2.6 Linguistically Responsive Teaching in Multilingual Classrooms

Development of a Performance-Oriented Test to Assess Teachers' Competence

Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Koch-Priewe, B., Köker, A., and Ohm, U.

Abstract

The *DaZKom-Video* project aims at performance-oriented measurement of German-as-a second-language (GSL)-competence of pre- and in-service teachers using video-based stimuli and oral responses. The present study is part of this project and focuses on the psychometric quality of the *DaZKom-Video* test instrument and evaluates the dimensionality of the construct *GSL competence (Deutsch-als-Zweitsprache-Kompetenz).* Additionally, we explore correlations between GSL competence and sociodemographic data as well as learning opportunities. The sample consists of pre- and in-service teachers from different school forms and universities across Germany. The items and scale analyses reveal good psychometric quality: The item-fit to the Rasch-model as well as the discrimination is satisfactory. The dimensional analyses show best model fit for the one-dimensional and three-dimensional models. Test persons with GSL experience such as in-service teachers (in contrast to pre-service teachers), test persons with teaching experience, those with an additional qualification in the field of GSL and test persons with many learning opportunities are statistically more likely to respond to items correctly. These results indicate a reliable and valid test instrument for performance-oriented measurement of GSL competence.

O. Zlatkin-Troitschanskaia et al. (Hrsg.), *Student Learning in German Higher Education* https://doi.org/10.1007/978-3-658-27886-1_7

Keywords

Video-based testing, teacher education, linguistically responsive teaching, teachers' competence

1 Introduction

Language is crucial in teaching and learning: Every subject-specific classroom uses language to represent content. Language has specific characteristics in content classrooms and every subject adds its own subject-specific register (e.g. written language or specific operators, Schmölzer-Eibinger 2013). For some pupils, especially but not only for those with German as a second language (GSL), it therefore can be challenging to apprehend content successfully. Teachers of all subjects must be aware that specific registers come with their own demands and may pose challenges for their pupils. Accordingly, teachers need competencies to support their pupils in gaining access to the content by teaching linguistically responsive (Thürmann and Vollmer 2013, Cummins 2001). Kimanen et al. (2019) implicate that teachers' competence in linguistically and culturally responsive teaching could increase teachers' self-esteem and therefore have an impact on teachers' motivation and commitment to teaching. How teachers must be trained concerning teaching multilingual learners is still unclear, however, and performance standards of teachers' competencies in this field are missing. Furthermore, understanding which learning opportunities may contribute to the development of teachers' competencies in teaching in linguistically diverse classrooms is still a desideratum (Baumann 2017).

The *DaZKom-Video*[1] project aims at evaluating and improving learning opportunities for pre- and in-service teachers in teaching multilingual learners and developed a performance-oriented test instrument. The theoretical and assessment framework includes the GSL competence model and the paper-pencil-test of the previous project (*DaZKom*[2]). To measure teachers' professional competencies and not solely capture teachers' knowledge, but operate as closely to performance as possible, it is necessary to use performance-oriented test formats (Aufschnaiter

1 The project *DaZKom-Video*: *Performanznahe Messung von Deutsch-als-Zweitsprache-Kompetenz bei (angehenden) Lehrkräften (2017–2020).*

2 The previous project *DaZKom* (2012–2015) was funded by the German Federal Ministry for Education and Research (BMBF) (01PK11010A, 01PK11010B).

and Blömeke 2010). Consequently, the test instrument uses video vignettes of teaching situations and asks for orally given responses that are then recorded.

This article examines the extent to which the developed performance-oriented test instrument is suitable for capturing the construct of GSL competence. It evaluates different scaling models concerning the dimensionality of the GSL construct. As indicators for convergent and discriminant validity, the correlations of the measured GSL competence and socio-demographic characteristics as well as GSL-related learning opportunities of test persons are examined in this study.

