peDOCS
Open Access Erziehungswissenschaften

DIPF

Smit, Robbert

Individual differences in beginning teachers' competencies - a latent growth curve model based on video data

Journal for educational research online 6 (2014) 2, S. 21-43

Quellenangabe/ Reference:
Smit, Robbert: Individual differences in beginning teachers' competencies - a latent growth curve model based on video data - In: Journal for educational research online 6 (2014) 2, S. 21-43 - URN: urn:nbn:de:0111-pedocs-96759 - DOI: 10.25656/01:9675

https://nbn-resolving.org/urn:nbn:de:0111-pedocs-96759
https://doi.org/10.25656/01:9675

in Kooperation mit / in cooperation with:

WAXMANN
www.waxmann.com

http://www.waxmann.com

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

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

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 6 (2014), No. 2, 21–43
© 2014 Waxmann

Robbert Smit

Individual differences in beginning teachers' competencies – A latent growth curve model based on video data

Abstract

As part of the cross-national project "Alpha", the classroom instruction of 73 beginning teachers from four teacher universities in Austria, Germany and Switzerland was recorded on video at two different points during their first year of teaching. The observed teaching techniques of the teachers were then rated by experts. A reanalysis of the qualitative rating data led to a reduced competence model. Derived from a confirmatory factor analysis, a model with three major teaching competencies – "motivating students", "pacing" and "facilitating" – was established. Based on a latent class growth curve model, we found individual differences at the first point of measurement, but no differences in growth. Adding covariates to the model revealed differences in growth. Beginning teachers with previous study experience showed significant growth in the three advanced teaching competencies compared to teachers without such experience.

Keywords

Teacher education; Beginning teachers; Teacher competencies

Individuelle Unterschiede in den Kompetenzen von Lehrpersonen im Berufseinstieg – Ein *Latent Growth Curve Model* auf der Basis von Videodaten

Zusammenfassung

Im Rahmen des länderübergreifenden Projekts „Alpha" wurde der Unterricht von 73 Studienabgänger/inne/n vierer pädagogischer Hochschulen aus Deutschland, Österreich und der Schweiz zu zwei Zeitpunkten des ersten Berufsjahres auf Video aufgenommen. Die Unterrichtsgestaltung wurde von Experten mittels eines Ratings evaluiert. Eine Re-Analyse der qualitativen Ratingdaten führte zu einem reduzierten Messmodell der Unterrichtskompetenz. Abgeleitet

Dr. Robbert Smit, Centre for Research & Development, University of Teacher Education St. Gallen, Müller-Friedberg-Strasse 34, 9400 Rorschach, Switzerland
e-mail: robbert.smit@phsg.ch

von einer konfirmatorischen Faktorenanalyse ergab sich ein Modell mit drei Unterrichtskompetenzen: „Motivierung der Lernenden“, „Lernsteuerung“ und „Lernbegleiter“. Mittels eines Latent Class Growth Curve Model wurden individuelle Kompetenzentwicklungen im ersten Berufsjahr geprüft. Es zeigten sich zwar individuelle Unterschiede zu Beginn, aber keine signifikanten Zuwächse an Kompetenz. Der Einbezug von Kovariaten erwies sich für die Entwicklung einer Gruppe von Berufseinsteigenden als bedeutsam. Lehrpersonen, welche Erfahrungen aus einem anderen Studium mitbrachten, zeigten im Gegensatz zu denjenigen ohne eine solche Erfahrung einen signifikanten Kompetenzzuwachs in den drei anspruchsvollen Kompetenzen.

Schlagworte
Lehrerausbildung; Berufseinstieg; Unterrichtskompetenz

1. Introduction

The transition from “student of teaching” to “teacher of students” has been an important topic in teacher research for many years (Feiman-Nemser, Schwille, Carver, & Yusko, 1999; Grossman, 1992; Veenman, 1984). It is during this transitional stage of “survival and discovery” that it will be decided whether a new teacher develops professional competence or whether a struggling beginner leaves the profession. The first year of teaching determines teaching effectiveness, job satisfaction and career length (Fantilli & McDougall, 2009). In the 1980s, Veenman (1984) and other researchers (Kagan, 1992; Müller-Fohrbrodt, Cloetta, & Dann, 1978) have referred to beginning teachers’ first year as a “reality shock”; however, this view has changed towards a more differentiated description of teachers’ professional development. The general decline in teachers’ competence and attitudes cited in the above-mentioned studies could not be replicated in more recent investigations (Baer et al., 2011; Pigge & Marso, 1997). In studies that examine student achievement in classes taught by first-year teachers, there is a general consensus, that teachers make significant progress in teaching quality and effectiveness during the first year and that significant progress will continue for the next two years before gradually tapering off (Henry, Bastian, & Fortner, 2011; Kane, Rockoff, & Staiger, 2008; Rivkin, Hanushek, & Kain, 2005). There is no clear explanation for this positive growth in beginning teacher’s first years of teaching (Henry et al., 2011); experts argue that the individual characteristics of new teachers should be considered when examining teacher growth (McNally & Oberski, 2003). Since we also look at teacher self-efficacy, it has to be mentioned that findings seem to be a little different here: Woolfolk and Burke (2005) observed a decline in teacher efficacy during the first year of practice. With reference to Bandura (1997), it must be assumed that these first years of teaching are also critical to the long-term development of teacher efficacy.

The tri-national project "Alpha" of four teaching universities from Austria, Germany and Switzerland aimed to further investigate the development of beginning teachers' with a multi-perspective methodological approach. Since primary teacher education is more easily comparable between the three countries than secondary, we aimed at 3rd and 4th grade teachers. This study focuses on data obtained through rating of the beginners' videotaped lessons. A first analysis showed no difference in teaching competence between two measurements in their first year of teaching (Kocher, Wyss, & Baer, 2013); neither for the whole construct nor for its four dimensions (see Table 1, p. 29). We used a repeated measures MANOVA with the means of the items of t1 (beginning) and t2 (end of year one) for each teacher. While this older method of analyzing change of time focuses on group differences, latent growth curve models (LGCM) accentuate the individual changes over time (Voelkle, 2007). Since a LGCM is based on structural equation modelling it is possible to verify the supposed structure (factors) of the measuring model for teaching competence as well. This step has not yet been completed in previous analyses. Furthermore, we hadn't included covariates like age (continuous) or gender (categorical) in our analysis of change. Considering the results of Henry et al. (2011), who compared two groups of teachers – stayers and leavers – to distinguish individual development during the first years of teaching, it might be beneficial to examine additional predictors. For these reasons, we believe that it might be worthwhile to reanalyze the rating-data in order to gain additional information about teachers' first year of teaching.

