.
- Reiss, K. (2012): Competence development a key concept of higher education. In: *Empirical Research in Vocational Education and Training*, vol. 4(1), pp. 73–76.
- Rost, J. (2004): Lehrbuch Testtheorie – Testkonstruktion. Bern: Huber.
- Rost, J. (2008): Zur Messung von Kompetenzen einer Bildung für nachhaltige Entwicklung. In I. Bormann/de Haan Gerhard (Eds.), *Kompetenzen der Bildung für nachhaltige Entwicklung. Operationalisierung, Messung, Rahmenbedingungen, Befunde*. Wiesbaden: VS, S. 61–73.
- Schmitt, N./Chan, D. (2006): Situational Judgment Tests: Method or Construct? In J. A. Weekley/R. E. Ployhart (Eds.), *Situational judgment tests. Theory, measurement, and application*. Mahwah, N.J: Lawrence Erlbaum, pp. 135–155.
- Soper, J. C. /Walstad, B. (1987): *The Test of Economic Literacy. Examiner's Manual*, Joint Council on Economic Education.
- Süß, H.-M. (1999): Intelligenz und komplexes Problemlösen. *Psychologische Rundschau*, 50(4), S. 220–228.
- Teichler, U. (2006): Was ist Qualität? Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung. In: *In Hochschulkonferenz*. Bonn: HRK, S. 168–184.
- OECD (2011): *Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Economics*. Retrieved from <http://dx.doi.org/10.1787/5kghtchw3nn-en> (15.03.2013).
- Weinert, F. E. (2001): *Leistungsmessung in Schulen*. Weinheim: Beltz.
- Winther, E. (2010): *Kompetenzmessung in der beruflichen Bildung*. Bielefeld.
- Wu, M. L./Adams, R. J./Wilson, M. R./Haldane, S.A. (2007): *ACER ConQuest Version 2.0. Generalised item response modelling software*. Camberwell Vic.
- Zlatkin-Troitschanskaia, O./Kuhn, C. (2010): *Messung akademisch vermittelter Fertigkeiten von Studierenden bzw. Hochschulabsolventen – Analyse zum Forschungsstand*. Johannes Gutenberg-Universität Mainz: Arbeitspapiere Wirtschaftspädagogik, Nr. 56.