Concerning the terminology, this paper uses 'teaching pupils with German as a second language (GSL)' synonymously to 'teaching multilingual learners' and 'teaching in linguistically diverse classrooms'. GSL competence therefore refers to the competency to teach multilingual learners/in linguistically diverse classrooms.

2 Theoretical Background

2.1 Structure of GSL Competence

A structural model of teachers´ GSL-competence served as the theoretical framework and basis for the performance-oriented test development. This model was devised in a previous project on the basis of document analysis, followed by an expert rating (Ehmke and Hammer 2018; Köker et al. 2015). Due to the premise that GSL competency is relevant for teaching all subjects in the mainstream classroom, the model consists of different partial competencies which are important for subject content integrated language facilitation and dealing adequately with multilingualism. The model illustrates the question of how subject content can be taught in such a way that the linguistic needs of all pupils, especially those with German as their second language, are taken into account (Carlson et al. 2018). In addition, the model depicts what teachers need to know to be able to act linguistically responsive. This knowledge refers to linguistic aspects, for instance, with regard to the specificity of specialized texts, aspects of second language acquisition and multilingualism as well as the didactic aspects of lesson planning and implementation. Since GSL competence is considered a generic competence in all subjects, it always refers to subject content.

The structural model discerns three dimensions: *subject-specific register*, *multilingualism,* and *didactics* (Ohm 2018). All dimensions are differentiated into subdimensions and further characterized by content-related facets.

The first dimension, *subject-specific register*, concerns language as a medium for classroom interaction and understanding technical concepts as well as a

learning object. It is assumed that all subjects have their specific registers that are essential for the students' knowledge construction and must therefore be taught explicitly in connection with subject contents (Schleppegrell 2004).

The second dimension, *Multilingualism,* includes the two subdimensions S*econd Language Acquisition* and *Migration* and focuses on the learning process (Ohm 2018). To support the (GSL) students according to their individual language learning levels, teachers require knowledge about the development of students' linguistic competence, for instance, regarding milestones of second language acquisition. Second language learning resp. the increasingly differentiated use of linguistic registers occur in the context of multilingualism. At the same time, multilingualism refers not only to external but also to internal multilingualism, such as the use of dialects. Therefore, the second dimension considers both linguistic diversity in school and dealing with heterogeneity.

The third dimension, *Didactics*, includes the subdimensions *formative assessment* in classrooms and *language facilitation* (Ohm 2018) and complements the first two dimensions. Based on knowledge about subject-specific registers and the specific characteristics of second language development, teachers plan and conduct their lessons. Consequently, the teaching process is at the focus of this dimension.

2.2 Indicators of Performance: Perception and Decision-Making

The situation-specific skills, *perception* and *decision-making* (Blömeke et al. 2015), serve as indicators of performance in GSL-relevant teaching situations. In research on teacher professionalization, perception often is defined by referring to the concept of *professional vision* by Goodwin (1994). He explains experts perceiving meaningful situations in their domain of expertise more selectively than novices do, as experts´ perception has been trained to that effect in years of practice. Van Es and Sherin (2002) define two components of teachers´ *professional vision: noticing* and *knowledge-based reasoning*. While the first term describes the filtered selective perception, the second refers to the ability to draw conclusions based on the situations perceived (Es and Sherin 2002). Therefore, experts are able to resort to their professional knowledge to classify the perceived into concepts and theories and react more precisely. In contrast, novices often deliver only a mere description of what they observed (Seidel et al. 2010).

Professional competence consists of professional knowledge and individual abilities, as well as motivational, volitional and social aspects and is dependent on domain-specific requests of action (Weinert 2001; Baumert and Kunter 2006).

Lindmeier (2013) captures the ability of *decision making* as *action-oriented* competency and thereby refers to the ability of teachers to use their professional knowledge spontaneously and immediately in situations (ibid.; see also Kuhn et al. in this volume).