2. The first year of teaching

There is a vast difference in the competence levels of beginning teachers. Competence can be seen as a complex ability construct that is context-specific, trainable, and closely related to real life (Koeppen, Hartig, Klieme, & Leutner, 2008). Some first year teachers possess a high level of competence when entering the field of teaching, whereas others possess basic rudimentary competence (Kane et al., 2008; Smit & Larcher, 2010). These differences are also evident in the years subsequent to the first year, according to Kane, Rockoff, and Staiger (2008), and are relevant predictors of whether new teachers will remain in the profession or stop teaching in the near future (Henry et al., 2011).

Local mentoring and induction programmes support a successful transition from the teacher university into the first year of employment in a school (Ingersoll & Strong, 2011; Nasser-Abu Alhija & Fresko, 2010). The quality of this support explains variations in the professional development of new teachers (Jensen, Sandoval-Hernàndez, Knoll, & Gonzalez, 2012). Interestingly, there is no relationship between mentoring or induction programmes and the amount of appraisal and feedback received by new teachers. Such feedback, however, is crucial for beginning teachers as they strive to develop and enhance both their teaching skills and their knowledge. In addition, the Organisation for Economic Co-operation and Development (OECD) Teaching and Learning International Survey (TALIS) (Jensen et al., 2012) presents significant differences in the quantity of professional development courses taken by new teachers across the 23 participating countries. Other fundamental elements of successful entry into the teaching profession include mentorship, support from colleagues and administrators, strong classroom management skills, student success, instructional resources, teaching assignment and workload, and parental contacts and support (Corbell, Reiman, & Nietfeld, 2008). Conversely, some of these aspects are issues of concern for many struggling first year teachers.

Teaching competence consists of a large number of discrete competencies (Blömeke & Delaney, 2012). Which of these teaching competencies are responsible for the differences between new teachers regarding the quality of instruction? Among the most challenging and important competencies are the ability to provide differentiated instruction in the classroom, the support and involvement offered by parents, and the teacher's skills with respect to classroom and time management (Fantilli & McDougall, 2009). Similar findings for classroom management and differentiated instruction as major concerns were presented in a Swiss study on stress factors for teachers in the transition from pre-service training to first year teaching (Zingg & Grob, 2002). The most problematic issue for the teachers, however, was the ability to balance their expectations with their personal established standards.

Upon entering the community of practitioners, the professional self-concept of a new teacher is confronted by the culture, expectations, traditions and ideas of the school environment and its team of professionals (Brunton, 2007). Accordingly, the new teacher is required to revise and adapt his or her professional self as a teacher and his or her knowledge and belief systems while also protecting personal interests and striving to perform the job in a way that satisfactorily meets the demands of the administration (Brunton, 2007). Levin, Hammer and Coffey (2009), for example, discovered that the professional context may cause first year teachers to place less emphasis on student orientation and more emphasis on classroom management, as the local school community also emphasizes this latter competency.

Stage theory describes teachers' developmental changes with increased classroom experience (Berliner, 1988; Fuller, 1969; Kagan, 1992). Conway and Clark (2003), among other scholars, continued research based on Fuller's "concerns theory", which includes the following three stages: 1) concerns about self, 2) concerns about tasks and 3) concerns about students and the impact of teaching. The results from this subsequent research shows that these stages no longer fall into a strict sequential order, as new teachers alternate among the stages as they strive to improve their level of professionalism (Watzke, 2007). According to Conway and Clark (2003), reverting to earlier stages entails heightened reflexivity and attention to the development of self-as-teacher. Watzke (2007) finds that of the three stages, impact-related concerns received the highest ratings during the two-year research programme with graduate student teacher candidates. The main impact concerns were "student growth" and "motivation", followed by "individual student differences" – diagnosing and teaching according to varied rates of student learning. So and Watkins (2005) analyzed the thinking of new primary school teachers. During their first year of classroom practice, the new teachers developed a slightly more constructivist thinking, thus placing the learner in the centre of their teaching aims. Most of the novice teachers, however, showed a learner-centred orientation from the beginning. This finding corresponds to Watzke's results, which argue that impact concerns rank high for most teachers from the moment they enter the teaching profession. It is worth noting that the new teachers in So and Watkins' sample simultaneously possessed several opposing teaching concepts. This phenomenon, however, did not appear to be problematic for the beginning teachers. The results from the TALIS study show that thinking and acting do not need to be congruent (Jensen et al., 2012). Although new teachers reported higher constructivist beliefs compared to more experienced teachers, they used fewer student-oriented and activity-enhanced teaching practices. That is, structured teaching practices were used more frequently by beginning teachers than by more experienced teachers.

3. Evaluating teaching competence

Although being a professional teacher implies that an individual possesses the appropriate pedagogical content knowledge (Shulman, 1986), it is also important for a teacher be able to effectively cope with a certain range of situations, thus demonstrating teaching competence. Consistent with Weinert's concept of competence (Weinert, 1999) and as applied in the TEDS-M framework (Tatto et al., 2008), we regard teacher quality as including context-specific cognitive dispositions that are acquired and required to successfully cope with teaching-related situations and tasks. Being a competent teacher involves the ability to select from and then orchestrate a set of competencies to achieve a particular end within a particular classroom situation (Blömeke & Delaney, 2012). Teacher education based on competencies has been utilized since the early 1970s in the United States (Popham, 1986), and it has also been a focus in many European teacher universities (Baer et al., 2009; Lunenberg, Snoek, & Swennen, 2000; McNally & Oberski, 2003). However, whether a competence or standards-based teacher education leads to higher teacher quality remains an unresolved question (Struyven & De Meyst, 2010).

In the past, measuring professional development consisted mainly of documenting teacher satisfaction, attitude change and/or commitment to innovation. Currently, however, the results of professional development or the degree to which professional development is successful is a major area of interest (Desimone, 2009). Video observations have the potential to assess teaching competence in complex teaching situations (Jacobs, Kawanaka, & Stigler, 1999; Seidel, Kobarg, & Prenzel, 2005). Clausen, Klieme, and Reusser (2003) successfully developed a video manual (with 94 items) for high inference rating of instructional quality to study cross-cultural differences in secondary schools (Grade 8) between Germany and Switzerland. In addition to items from process-product research into teaching (Brophy & Good, 1986), aspects of instruction for problem solving from the Trends in International Mathematics and Science Study (TIMSS) video-study were integrated. The existing possibilities to record lessons and then extensively analyze them enable more valid evaluations of teacher performance. Kersting, Givvin, Thompson, Santagata, and Stigler (2012) show that scores from video-analysis predict student learning. However, it remains a question whether the amount of taped classroom lessons and the selected competencies within studies with video analysis are sufficient to be considered a representative sample size. In addition, Kersting, et al. (2012) focused on competencies closely related to introducing a particular concept (fractions), whereas other projects covered more general competencies independent of closed topics. To cover both more general and subject-related competencies, M-Scan appears to be a promising tool (Merritt, Rimm-Kaufman, Berry, Walkowiak, & Larsen, 2011). The research group that developed the instrument combined knowledge tests for teachers and students to detect relations between instructional quality, student achievement and classroom context factors. This is important because the quality of the demonstrated teaching competence is subject to the existing conditions in the classroom and the school. For example, school climate, student-teacher relationships or student backgrounds may explain why teachers with similar competence demonstrate varying performance in different classes.

