Early childhood teachers' mathematics-specific professional competence and its relation to the quality of learning support

Competência profissional de professores que ensinam matemática e sua relação com a qualidade instrucional

Anke Lindmeier[1]

Selma Seemann[2,*]

Andrea Wullschleger[3,*]

Anuschka Meier-Wyder[4,*]

Miriam Leuchter[5]

Franziska Vogt[6]

Elisabeth Moser Opitz[7]

Aiso Heinze[8]

[1] PhD in Mathematics Education from the Technical University of Munich, Germany. Professor at the University of Jena, Germany. Email: anke.lindmeier@uni-jena.de. **ORCID:** http://orcid.org/0000-0002-2819-4375

[2] Master in Educational Science from the Christian Albrecht Universität zu Kiel, Germany. PhD student at the IPN – Leibniz Institute for Science and Mathematics Education Kiel, Germany. Email: seemann@leibniz-ipn.de. **ORCID:** http://orcid.org/0000-0002-5577-4826

[3] PhD in Educational Science from the University of Zurich, Switzerland. Postdoctoral researcher at the University of Zurich, Switzerland. Email: awullschleger@ife.uzh.ch. **ORCID:** http://orcid.org/0000-0003-4390-3835

[4] PhD in Education Studies University of Zurich, Switzerland. Senior lecturer at Interkantonale Hochschule für Heilpädagogik Zürich, Switzerland. Email: anuschka.meier@hfh.ch. **ORCID:** https://orcid.org/0000-0003-0703-2403

[5] PhD in Pedagogical Psychology from the University of Zurich, Switzerland. Professor at the University of Koblenz-Landau, Germany. Email: leuchter@uni-landau.de. **ORCID:** http://orcid.org/0000-0002-7962-6561

[6] PhD in Educational Research from Lancaster University, UK. Professor at the St.Gallen University of Teacher Education. Email: Franziska.vogt@phsg.ch. **ORCID:** http://orcid.org/0000-0002-2023-0431

[7] PhD in Special Education from the University of Freiburg/Switzerland. Professor at the University of Zurich, Switzerland. Email: elisabeth.moseropitz@uzh.ch. **ORCID:** http://orcid.org/0000-0002-5243-4770

[8] PhD in Mathematics Education from the Carl von Ossietzky-Universität Oldenburg. Professor at the IPN – Leibniz Institute for Science and Mathematics Education Kiel, Germany. Email: heinze@leibniz-ipn.de. **ORCID:** https://orcid.org/0000-0002-7408-0395

* These authors contributed equally to this work.

Abstract

Teacher expertise research uses different approaches, among them competence models. The model of mathematics-specific professional competence in this study differentiates action-related and reflective competence, which pertain to different types of professional demands according to dual processing theories and should positively impact teaching quality. The cross-sectional study with early childhood teachers investigates the structure of teacher competence and its relations to the quality of learning support in line with the notion of adaptive teaching when playing board games. The findings support a differentiation of competences according to the model (structural validity evidence). Expected relations between competences and the quality of the learning support are partially supported (predictive validity evidence). The study hence adds to research on teacher expertise by developing a specific understanding of competence in line with existing theories. Affordances of using competence models, as well as difficulties and open questions with relation to the emerging approach are discussed.

Keywords: Teacher competence; Adaptive teaching; Validation study; Situated assessment

Resumo

A pesquisa da expertise de professores usa diferentes abordagens, dentre elas modelos de competência. O modelo de competência profissional específica da matemática neste estudo diferencia a competência relacionada à ação e a competência reflexiva, que pertencem a diferentes tipos de demandas profissionais de acordo com as teorias de processamento dual e devem impactar positivamente a qualidade do ensino. O estudo transversal com professores da Educação Infantil investiga a estrutura de competência docente e suas relações com a qualidade do suporte à aprendizagem em consonância com a noção de ensino adaptativo utilizando jogos de tabuleiro. Os resultados apóiam uma diferenciação de competências de acordo com o modelo (evidências de validade estrutural). As relações esperadas entre as competências e a qualidade do suporte à aprendizagem são parcialmente suportadas (evidências de validade preditiva). O estudo, portanto, contribui à pesquisa sobre a experiência do professor, desenvolvendo uma compreensão específica de competência de acordo com as teorias existentes. São discutidos os recursos do uso de modelos de competência, bem como as dificuldades e questões em aberto em relação à abordagem emergente.

Palavras-chave: Competência do professor; Ensino adaptativo; Estudo de validação; Estudo de validação.