The *DaZKom-Video* test instrument uses test items which pick up this theoretical structure. Video-stimuli present authentic teaching situations with specific *perception* and *decision-making* tasks to measure the German-as-a-Second-Language competence of the (pre-service) teachers as closely to performance as possible. The video-vignettes were carefully chosen after reviewing many teaching situations (video and audio recordings) and were validated for their authenticity by experts of universities and teacher experts in schools.

3 Research Questions

Based on the theoretical framework, three research questions will be answered in this paper:

1. How satisfactory is the *DaZKom-Video* test instrument´s psychometrical quality?

Based on the evaluation of data of a first pilot study (Lemmrich et al. 2019), it can be assumed that the data will have a good item fit to the Rasch-model. Additionally, the discrimination in the first pilot study was satisfactory, even though the items were too difficult for the respective sample then.

2. Which dimensional structure can be determined for the test instrument?

Considering the results of the paper-pencil test of the previous project (*DaZKom*) it can be assumed that the construct *GSL competence* is structured multidimensionally (Hammer et al. 2015), also in a performance-oriented test. However, taking the results of the first pilot study into account (Lemmrich et al. 2019), it is possible that the theoretically determined dimensions do not appear clearly separated. As the two-dimensional analysis (*perception (1)* and *decision-making (2)*) showed strong correlations between these two dimensions in the first pilot study (ibid.) we expect a similar result in the second pilot study. Referring to GSL competence, three dimensions can be determined based on the theoretical assumptions (*subject-specific register (1), multilingualism (2), didactics (3)*, see *DaZKom-Model* in Köker et al. 2015, Ohm 2018; Section 2), which either appear as one-dimensional

or can be determined separately. In the first pilot study, results showed a medium correlation between dimension 1 and 3 and a low correlation between both of these and dimension 2. As the present pilot study was carried out with a revised test instrument and a different sample including more in-service teachers than the first pilot study, the results are expected to clarify the dimensionality of the construct.

3. Which correlations can be identified between the GSL competence of the test persons, the sociodemographic characteristics and learning opportunities?

The correlations examined give insight into the convergent and discriminant validity. We assume that test persons who had GSL-relevant learning opportunities are more likely to respond correctly. Additionally, experienced teachers are expected to pass the performance-oriented test more successfully than unexperienced teachers or pre-service teachers.

4 Methods

4.1 Test Instrument

The project team collected data in a standardized way at four universities and six schools in four German federal states (Berlin, Lower Saxony, North-Rhine Westphalia, Hamburg). The conductors first introduced the test closely following a manual, covering both the test procedure as well as the test´s background. In their introductions, the conductors invariably pointed out the items´ GSL focus. Afterwards, the test persons took two of four sub tests. Each subtest contained four video stimuli. Headsets and tablets were provided to enable the participants to take the performance-oriented test independently. After watching an example item including a sample solution, eight video-vignettes were shown, each followed by two items on the situation-specific skills *perception* and *decision-making*. The video-vignettes used as prompts take 30 seconds to three minutes, each dealing with authentic, GSL-relevant teaching situations that were taken from different subject lessons at secondary school classes. Subjects are Mathematics, German, Ethics and Science. All situations were matched both to the dimensions and subdimensions of the DaZKom-Model (Köker et al. 2015). Their adequacy, GSL relevance and typicality for the chosen dimension were assessed and approved in expert ratings with $N = 3$ university experts in the field.

To best simulate the immediacy and spontaneity of teacher behavior in real situations and thereby achieve a higher ecological validity, only open-response

items on the test persons´ perception (*What do you perceive?*) and on their simulated (re-)action (a. *You are the teacher in this situation, how do you react word for word?*, or b. *How would you act in this situation if you were the teacher?)* were used, each asking for an oral response using the headsets´ built-in microphones. To enable these kinds of open responses, the video vignettes usually closed with a student´s statement the test persons could react to. At last, the test persons took an additional questionnaire both on their learning opportunities and sociodemographic data (Section 4.3).