To summarize, video-analysis is a promising research tool that can facilitate deep analysis of individual teaching competence and its development during the first year of teaching. Yet, results based on data from rating processes should be analyzed more thoroughly to detect individual developments.

Research questions:
1) Can we validate the hypothetical competence construct and its dimensions with the help of structural equation modelling?
2) Does growth in competence differ across individual teachers during their first year of practice?
3) Is individual growth affected by covariates?

4. Method

4.1 Design

The present study was part of the Lake of Constance-IBH project "Aller Anfang ist schwer (ALPHA)" (Beginning is always difficult) and was conducted by four teacher education universities from Austria, Germany and Switzerland. The multi-perspective approach combined, among others, self-estimations of acquired teaching competencies and teacher-specific self-efficacy as well as expert estimations of teaching competencies. Our approach also utilized a vignette (situated problem) to measure the teachers' planning competence. We included students' achievement data from a standardized mathematics test and students' perceptions of instruction as dependent variables. The project began in the autumn of 2009 and lasted for two years. Its design was exploratory and longitudinal. We measured teaching competence at the beginning and at the end of the teachers' first year of teaching. Every lesson needed to begin with some type of instruction, and the subject had to be mathematics because the curricula of the three countries are similar for this subject. The research design did not focus on cross-country analysis, as the sample size of four institutions would be insufficient in each case.

4.2 Sample

It was difficult to find participating volunteers in one wave; therefore, we collected the data using two cohorts of beginning primary school teachers. The sample included a total of 73 participants (62 female and 11 male) who were nearly equally divided from each of the four teacher institutions. Of the participants, 38 were Swiss teachers (from two universities), 20 came from Austria and 18 from Germany. Of the teachers, 82 % started with a class grade 3 or 4. There may be a certain bias in the sample, as participation was voluntary. For example, beginning teachers with less self-confidence or low grades on the final teacher exam were likely under-represented. Furthermore, three teachers only participated up to the first measuring point. Of the teachers, 13 had completed an apprenticeship and 17 had studied another subject before entering teacher education (previous study experience). The beginning teachers' mean age was 25 (SD = 3.2).

4.3 Instruments

We used a procedure that allowed for us to indirectly measure an individual teacher's competence, as we deduced the level of teaching competence by teaching performance, which was recorded on video. Every point of measurement consists of one videotaped lesson (approximately 45 minutes). The subject had to be mathematics but the theme was free to choose. Every lesson was rated on 34 items with high inference rating scales. The 34 items belong to four competencies and eight sub-competencies under the overarching category of teaching competence (see Table 1). The rating manual originates from the project "Reaching standards in teacher education" (Baer et al., 2009) and is based on the research of Clausen, Reusser, and Klieme (2003), who used mainly scales derived from the TIMS-Study. The current research and that of Clausen et al. (2003) follow a research line based upon the work of Weinert, Schrader, and Helmke (1989). Weinert et al. (1989) expanded the process-product paradigm with their research findings and identified major characteristics of instructional quality and effectiveness with a high impact on student outcomes.

Two rating teams (each with two raters) evaluated each lesson on a 6-point rating scale that ranged between the poles "applies not at all" and "applies completely". The video data (143 tapes) of the participants from the four institutions and the two time points were equally distributed between the two teams of raters. Before the teams started working apart the rating was conducted within the whole group of four until a sufficient agreement upon the criteria among the raters was achieved. Kappa measures above .80 were expected for all items of the manual.

The inter-rater reliability with respect to the quality of the teaching competence values was calculated using generalizability theory (GT) (Brennan, 2000; Cardinet, Johnson, & Pini, 2010). The generalization approach allows the impact of different types of errors to be estimated by tracing the observed variance to different potential sources of variance (facets). More specifically, the approach distinguishes the amount of variance contributed by the facets "teaching competence", "item" and "rating group". The generalization coefficient expresses the quality of the rated values for teaching competence and was calculated using the EduG programme (Cardinet et al., 2010). Whereas an absolute G coefficient indicates how well a measurement procedure has located the objects under study on a scale, irrespective of where fellow objects are placed, a relative coefficient – comparable to Cronbach's α coefficient – indicates how well the procedure has ranked the objects.

Table 1: Items of the video-rating manual for teaching competence

Competencies	Sub-competencies	Items
A) Efficiency of instruction	Time on task	A11 Lesson is based on time schedule
		A12 Teacher is gaining students' attention before starting a lesson
		A13 Changing between different methods of instruction and learning goes smoothly – minimal loss of time
		A14 Teacher possesses high classroom management skills
	Quality of organization	A21 Teaching and learning material is ready
		A22 Teaching media (including blackboard and overhead projector) support student learning
		A23 Effective organization of student centred learning activities
		A24 Teacher uses a variety of teaching and learning methods
B) Student orientation	Handling of mistakes	B11 Teacher pays attention to not putting too much significance to mistakes.
		B12 Teachers never makes fun of students mistakes
		B13 Students are not afraid of making mistakes
	Motivation	B21 Teacher instructs actively and with commitment
		B22 Students are motivated and work with great concentration
		B23 Teacher encourages students
		B24 Teacher sparks interest at the beginning of the lesson
		B25 Teacher gives verbal/nonverbal impulses
		B26 Teacher acts with goodwill while interacting with students
C) Cognitive activation	Pacing	C11 Teacher introduces new subject matter without overstraining students
		C12 Students get enough time for completing tasks
		C13 Teacher supports students learning during phases of self-directed learning
		C14 Teacher adjusts instruction to students' needs
	Teacher as facilitator	C21 Teacher helps students to formulate or realize their own ideas
		C22 Teacher checks – without judging – when student formulates unclear or incomplete ideas
		C23 Teacher supports students solving problems
		C24 Teacher gives individual feedback
		C25 Teacher assists students' learning with scaffolding
		C26 Teacher assists students' learning with coaching
D) Clarity and structuring	Classroom discourse	D11 Teacher uses high quality content knowledge language
		D12 Teacher adapts language to students' level
		D13 Teacher possesses different strategies using students' answers for classroom discourse
		D14 Teacher uses questions in a learning fostering way
	Learning goals	D21 Lesson is apparently based on learning targets
		D22 Orders and instructions are clearly formulated
		D23 A connection between different phases of instruction is visible

As every teacher was rated on every item, but the teachers were shared between two rating groups, we used a two-level design with individuals and items nested in rating groups. The G coefficients for relative measurement are sufficient (> .80) to measure change of teaching competence (t1: G rel .93 and G abs. 68; t2: .91 and .62) (Cardinet et al., 2010). For the measuring point t1, 4.5 % (t2 0.5 %) of the error variance was due to the interaction rating group x item, and 12.9 % (t2 16.2 %) was due to the interaction teacher x rating group. The judgment error components were relevant and caused bias in the competence values. Because the teachers' videos were exchanged between the two rating groups and the two measure points, part of the bias was randomly distributed between all teachers' competence values.