Introduction

Research aims to understand the nature of teacher expertise and its effects on instructional processes. Although some studies could establish connections between teacher characteristics and instructional quality, the findings are still mixed. One explanation starts from the consideration that the predominant research focus (in quantitative approaches) on professional knowledge as declarative, decontextualized knowledge may be too limiting so as to capture relevant aspects of teacher expertise. Alternative approaches model different aspects of teacher expertise through delineating constructs of teacher competence closely connected to professional demands. Their use in quantitative approaches is still under development. So far, the question remains unanswered as to whether assessment tools can capture the richness of such competence constructs and whether they are actually suited to trace instructional effects of teacher expertise.

The article addresses this research gap and reports a study on mathematics-specific professional competence of early childhood teachers as knowledge-based dispositions to master professional demands. The study focuses on aspects of validity of the assessment approach. We first present the research background and discuss conclusions for the structure, assessment, and instructional effects from a theoretical perspective. As a means to investigate

validity aspects, we relate teachers' competence scores to performance measures of instructional quality when playing board games in kindergarten.

Theoretical Background

Shifting the view from teacher knowledge to teacher competence

Teacher knowledge has become an important area of research in the last decades. Based on the seminal differentiation between content and pedagogical content knowledge (CK, PCK, Shulman, 1986), an area of research developed with focus on components of teacher (subject-specific) knowledge (e.g., Petrou & Goulding, 2011; Hill et al., 2008). Empirical investigations revealed important findings on the structure of mathematics teachers' knowledge, its development in teacher education, and international differences (e.g., Blömeke & Delaney, 2012). Hereby, PCK is considered one of the most influential constructs (Depaepe, Verschaffel & Kelchtermans, 2013).

Whereas some researchers could establish a positive relation between measures of mathematics teacher knowledge and indicators of high quality of instruction (e.g., Baumert et al., 2010; Early et al., 2007; Santagata & Lee, 2018; Hill et al., 2008), others argue that the link between knowledge and high-quality instruction is more complex (e.g., Putnam, & Borko, 2000). This relative weakness calls for a critical review of perspectives on teacher knowledge and resonates with certain challenges identified in this field of research.

I) Teacher knowledge cannot be exhaustively described as declarative, decontextualized, propositional knowledge. It encompasses other knowledge sources, which might be inert or of a different quality (e.g., Shulman, 1986; Putnam & Borko, 2000). Hence, research should apply a broader understanding of knowledge, especially considering knowledge of practitioners, e.g., case knowledge.

II) Although teacher knowledge is the basis, teachers' actual decision-making is the crucial point of teacher expertise. Yet knowledge (as exhibited in a teacher knowledge test) may be inert in actual decision-making. Hence, research should account for the usability of knowledge in relevant decision-making processes (e.g., Blömeke et al., 2015; Kersting, Givvin, Thompson, Santagata & Stigler, 2012).

III) Teaching is an act of complex problem solving requiring rational, reasoned decision-making processes and quick, intuitive decisions (e.g., Schön, 1983; Corno & Snow, 1986). Dual processing models of cognition suggest that successful problem solving relies on different memory and reasoning systems according to the vastly different contextual affordances (e.g., Evans, 2008), in short, *slow* and *fast thinking*. Research should hence account for these differences.

Although synthesized from different lines of research, these issues share a common core and address a possible inappropriateness to narrow down teacher expertise to teacher *knowledge*. A recent suggestion to address these challenges complements teacher professional *knowledge* models with a model of teacher professional *competence* without

resigning from a strong cognitive perspective (Knievel et al., 2015; Jeschke et al., 2019). It stands in line with other attempts to extend the understanding of PCK towards a *contextualized* or *professionally situated* (e.g., Manizade & Mason, 2011) teacher knowledge. But unlike others, it uses the concept of professional competence to avoid the further blurring of the PCK construct. Further refinements are based on dual processing theories of cognition. Both points of reference will be outlined briefly in the following.

The notion of competence originated to address discrepancies between tests of isolated traits (e.g., knowledge) and the complex demands of real-life performance (Blömeke et al., 2012). Different understandings evolved, especially whether competence is to be understood as cognitive characteristic (cognitive disposition) of the person causing high performance on criterion tasks or as performance of persons on criterion tasks (McMullan et al., 2003). Following the dispositional perspective, we understand competences as "complex ability constructs that are context-specific, trainable and closely related to real life" (Koeppen et al., 2008, p. 61). In our case, mathematics-specific professional demands of teachers constitute the real-life criterion tasks. Competence hence encompasses different knowledge sources, skills (e.g., noticing, decision-making), and affective-motivational characteristics knit together in a manner demanded by the profession.