4.2 Coding Manual

The coding manual, required to measure difference in quality in the test persons´ responses reliably, was developed in a complex multi-step process. It included expert ratings ($N = 6$) with $n = 3$ researchers from the field (PhD and above) and $n = 3$ experienced teachers (different subjects, 30+ years of experience in schools while also working as instructors offering further training in GSL). The experts were taken through the same procedure as the regular test persons: they first watched the video vignettes before giving oral responses to the items on their situation-specific skills, independently and without a project member being present. Their responses were audio-taped and afterwards used for a qualitative content analysis employed to identify codes (Nimz, Hecker & Köker 2018). These codes were then ascribed to the scores 0, 1 and 2 in the project team. Based on this allocation, the actual item specific coding manuals were designed. Several rounds of optimization followed. The manuals were tested using data from the first pilot study conducted in Germany ($N = 137$; $n = 40$ in-service teachers, $n = 83$ pre-service teachers, $n = 11$ scientists). All responses were double coded by two independent raters. During the rating, doubtful cases were identified and discussed in the project team, more anchors added and necessary adjustments to the manuals made. Finally, the team achieved a satisfactory interrater agreement (Cohen´s Kappa $\varkappa = .76$). The final manuals include closed descriptions of the values in addition to many anchors. In the process of scoring the responses, only the best aspects in every response count, the rest is not taken into account. A maximum of two points can be reached in each item. For the analysis performed in this study, code 2 and code 1 were merged into code 1.

4.3 Additional Questionnaires: Sociodemographic Data and Learning Opportunities

After giving their oral responses, the test persons filled in additional questionnaires on sociodemographic characteristics and learning opportunities. The first questionnaire presented questions on gender, subjects of studies, teaching experience, etc. (6 items). The learning opportunities questionnaire contains two scales: One scale of 16 items on GSL-relevant topics that the test persons might have discussed in their studies/teacher training ($\alpha = 0.91$; Ehmke and Lemmrich 2018). Furthermore, it contains a scale of eight items ($\alpha = 0.83$, ibd.) on GSL-relevant actions the test persons might have taken before in their learning contexts (e.g. studies, further teacher training, etc.). Both scales use a five-point Likert scale ((1) *never* (2) *in one session*, (3) *in in several sessions*, (4) *in a module*, (5) *in several modules*. Additionally, there were three single items that aimed for answers on GSL-specific teaching experience, additional qualifications and experience in research in the field of teaching multilingual learners.

4.4 Sample

The sample size of this study was $N = 184$ test persons, 79 % female and 21 % male; 51.2 % were pre-service teachers (teacher students) from German universities and 39.6 % were in-service teachers from different areas of Germany and different types of schools; 8.2 % were "other", such as scientists, teacher educators and students of other disciplines. The test persons studied/taught different kinds of subjects (Math, German, English, Physics, Biology, Chemistry, Geography, Social Studies/Science, History, Music, Arts, Religion, Politics, Sports); 25.6 % had additional qualifications in the field of GSL/teaching multilingual learners (e.g. GSL-certificate); 88.6 % had German as their first language, while 11.4 % of the test persons had a different first language, which mostly were Turkish (3.8 %) and Russian (1.6 %).

4.5 Statistical Procedure

The collected data (coding of the *DaZKom-Video* test, sociodemographic data and learning opportunities) were analyzed with SPSS 25 (IBM 2018), including frequencies and information on the sample. The psychometric analysis of items and scales was carried out in *ConQuest* (Adams et al. 2015) on the basis of the

Rasch-model (Rost 2004). The GSL competence of the test persons was determined with the Weighted Likelihood Estimates (WLEs). WLEs take two factors into account: Firstly, the correct reply to an item depending on the respondents' abilities (in this case GSL competence); secondly, the item difficulty (Wilson 2005). Hence, the WLE does not map the actual responses of the test persons, but the probability of a correct answer based on the test persons´ skills and the item difficulty (ibid.). To examine the dimensionality of the construct GSL-competence, a one, a two and a three-dimensional model were estimated with *ConQuest*.