As a covariate, we measured teachers' self-efficacy in a questionnaire with 3 items based on Schwarzer & Jerusalem (1999). The Likert scale (1 = don't agree – 4 = fully agree) shows a satisfactory Cronbach's α of .81 and a mean of 3.30 (*SD* = .31) at the beginning of teachers' experience with teaching. A sample item is "I know that I can manage to impart the subject matter relevant for the next test even to the most problematic student".

4.4 Confirmatory factor analysis and bootstrapping

The configural structure of the construct "teaching competence" was tested on the basis of a confirmatory factor analysis (CFA) model using Mplus 7 (L. K. Muthén & Muthén, 2012). We used the default estimator FIML, which includes missing values to estimate the parameters, and applied data bootstrapping for final estimates. With small sample sizes or non-normal data, bootstrapping can help to obtain estimates that are nearly identical to FIML-estimates (Hox, 2010). In bootstrapping, random samples are repeatedly drawn with replacement from the observed data. This procedure is repeated at least 1,000 times to produce sufficient accuracy for the parameter estimates. We chose an *N* of 5,000 samples, as did Kersting et al. (2012). Model fit was evaluated using several goodness-of-fit measures, including chi-square, Bentler's comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA). CFI and TLI values of 0.95 or above and RMSEA values of 0.05 or below indicate a good model fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003).

4.5 Investigating measurement invariance

To test for measurement invariance between the two time points, we conducted a series of CFA following an approach that is well established in the literature on structural equation modelling (Meredith, 1993; Vandenberg & Lance, 2000). In the first step, we tested a configural invariance model, in which the same pattern of fixed and free factor loadings was specified for each time. In the second step, we tested a metric invariance model, in which factor loadings for like items were set to be invariant across time. Third, we tested a scalar invariance model, which required the intercepts of like items' regressions on the latent factor to be invariant across time. The grade of measurement invariance can be determined by comparing the different models using a chi square-difference test.

4.6 Growth modelling and Bayesian estimation

Most analyses offer information on how group scores (mean differences) change or the extent to which the rank orderings of individual scores (correlations) are similar over time, i.e., inter-individual changes. These analyses do not, however, offer information on changes that might occur within individuals, i.e., intra-individual changes. Growth modelling is an approach that eliminates many of the problems that have traditionally plagued the measurement of change (Gibbons, Gibbons, Hedeker, & DuToit, 2010). In latent class growth modelling, different time points are considered as classes that define different trends over time in the item probabilities. For example, a trend might be linear with an intercept and a slope (B. O. Muthén, 2001). Another approach to analyze longitudinal data is growth mixture modelling, which typically captures individual differences in development by random effects. These random effects represent continuous variation across individuals in growth features such as initial status and rate of change.

Especially for small sample sizes and non-normal distributed samples, Bayesian analysis is an attractive alternative to ML estimation (B. O. Muthén, 2010; B. O. Muthén & Asparouhov, 2012). The Bayesian estimates are obtained as means, modes, or medians of their posterior distributions. Priors can help to optimize small variance parameters. Mplus uses a series of default priors. Bayesian exploration of model fit can be performed in a flexible way using Posterior Predictive Checking. An excellent fitting model is expected to have a PPP value approximately 0.5 and an *f* statistic difference of zero falling close to the middle of the confidence interval. A positive lower limit is in line with a low PPP and indicates poor fit. A 95 % confidence interval is produced for the difference in the *f* statistic for the real and replicated data (B. O. Muthén & Asparouhov, 2012).

5. Results

5.1 CFA of teaching competence

Our first step consisted of verifying the four assumed dimensions of the video manual as part of a general construct "teaching competence" using a bootstrapping procedure. The fit indices of this model were not satisfactory: Chi-squared = 1131.36, df = 523, CFI = .78, TLI = .77 and RMSEA = .10. To gain further information for improvements, we conducted an EFA with an oblique rotation of geomin and with 1 to 4 factors to be extracted. We also checked for outliers using SPSS and adjusted a few measures for which the differences between the next higher or lower rating were so slight as to render them negligible. Fit indices for the four-factors model were best, but only two of the four factors showed clear factor loadings of more than two items of the same sub-dimension from the video manual. As a consequence, we dropped the dimensions "efficiency of instruction" and "clarity and structuring", which were originally part of the video rating manual for further CFA analysis. To establish a model with good or acceptable model fit indices, we further limited the model to 3 sub-dimensions of the total assumed 7 from the rating manual. Because of problems such as negative (residual) variances, not all of the assumed items of a sub-dimension were included in the final model. The final factors were motivating (B22, B24), pacing (C12, C14) and facilitator (C13, C21, C23, C25). All fit indices were either good or acceptable: Chi-squared = 134.56, df = 97, CFI = .95, TLI = .94 and RMSEA = .07. Two items focus on the teacher's competency to motivate and interest students through his or her instruction. The next two items centre on the teacher's competency to adapt his or her teaching, which implies a high diagnostic competency. Lastly, four items cover the teacher's competency as a coach or facilitator during phases of self-directed learning or classroom dialogue. Relating to our research question 1, we couldn't confirm the hypothesized structure of our construct of teaching competence with four dimensions. Of note, we are aware that more relevant dimensions of teaching competence exist; however, we were not able to measure them in a reliable and valid manner. Based on our data from the video rating, we could only establish a reduced model with three competencies. This does not affect the goal of our second research question, i.e., to investigate individual developments of teaching beginners on these competencies.