Following the approach to model teacher competence with close relation to professional demands, we argue that two vastly different types of demands must be distinguished: 1) pre- and post-instructional demands that allow for analytical, reasoned decision processes, and 2) in-instructional demands, which require teachers to act in-the-moment (Schön, 1983). Dual processing theories of the mind (Evans, 2008) lead to the assumption that mastering these fundamentally different demands requires differently catered cognitive resources. We hence propose to distinguish action-related competence (AC) which refers to the cognitive dispositions to master subject-specific demands in instruction, and reflective competence (RC) which refers to the cognitive dispositions to master subject-specific demands of pre- and post-instructional tasks (Knievel et al., 2015; Lindmeier, 2011). In line with Blömeke et al. (2012), we understand professional knowledge (CK and PCK) as the basis for competence (shortly basic knowledge, BK), but do not further differentiate situation-specific skills. The conceptual model on the level of teachers is displayed in the upper half of Figure 1.
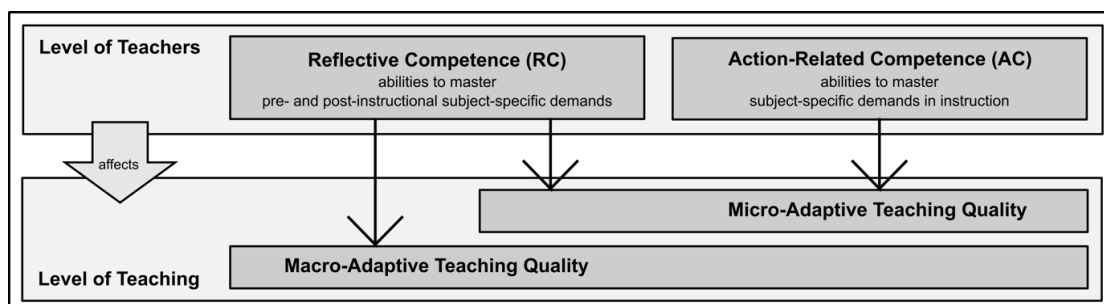

Figure 1 – Conceptual framework of teacher competences (AC, RC) and its hypothesized effects on aspects of teaching quality.

This conceptual framework allows addressing the previously raised challenges as follows: The touchstone of competence is whether teachers are equipped to master the professional demands and not whether they possess (potentially inert) professional knowledge. At the same time and in close continuation of research on teacher knowledge, action-related and reflective competence are understood to ground on professional knowledge. The differentiation between fast and slow demands is fundamental to the model. Comparing in-instructional demands with the demands of planning and post-processing mathematical learning situations, the latter can be considered to be slow processes, distal from the contextual affordances of interacting with children in situations, and relying on analytical thinking processes. Dispositional knowledge is described as main resource for slow, analytical processes in dual processing theories. Hence, one can assume that corresponding reflective competence (RC) should be more strongly intertwined with the professional knowledge base (BK) than action-related competence (AC). In order to master in-instructional demands, quick thinking processes are required. Such processes typically rely on other cognitive structures than dispositional knowledge and allow intuitive decision-making (Evans, 2008).

Professional competence of early childhood teachers

The competence model originates from research on schoolteachers (Lindmeier, 2011). In order to use this approach in the early childhood education context, it is necessary to argue why it makes sense to transfer the model between contexts. Children acquire mathematical knowledge and abilities on numbers and basic operations while in kindergarten (Krajewski & Schneider, 2009). These early mathematical abilities predict later achievement in mathematics (Curby et al., 2009; Krajewski & Schneider, 2009). But in contrast to the school context, mathematical learning opportunities in kindergarten are often connected to play or everyday situations and seldom clearly marked as such (Walsh, Sproule, McGuinness, Trew & Ingram, 2010). This is especially apparent in countries where kindergarten is understood to be part of the early education and care system and hence disparate from school (e.g., Germany: Gasteiger, 2012).