5 Results

5.1 Psychometric Quality

Table 1 shows the results of the IRT-Scaling. The first criterion for psychometric quality is the item fit based on the Rasch-model, which is very good ($1.0 < MNSQ^3 < 1.25$; OECD 2005). A few items show a lower item fit ($Min = 0.88$; $Max = 1.10$), but at least the maximum item fit is located within the satisfactory range. The second criterion is discrimination, which indicates how reliable the instrument distinguishes between test persons with high and low ability estimates. Three items that showed a discrimination of under 0.3 were excluded from all analysis. Thereafter, the average discrimination of all items was satisfactory ($M = 0.43$; $SD = 0.09$). The overlap of the average item difficulty ($M = 0.88$; $SD = 0.78$) and the respondents' abilities ($M = 0.00$; $SD = 1.08$) is not fully satisfactory yet. Therefore, the items either are too difficult or the respondents' abilities not sufficient enough. The EAP-reliability of the one-dimensional scaling amounts $\alpha = 0.62$.

3 *Weighted Meansquare Fit Statistics (*MNSQ): indicates if the expected and the observed likelihood to respond correctly relate.

Table 1 Item and scale indices

	M	SD	Min	Max
Weighted fit MNSQ	1.00	0.05	0.88	1.10
Discrimination	0.43	0.09	0.30	0.58
Item difficulty	0.88	0.78	-0.53	2.79
WLE	0.00	1.08	-2.87	3.28

5.2 Verification of Dimensions

The dimensionality of the test instrument was tested with one-, two- and three-dimensional IRT models. Table 2 shows the fit-indices of the comparison of the three models: the deviance, the AIC, BIC and the CAIC. All four of them are criteria that give information about the goodness of fit of the estimated model. They take into account the empirical data as well as the model (Moosbrugger and Kelava 2007, p. 390). All these criteria should be as low as possible (ibid.). The BIC and the CAIC are at the lowest level for the one-dimensional model. The deviance and the AIC are at lowest for the three-dimensional model. The construct *GSL competence* therefore is likely to be a one-dimensional or three-dimensional construct rather than a two-dimensional construct. The reliability of the three-dimensional model was not satisfactory due to a very low number of items per dimension. The following analyses therefore are operated with data of the one-dimensional model.

Table 2 Comparison of dimensions

	N	Parameter	Deviance	AIC	BIC	CAIC
1-dimensional	176	30	2606	2666	2761	2791
2-dimensional	176	32	2607	2671	2773	2805
3-dimensional	176	35	2585	2655	2766	2801

Looking at the two-dimensionally modeled construct, the two assumed dimensions *perception* and *decision-making* (Section 2.2) highly correlate ($r = 0.97$, Figure 1). It therefore can be assumed that the construct *GSL competence* is not separable into these two dimensions. The three-dimensional modeled construct (Section 2.1) shows moderate correlations: dimension 1 (*subject specific register*) and dimension 2 (*multilingualism*) correlate by $r = 0.46$, dimension 2 and 3 (*didactics*) by $r = 0.51$ and dimension 1 and 3 by $r = 0.55$. This result indicates that these three theoretical dimensions could be separated within the construct *GSL competence*.