5.2 Growth of teaching competence

As a next step, we examined the metric invariance over time following a procedure recommended by Geiser (2012) as a precondition to estimate growth in teaching competence during the first year of teaching. Table 2 shows the results of the CFA analyses. A first model must show that the number of factors and loading pattern remain the same from time 1 to time 2. In a second model, we needed to test for weak factorial invariance. We therefore constrained the factor loadings to be equal over time. Lastly, for strong factorial invariance, we constrained the intercepts to remain the same from t1 to t2. Chen, Sousa, and West (2005) recommend to proceed with second-order models in a hierarchical manner and to test measurement invariance for first- and second-order factors. Consequently, we present results first for constraints on first-order variables only and then for first- and second-order variables together. The fit indices indicate that configural, metric and scalar models were acceptable for the measurement of teaching competencies over time. Because the Chi-squared difference tests for the comparison of the three types of invariance models were not significant, at least not for first-order only factor model comparisons, we used constraints for intended growth modelling to make valid inferences about the differences between latent factor means in the model.

Table 2: Series of CFA models investigating measurement invariance between the two time points of teaching competencies

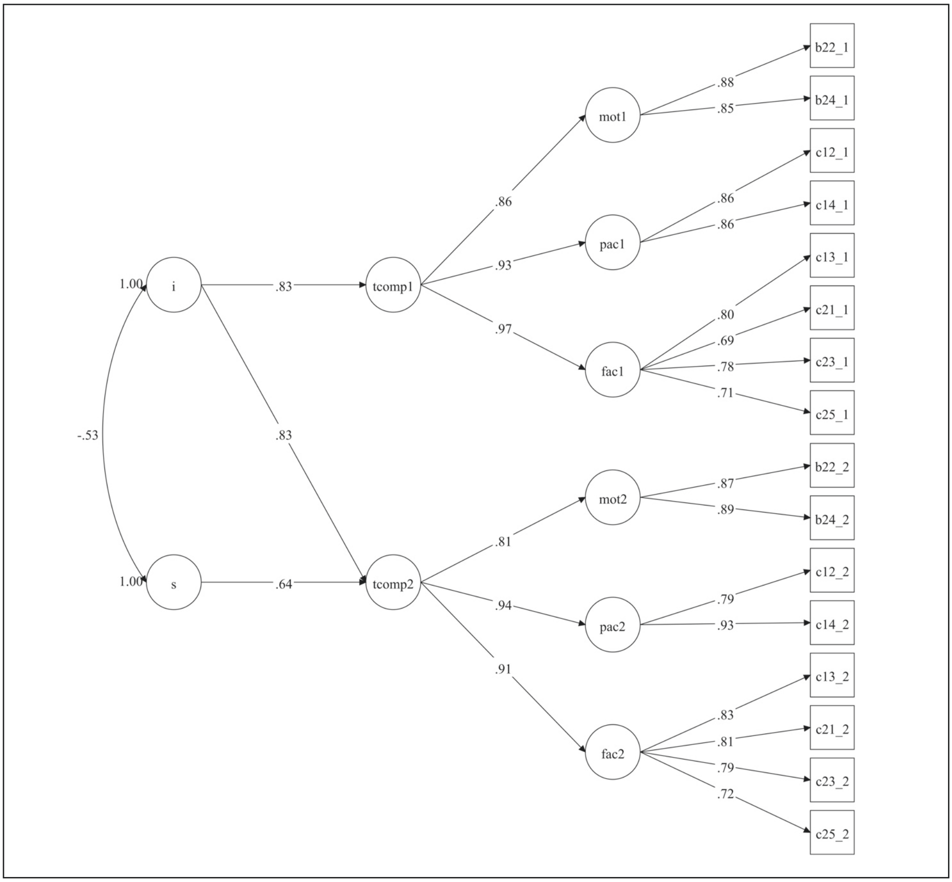
Parameters constrained to be equal	χ^2	df	CFI	TLI	RMSEA	$\Delta\chi^2$
1. Unconstrained (configural invariance)	129.63	97	.954	.943	.068	-
2. Factor loadings first-order only (metric invariance)	140.68	102	.946	.936	.072	n.s.
3. Factor loadings first and second-order (metric invariance)	141.93	104	.947	.939	.071	n.s.
4. Intercepts first-order only (scalar invariance)	148.99	106	.940	.932	.075	$p < .05$
5. Intercepts first and second-order (scalar invariance)	149.31	108	.942	.936	.072	n.s.

Note. CFA = confirmatory factor analyses; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; model comparison with Chi-squared test.
$\Delta\chi^2$ 5 – 3: n.s., $\Delta\chi^2$ 4 – 2: n.s. Bootstrapping estimates.

To examine growth of the three competencies, a 2-order factor Latent Growth Curve Model (LGCM) was applied using Mplus 7 (B. O. Muthén, 2001). We used our CFA model with scalar invariance as a basis and added the two factors intercept and slope (Figure 1) and applied Bayesian estimation with a Gibbs-Algorithm. The first higher order factor (the intercept factor; i) describes the initial level of the teaching competencies (intercept mean) and individual differences in the initial level (intercept variance). The intercept is a constant for any given individual across time; therefore, the factor loadings for teaching competencies measures were set at 1 for each wave. The second factor in LGCM (the slope factor; s) describes the rate of change (slope mean) and individual differences in growth patterns (slope variance). The intercepts of the response variable (teaching competencies) at the two time points are constrained to zero; as a result, the mean of the intercept factor is an estimate of the common intercept (true score variable for the first point of measurement). The successive loadings for the slope factor define the slope as the linear trend over time (Duncan, Duncan, & Strycker, 2006). The mean of the slope factor is an estimate of the common slope (Hox, 2010; Meredith &

Tisak, 1990). Individual deviations from the true score intercept variable are modelled by the variance of the slope factor. The correlation between the intercept and the slope indicates whether an individual with a high intercept also displays a high slope.

Figure 1: Constrained LGCM model with scalar invariance on first-order factors for three teacher competencies (standardized estimates from Bayesian analysis)



After conducting estimations with different iterations to check convergence and PSR value, the output of the LGCM model (Figure 1) analysis showed a stable result. The PPP value amounted to .11 and an *f* difference of 53.04. The number of free parameters was 52, the deviance (DIC) was 2277.85, the Bayesian (BIC) was 2407.69, and the estimated number of parameters was 43.72. The plots to control for "burn in" and the posterior distribution were satisfactory. The results show that there were significant individual differences in the means of the intercept but not in the slope. This indicates that the beginning teachers showed different levels of teacher competencies at time point 1 and 2 but no differences in growth (or decline). The correlation between intercept and slope was not significant. With regard to our research question 2, it can be stated that there are no significant individual differences in the rate of change in competence for our beginning teachers in their first year of teaching.