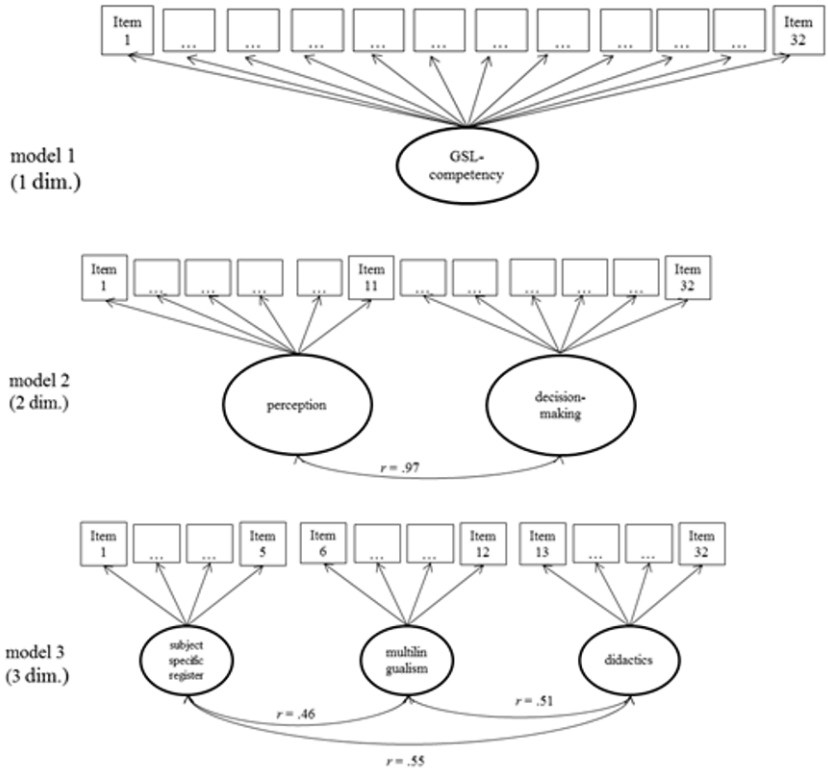

Figure 1 Comparison of models

5.3 Correlations between GSL Competence, Personal Characteristics and Learning Opportunities

Table 3 shows correlations between sociodemographic characteristics and GSL-relevant learning opportunities with the weighted likelihood estimates (WLE). There is no statistically significant difference between the female and male test persons probability to response correctly. The results show a tendency that test persons with German as their first language are more likely to respond correctly, even though the effect is statistically non-significant. Teachers' responses are more often correct than those of students' (pre-service teachers). This effect is confirmed by the statistically significant correlation of teaching experience with the

WLE. Also, test persons who have a GSL-certificate, which is a specific qualification for teaching multilingual learners, are more likely to respond correctly than those without. Lastly, learning opportunities (sum score of both scales) correlate with the WLE: The more GSL-relevant learning opportunities test persons had, the more likely they were to respond correctly.

Table 3 Correlations between the results of the DaZKom-Video-test and sociodemographic data

	WLE
Gender (male = 0; female = 1)	0.14
Status (teacher student = 0; teacher = 1)	0.04
GSL – certificate[1]) (no = 0, yes = 1)	0.20**
Teaching experience GSL[2]) (no = 0; yes = 1)	0.16*
Research GSL[3]) (no = 0; yes = 1)	0.09
Scale GSL-learning opportunities[4])	0.18*

* The correlations are statistically significant at a level of .05 (two-sided).
** The correlations are statistically significant at a level of .01 (two-sided).
1) additional certificate that proves the attendance of a specific course on teaching students with German as a second language
2) experience in teaching students with German as a second language
3) research experience in the field of German a s a second language/teaching multilingual learners
4) learning opportunities in the field of German a s a second language/teaching multilingual learners

6 Discussion and Outlook

This study has been carried out to evaluate the psychometric quality and the dimensionality of the performance-oriented test instrument. Additionally, correlations of GSL competence with sociodemographic data and learning opportunities were examined. The main results were the following:

1. The test instrument shows a good fit to the Rasch-model. Three items were excluded from the test due to a low discrimination value (< 0.3). The discrimination was satisfactory. At the same time, the test did not fit perfectly to the ability of the test persons, which indicates that the items were too difficult.
2. The dimensional analyses showed the best model fits for the one-dimensional and the three-dimensional models. The three-dimensional structure represents

the theoretical structure of the GSL assessment framework. The two dimensions of perception and decision making could not be separated statistically from each other.

3. The correlations between the WLE of the test persons concerning GSL competence and the sociodemographic data/learning opportunities indicate a higher likelihood of responding correctly for test persons who are experienced teachers. Teaching experience as well as more learning opportunities correlated statistically significant with the WLE.