In a second LGCM, we added the covariates age, gender, working experience, study experience and teacher self-efficacy as factors both to the intercept and the slope. This second model had a slightly better fit. The output of the LGCM model analysis showed a PPP value of .35 and an *f* difference of 16.97. The number of free parameters was 55, the deviance (DIC) 2281.00 and the Bayesian (BIC) was 2417.19. The estimated number of parameters (PD) was 47.29. To assess the convergence of the MCMC chain, we examined the PSR statistics, which ended after 60,000 iterations at 1.018 (close to 1). We also checked the plots; we especially controlled for a "burn in" of the two chains as a pattern for convergence. Table 3 shows the results for the effects of the covariates on the intercept and the slope. Although there were no significant results for the intercept, there was a significant effect of having study experience on slope. That is, beginning teachers with study experience show a stronger growth in teacher competencies than those without study experience. Teachers with a study experience had a higher mean age, and this group included a higher percentage of men; however, these two variables caused no effect on intercept and slope in the growth model. Interestingly, according to the intercept, the group with study experience began with a lower level of competence than the group without such experience.

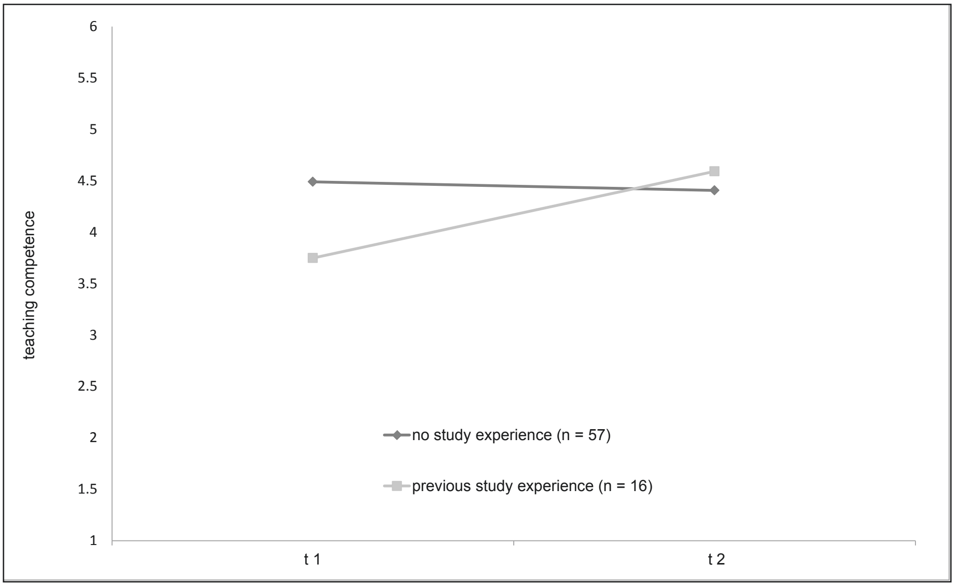
Table 3: LGCM model of three teacher competencies with covariates (standardized estimates from Bayesian analysis)

	Estimate	Posterior S.D.
Intercept	8.79*	3.25
Slope	.17	.37
Slope with Intercept	-.46	.41
Intercept on		
Age (higher)	.08	.20
Gender (male)	.08	.17
Working experience (yes)	-.14	.20
Study experience (yes)	-.24	.18
Teacher self-efficacy (higher)	-.13	.14
Slope on		
Age (higher)	-.36	.23
Gender (male)	.15	.20
Working experience (yes)	.02	.23
Study experience (yes)	.49*	.20
Teacher self-efficacy (higher)	.21	.17

*$p < .001$

With the help of a Growth Mixed Model (GMM), based on the LGCM and previous study experience as a grouping variable, we tested whether the slopes of the two groups were significant (Figure 2). Whereas the mean of the slope for the group without study experience was not significant (M = -.01; posterior SD = .23; n = 57), the mean of the slope of the group with study experience was significant (M = 2.69, posterior SD = 1.15, p < .001, n = 16). There were no significant differences between the group means at both times, but the group of study experience has at t2 a higher mean than the group without, whereas at t1 it was the opposite. An examination of the information on their previous studies revealed a wide range of fields, e.g., History, Economics, Art and Medicine.

Figure 2: Slopes of beginning teachers with and without previous experience in other study subjects



6. Discussion

The results of the current study indicate that it is often worthwhile to reconsider missing effects, such as those that we encountered both in this project and the preceding project "Reaching standards (in teacher education)" (Baer et al., 2011), when examining rated video data from the first year of teaching. We examined in this study the individual development of beginning teacher's competence. Therefore we constructed a LGCM, including covariates as well. Before we could perform a LGCM, we had to make sure that the measuring model was invariant over time; the concept of teaching competence assessed should be the same at both points of measurement. Our analysis showed only partial validity of the factorial structure of the original competence model used for the video rating.

At this point we would like to discuss two threats to reliability and validity relevant to our research that are not located in the operationalization of the construct but in the testing condition and in the rating of the video material. They might help to explain the methodological and other results of our study. First, we used an open, authentic setting for our performance assessment of teaching competence as suggested by Shavelson (2012). Choosing a real-life setting implies some issues with the standardization of the testing condition: Classes are different in their composition. A highly diverse class might force a teacher to spend more time on classroom management, having less time for instruction. Therefore, the reality of the teaching situation might hinder a teacher from showing his or her full potential when measuring competence. Novice teachers are perhaps even more affected by these varying circumstances. A study by Buddin and Zamarro (2009) presented positive effects of teacher quality on student achievement in different schools with high- and low-test scores and varying student backgrounds. Variation in student achievement between schools could be explained largely by student background and other student characteristics. In addition, it is worth noting that schools with high numbers of students from less-educated families have a higher turnover of teachers and a larger proportion of beginning teachers (Rivkin et al., 2005). Consequently, it is important to include student characteristics when comparing teacher competencies. Integrating criteria for the teaching context could add to validity of the video-rating. Second, rater bias is a severe problem for the reliability of the measurements (Praetorius, Lenske, & Helmke, 2012). We suppose, that in our study agreement between the rating teams drifted a little apart over time. Hence, to maintain a high reliability it is recommended to keep up the rater-training for the whole rating period. This could be achieved by conducting some video-ratings in the whole group from time to time. Praetorius et al. (2012) make several suggestions how to raise the quality of rating measures. Among them is the recommendation to use a higher number of raters, which could add to reliability. Since video-rating is time-intensive the high costs will make it difficult to implement this.