The psychometric analysis showed that some items were still too difficult for the test persons. However, compared to the first pilot study (Lemmrich et al. 2019), the item fit increased. After the first evaluation, the test instrument was revised: Stimuli that did not work were excluded and more context information was added to increase the specificity of the situations. Additionally, the current sample consists of half students (pre-service teachers) and half teachers, whereas the first pilot study was carried out with a majority of students, who are less experienced than teachers and therefore not as likely to perform successfully in a performance-oriented test instrument (Section 2).

Concerning the dimensionality of the construct, there is no evidence for a two-dimensional construct. *Perception* and *decision-making* cannot be determined separately for the construct. This result is in line with results of a first pilot study (Lemmrich et al. 2019) and is compliant with theories of cognition psychology as well as other studies on professional vision and performance (Niesen 2018; Hommel et al. 2019). Another possible explanation for the results is the fact that the items corresponding to either *perception* or *decision-making* both demand answers that are based on one and the same video vignette. The results indicate a one-dimensional construct in a performance-oriented environment. However, in the three-dimensional modeling (three dimensions of GSL competence: *subject specific register, multilingualism* and *didactics;* Section 2), the three dimensions show only medium correlations, which indicates that a differentiated analysis might be reasonable. Further analysis and evaluation will examine the dimensionality in detail with the final test instrument and a larger sample with a higher number of teachers. The standardization study will examine a sample of $N = 300$ ($N_{teachers} = 150$) with a revised test instrument (12 final video vignettes in one test version).

Looking at the correlations of GSL-competence with sociodemographic data and learning opportunities, the results comply with the theoretical background. Experience and learning opportunities correlate with the WLEs of the test persons: test persons with teaching experience as well as with additional qualifica-

tions in the field of GSL or learning opportunities are more likely to respond correctly to the items.

The instrument contains 16 vignettes that show very short teaching examples. It therefore offers only a selection of possible contexts and cannot portray the broad variety and complexity of everyday teaching situations. This study aimed at evaluating the psychometric quality and dimensionality of the test instrument and did not analyze descriptive values. Future studies (with an equally high psychometric quality) will take this into account and identify GSL-relevant learning opportunities to then comprehensively train teachers in this field and prepare them optimally to teach in linguistically diverse classrooms.

Funding

DaZKom-Video was funded by the KoKoHs-program of the German Federal Ministry for Education and Research (01PK16001A, 01PK16001B).

References

Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). ACER ConQuest: Generalised Item Response Modelling Software [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.

Aufschnaiter, C. v., & Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften, 16,* pp. 361–367.

Baumann, B. (2017). Sprachförderung und Deutsch als Zweitsprache in der Lehrbildung. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Eds.), *Deutsch als Zweitsprache in der Lehrerbildung* (pp. 9–26). Münster: Waxmann.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9 (4)*, (pp. 469–520).

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

Carlson, S. A., Köker, A., Rosenbrock-Agyei, S., Ohm, U., Koch-Priewe, B., Hammer, S., Fischer, N., & Ehmke, T. (2018). DaZKom – a Structure Model of Pre-service Teachers' Competency for Teaching German as a Second Language in the Mainstream Classroom. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 261–283). Münster and New York: Waxmann.

Cummins, J. (2001). *Negotiating Identities: Education for Empowerment in a Diverse Society*. Second Edition. Los Angeles: California Association for Bilingual Education.

Ehmke, T., & Hammer, S. (2018). Skalierung und dimensionale Struktur des DaZ-Kom-Testinstruments. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 129–148). Münster: Waxmann.

Ehmke, T., & Lemmrich, S. (2018). Bedeutung von Lerngelegenheiten für den Erwerb von DaZ-Kompetenz. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 201–219). Münster: Waxmann

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96 (3),* pp. 606–633.