After reanalyzing the qualitative data from 34 rated items with quantitative methods, we could only use a few items to build three reliable factors in a longitudinal CFA model. In the next few lines we will further discuss these three factors and their relevance in teacher development. The teachers showed considerable individual differences regarding the teaching competencies at the beginning of their first year but not in growth. Motivating, pacing and facilitating are three important competencies with a high impact concern (Watzke, 2007). According to Watzke's study (2007), based on stage concern theory, impact concerns were most troublesome to beginning teachers. Those topics or competencies focus on student academic growth as a result of cognitive activation (see Baumert et al., 2010) and student motivation. A feature of cognitive activation and impact concern that appears to be difficult for new teachers is adapting teaching to students' individual differences and scaffolding individual learning. Baumert and colleagues (2010) identified cognitive activation and individual learning support, but not classroom assessment, as mediating factors that explain the differences in student achievement. According to Reynolds (1992), classroom management ceases to be a major issue for beginning teachers by the end of their first year of teaching. However, with respect to differences in teachers' content knowledge as a precondition for cognitive activation (see Baumert et al., 2010), impact concerns remain troublesome for a longer period than the first year of teaching. This could be a reason that we did not discover growth for the three competencies in the current analysis. Although we focused on primary school teachers, these three factors are issues in secondary school education as well (Watzke, 2007). From a methodological point of view more measurements within the first year would have been useful. Using latent growth curve modeling as a methodology is especially beneficial when considering that the means of the whole sample may conceal individual developments. We couldn't achieve the full potential of this method, however; more than two time points would have allowed testing for different shapes of the developmental trajectories and not just the shape of a straight line.

We identified (with the help of a covariate – previous study experience) two groups of beginning teachers who showed significant differences in their competence trajectories. These teachers showed differences in their professional development. Most significant, only the beginning teachers with a background from other study areas were able to learn from their experiences during their first year of teaching and progressed. Although the current teachers all underwent the same education program, results from an American study (Owings et al., 2006) on the quality of teachers with alternative certification might support the present findings. In this American study, supervising administrators overwhelmingly indicated that these teachers performed better in all instructional areas than traditionally prepared teachers with comparable teaching experience. Findings from the current study allude that teachers who enter their career with study experience perhaps collaborate more with other school team members. Positive interactive teaching approaches were also reported by an English study on the programme Teach First for graduate students (Muijs, Chapman, & Armstrong, 2013) holding a previous degree from a different field. To further enlarge our knowledge about how those freshly graduated teachers showing growth were able to develop their competencies, qualitative methods such as interviews would be appropriate. Contrary to theory (Tschannen-Moran & Hoy, 2001), teaching-efficacy was not helpful in explaining individual differences in teaching competencies.

We would like to end with a reference to the research of Kersting et al. (2012), who also worked with video observation to measure teacher quality; however, they added a knowledge test to their study. In line with this group, we believe that video analysis is a promising tool for measuring teaching quality. Our entire project "Alpha" combined several measures, such as attitudes, planning and instruction quality, in addition to student achievement in a longitudinal setting, as suggested by Kersting et al. (2012). This research group considers videos as helpful in assessing what teachers know and how they apply their knowledge. Similarly, we use the term competence to measure teachers' performance as presented in the videos and based on their subject and content knowledge. Kersting et al. (2012) applied their video measures to demonstrate mediating effects of instructional quality on student achievement affected by teacher content knowledge. Domain-specific processes have large effects on cognitive outcomes (Baumert et al., 2010; Hill, Rowan, & Ball, 2005; Seidel & Shavelson, 2007). Testing beginning teachers on their subject and content knowledge could be a significant addition to our future research.

Acknowledgments

We would like to thank the reviewers for their thoughtful comments. This research was funded by a grant from the Internationale Bodensee Hochschule. The research team of the ALPHA project consisted of Matthias Baer, Gabriele Böheim, Veronika Fain, Titus Guldimann, Petra Hecht, Sara Helfenstein, Mirjam Kocher, Diemut Kucharz, Katja Mackowiak, Robbert Smit, and Corinne Wyss.

References

Baer, M., Doerr, G., Fräfel, U., Kocher, M., Küster, O., Larcher, S., Müller, P., Sempert, W., & Wyss, C. (2009). Competences and standards in teacher education in Switzerland and Germany – Do prospective teachers become more competent through teacher training. In F. Achtenhagen, F. Oser, & U. Renold (Eds.), *Teachers' professional development: Aims, modules, evaluation* (pp. 145–174). Rotterdam, Netherlands: Sense.

Baer, M., Kocher, M., Wyss, C., Guldimann, T., Larcher, S., & Dörr, G. (2011). Lehrerbildung und Praxiserfahrung im ersten Berufsjahr und ihre Wirkung auf die Unterrichtskompetenzen von Studierenden und jungen Lehrpersonen im Berufseinstieg [Teacher training and practical experiences in the first year of work and their effects on the teaching competencies of students and teachers at the start of their career]. *Zeitschrift für Erziehungswissenschaft, 14*(1), 85–117.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W.H. Freeman and Company.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.

Berliner, D. C. (1988). *The development of expertise in pedagogy*. Washington, DC: AACTE Publications.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM Mathematics Education,* 223–247.

Brennan, R. L. (2000). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement, 24*(4), 339–353.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York, NY: Mcmillan.

Brunton, K. (2007). *Beginning teachers and their reality of teaching: Negotiating the micro-political world*. Paper presented at the the bi-annual conference of the Inter national Study Association on Teachers and Teaching (ISATT), Brock University, St. Catherines, Canada.

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics, 66*(2), 103–115.

Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying generalizability theory using EduG*. New York, NY: Routledge.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal, 12*(3), 471–492. doi: 10.1207/s15328007sem1203_7.

Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilung: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz [Using high-inference ratings to assess quality of instruction. A comparison between Germany and the German-speaking part of Switzerland]. *Unterrichtswissenschaft, 31,* 122–141.

Conway, P. F., & Clark, C. M. (2003). The journey inward and outward: A re-examination of Fuller's concerns-based model of teacher development. *Teaching and Teacher Education, 19*(5), 465–482.

Corbell, K. A., Reiman, A. J., & Nietfeld, J. L. (2008). The perceptions of success inventory for beginning teachers: Measuring its psychometric properties. *Teaching and Teacher Education, 24*(6), 1551–1563.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Towards better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling concepts, issues, and applications*. Mahwah, NJ: Erlbaum.

Fantilli, R. D., & McDougall, D. E. (2009). A study of novice teachers: Challenges and supports in the first years. *Teaching and Teacher Education, 25*(6), 814–825.

Feiman-Nemser, S., Schwille, S., Carver, C., & Yusko, B. (1999). A conceptual review of literature on new teacher induction. Washington, DC: National partnership for excellence and accountability in teaching.

Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal, 6*(2), 207–226.

Geiser, C. (2012). *Data analysis with Mplus (Methodology in the social sciences)*. New York, NY: Guilford Press.

Gibbons, R., Gibbons, D., Hedeker, S., & DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual review of clinical psychology, 6*(1), 79–107.

Grossman, P. (1992). Why models matter: An alternate view on professional growth in teaching. *Review of Educational Research, 62*(2), 171–179.

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and leavers: Early-career teacher effectiveness and attrition. *Educational Researcher, 40*(6), 271–280.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Ingersoll, R. M., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers. *Review of Educational Research, 81*(2), 201–233.

Jacobs, J. K., Kawanaka, T., & Stigler, J. W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research, 31*, 717–724.

Jensen, B., Sandoval-Hernàndez, A., Knoll, S., & Gonzalez, E. (2012). The experience of new teachers: Results from TALIS 2008. Paris, France: OECD Publishing.

Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research, 62*(2), 129–169.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*(6), 615–631.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–589. doi: 10.3102/0002831212437853.

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie, 216*(2), 61–73. doi: 10.1027/0044-3409.216.2.61.

Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. *Journal of Teacher Education, 60*(2), 142–154.

Lunenberg, M., Snoek, M., & Swennen, A. (2000). Between pragmatism and legitimacy: Developments and dilemmas in teacher education in The Netherlands. *European Journal of Teacher Education, 23*(3), 251–260. doi: 10.1080/02619760120049139.

McNally, J., & Oberski, I. (2003). Right at the start: An agenda for research and development in teacher induction. *Teacher Development, 7*(1), 59–74. doi: 10.1080/13664530300200186.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. doi: 10.1007/BF02294825.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107–122. doi: 10.1007/BF02294746.

Merritt, E. G., Rimm-Kaufman, S. E., Berry, R. Q., Walkowiak, T. A., & Larsen, R. A. (2011). *The contribution of mathematics instructional quality and class size to student achievement for third grade students from low income families*. Paper presented at the Society for Research of Educational Effectiveness Conference, Washington DC.

Muijs, D., Chapman, C., & Armstrong, P. (2013). Teach First: Pedagogy and outcomes. The impact of an alternative certification programme. *Journal for Educational Research Online, 4*(2), 29–64.

Müller-Fohrbrodt, G., Cloetta, B., & Dann, H.-D. (1978). *Der Praxisschock bei jungen Lehrern: Formen, Ursachen, Folgerungen: eine zusammenfassende Bewertung der theoretischen und empirischen Erkenntnisse* [The practice-shock of young teachers: Forms, reasons, conclusions: A comprehensive assessment of theoretical and empirical findings]. Stuttgart, Germany: Klett.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–34). Mahwah, NJ: Lawrence Earlbaum.

Muthén, B. O. (2010). *Bayesian analysis in Mplus: A brief introduction*. Retrieved from https://http://www.statmodel.com/download/IntroBayesVersion3.pdf

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335. doi: 10.1037/a0026802.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus: Statistical analysis with latent variables: User's Guide* (7th ed.). Los Angeles: Muthén & Muthén.

Nasser-Abu Alhija, F., & Fresko, B. (2010). Socialization of new teachers: Does induction matter? *Teaching and Teacher Education, 26*(8), 1592–1597.

Owings, W. A., Kaplan, L. S., Nunnery, J., Marzano, R., Myran, S., & Blackburn, D. (2006). Teacher quality and troops to teachers: A national study with implications for principals. *NASSP Bulletin, 90*(2), 102–131.

Pigge, F. L., & Marso, R. N. (1997). A seven year longitudinal multi-factor assessment of teaching concerns development through preparation and early years of teaching. *Teaching and Teacher Education, 13*(2), 225–235.

Popham, W. J. (1986). Teacher competency testing: The devil's dilemma. *The Journal of Negro Education, 55*(3), 379–385.

Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and instruction, 22*(6), 387–400. doi: 10.1016/j.learninstruc.2012.03.002.

Reynolds, A. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research, 62*(1), 1–35.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23–74.

Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen* [Scales to measure teacher- and student attributes]. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin. http://www.psyc.de/skalendoku.pdf.

Seidel, T., Kobarg, M., & Prenzel, M. (2005). *How to run a video study*. Münster, Germany: Waxmann.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499.

Shavelson, R. J. (2012). Assessing business-planning competence using the Collegiate Learning Assessment as a prototype. *Empirical Research in Vocational Education and Training, 4*(1), 77–90.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *American educational research journal, 15*(2), 4–14.

Smit, R., Helfenstein, S., & Guldimann, T. (2013). Effekte der Unterrichtskompetenz von Berufseinsteigenden auf das Schülerbefinden und die Schülerleistungen [Effects of first year teachers' competence on students' perceptions and achievement]. *Unterrichtswissenschaft, 2*, 153–172.

Smit, R., & Larcher, S. (2010). Unterrichtsvideografie als Entwicklungsgrundlage für ein Kompetenzmodell "Unterrichtskompetenz" [Teaching videos as a basis for the development of a competence-model for "teaching competence"]. *Lehrerbildung auf dem Prüfstand, 3*(2), 199–216.

So, W. W. M., & Watkins, D. A. (2005). From beginning teacher education to professional teaching: A study of the thinking of Hong Kong primary science teachers. *Teaching and Teacher Education, 21*(5), 525–541.

Struyven, K., & De Meyst, M. (2010). Competence-based teacher education: Illusion or reality? An assessment of the implementation status in Flanders from teachers, and students, points of view. *Teaching and Teacher Education, 26*(8), 1495–1510. doi: 10.1016/j.tate.2010.05.006.

Tatto, M. T., Ingvarson, L., Schwille, J., Peck, R., Senk, S. L., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary Mathematics*. Conceptual Framework: International Association for the Evaluation of Educational Achievement. Amsterdam.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805. doi: http://dx.doi.org/10.1016/S0742-051X(01)00036-1.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. doi: 10.1177/109442810031002.
Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*(2), 143–178.
Voelkle, M. C. (2007). Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Science, 49*(4), 375.
Watzke, J. L. (2007). Longitudinal research on beginning teacher development: Complexity as a challenge to concerns-based stage theory. *Teaching and Teacher Education, 23*(1), 106–122.
Weinert, F. E. (1999). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45–66). Seattle WA: Hogrefe & Huber.
Weinert, F. E., Schrader, F. W., & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research, 13*(8), 895–914.
Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*(4), 343–356.
Zingg, C., & Grob, U. (2002). Belastungswahrnehmungen von Primarlehrpersonen im Kanton Zürich beim Berufseinstieg [Perceived stress of primary school teachers in the Canton of Zurich at career entry]. *Beiträge zur Lehrerbildung, 20*(2), 216–226.