Hammer S., Carlson, S. A., Ehmke, T., Koch-Priewe, B., Köker, A., Ohm, U. Rosenbrock, S., & Schulze, N. (2015). Kompetenz von Lehramtsstudierenden in Deutsch als Zweitsprache. Validierung des GSL-Testinstruments. *Zeitschrift für Pädagogik*, Beiheft 61, pp. 32–54.

Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W. (2001).The theory of event coding (TEC): A framework for perception and action planning. In: Behavioral and Brain Sciences 24, H. 5, S. 849–878.

Kimanen, A., Alisaari, J., & Kallioniemi, A. (2019). In-service and pre-service teachers' orientations to linguistic, cultural and worldview diversity. In: J*ournal of Teacher Education and Educators*, 8 (1), 35–54.

Köker, A., Rosenbrock-Agyei, S., Ohm, U., Carlson, S. A., Ehmke, T., Hammer, S., Koch-Priewe, B., & Schulze, N. (2015). DaZKom – Ein Modell von Lehrerkompetenz im Bereich Deutsch als Zweitsprache. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttke (Eds.): *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie früh-pädagogischer Fachkräfte* (pp. 177–206). Bad Heilbrunn: Klinkhardt.

Lemmrich, S., Bahls, A., & Ehmke, T. (under review). Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften.

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 T. Ehmke, P. Kuhl & M. Pietsch (Eds.), *Lehrer. Bildung. Gestalten. Beiträge zur empirischen Forschung in der Lehrerbildung* (pp. 188–202). Weinheim und Basel: Beltz Juventa.

Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Eds.), *Videobasierte Kompetenzforschung in den Fachdidaktiken. Fachdidaktische Forschungen, Vol. 4,* pp. 45–62. Münster: Waxmann.

Moosbrugger, H., & Kelava, A. (2007). Testtheorie und Fragebogenkonstruktion. Heidelberg: Springer Medizin Verlag.

Nimz, K., Hecker, S. – L., & Köker, A. (2018). Videobasierte Messung von DaZ-Kompetenz bei Lehrkräften. In C. Caruso, J. Hofmann, A. Rohde & K. Schick (Eds.), Sprache im Unterricht. Ansätze, Konzept, Methoden (pp. 439–452). Trier: WVT.

Niesen, H. (2018). Förderung mehrsprachigkeitssensibler professioneller Handlungskompetenz angehender Englischlehrkräfte. In: Zeitschrift für Interkulturellen Fremdsprachenunterricht: Didaktik und Methodik im Bereich Deutsch als Fremdsprache 23, H. 1, S. 121–134.

OECD (2005): PISA 2003. Technical report. Paris: OECD.

Ohm, U. (2018). Das Modell von DaZ-Kompetenz bei angehenden Lehrkräften. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 73–91). Münster: Waxmann.

Rost, J. (2004). Testtheorie Testkonstruktion (second edition). Bern: Huber.

Schmölzer-Eibinger, S. (2013). Sprache als Medium des Lernens im Fach. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Eds.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 25–40). Münster: Waxmann.

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer" – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In M. Bayrhuber, T. Leuders, R. Bruder & M. Wirtz (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (pp. 296–306). Weinheim und Basel: Beltz.

Thürmann, E., & Vollmer, H. J. (2013). Schulsprache und Sprachsensibler Fachunterricht: Eine Checkliste mit Erläuterungen. In C. Röhner & B. Hövelbrinks (Eds.), *Fachbezogene Sprachförderung in Deutsch als Zweitsprache. Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen* (pp. 121–233). Weinheim: Beltz Juventa.

van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education, 10 (4)*, pp. 571–596.

Schleppegrell, M. J. (2004). The language of school: A functional linguistics perspective. Mahwah: Lawrence Erlbaum Associates.

Weinert, F. E. (2001). Concept of Competence: A Conceptual Clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45–66). Göttingen: Hogrefe.

Wilson, M. (2005). Constructing Measures. An Item Response Modeling Approach. New York and London: Psychology Press, Taylor & Francis Group.