The importance of child-orientated mathematical learning offers in kindergarten is underpinned by research (e.g., Stebler, Vogt, Wolf, Hauser & Rechtsteiner, 2013; Krajewski & Schneider, 2009). Research could also show that teachers play an essential role in planning, offering, and supporting mathematical learning (e.g., Klibanoff et al., 2006; Walsh et al., 2010). In line with the concept of early learning as a social process, mathematical skills develop through interactions of learners and teachers in situations that challenge a child's skills in the zone of proximal development. The teacher hence has to offer apt learning opportunities and support the child's learning, e.g., through scaffolding (van de Pol, Volman & Beishuizen, 2010). Similar to schoolteachers, it is emphasized that early childhood teachers need specific content knowledge and pedagogical content knowledge to succeed at this, e.g., for handling counting errors productively or using apt mathematical language. At the same time, early childhood teachers' education and training are institutionally fragmented

and less standardized in respect to relevant professional knowledge (Gasteiger & Benz, 2018; Walsh et al., 2010). As a result, the professional knowledge for early childhood teachers regarding mathematical learning is difficult to describe. However, in recent years a clear increase in empirical and theoretical knowledge has been generated in key content areas such as numbers and operations (Gasteiger & Benz, 2018; Bruns et al., 2017; Blömeke et al., 2017; Klibanoff et al., 2006).

In light of these findings, we argue that the model of professional competence for teachers can be transferred to the early childhood education context, however, one must take into consideration the particularities of early mathematical learning and the specifics of the context. Hence, we understand action-related competence (AC) of early childhood teachers as the cognitive dispositions to support mathematical learning processes in play or everyday situations. It especially refers to abilities to recognize age-adequate mathematical affordances in situations and to react spontaneously and adaptively to support the learning processes. Reflective competence (RC) of early childhood teachers refers to the cognitive dispositions to plan, prepare, and anticipate mathematical learning opportunities as well as to reflect learning situations. RC enables to systematically use observations from play-based or everyday situations to diagnose and review individual mathematical development and prepare further learning opportunities. Both competence components are based on specific professional knowledge (BK) about the characteristics, affordances, and difficulties of mathematical development in the early years. Professional knowledge is hence an important part of the competences.

Assessing professional competence of early childhood teachers

The presented model of teacher competence is motivated by the fundamental assumption that cognitive resources of teachers are strongly catered according to the professional demands. Thus, the assessment problem for competence boils down to the question of how to implement the professional demands into test instruments under two constraints: Demands should be close to practice so that they elicit cognitive processes according to the theoretical differentiation, and, at the same time, standardized for the use in research, e.g., for the comparison of (groups of) teachers (Shavelson, 2010). Especially the former requirement leads to the fact that typical test rationales used to assess teacher knowledge, for example paper-pencil tests, are not applicable.

The assessment approach for teacher competence following Lindmeier (2011) suggests mirroring closely the key characteristics used to differentiate between the two components of competence which relate to fast and slow professional demands. AC are assessed accordingly in a test with time constraints using video clips of realistic instructional situations as stimuli. This approach led to a high face validity in prior studies and professional knowledge and competence components could be empirically separated for teachers at different educational levels (Hepberger, et al., 2020; Knievel et al., 2015; Jeschke et al., 2019). Prompts for the teachers to act as if they were in this situation as well as capturing the reactions in a naturalistic response format by voice recording, should ensure

that the fast demand takes effect (see AC sample item in the Appendix). In contrast, the assessment of RC aims to stimulate analytical, instruction-related reasoning processes, mirroring pre- and post-instructional professional tasks. Video or written vignettes similarly represent realistic instructional situations. However, prompts and assessment modes differ according to the nature of RC: The teachers are asked to diagnose children's thinking or analyze the affordances of situations or materials. To ensure an engagement in slow cognitive processes, the assessment of RC uses a written response format and no time limitations.

These methical choices aim for a proximal implementation of defining professional demands to enhance the validity of the measures in respect to content and cognitive processes. The methods section provides details on the procedure implemented. In the Appendix, we present sample items together with their scoring rationale in order to illustrate the high coherence between construct, professional demand, and assessment approach.

Quality of learning support as an effect of professional competence

Teachers with higher competences (AC, RC) should be able to provide better mathematical learning situations for children, but studying these relationships requires a clear conceptualization of what "better" means. We will first review the concept of quality of educational processes with a special focus on the quality of mathematical learning support. In the tradition of school effectiveness research, a variety of converging approaches have emerged to describe *dimensions of teaching quality* with varying degrees of focus on mathematics (Praetorius & Charalambous, 2018). For example, the mathematical quality of instruction (MQI) approach describes elements that negatively (e.g., mathematical errors) or positively (e.g., use of math language) affect the quality of mathematics instruction in school (Hill et al., 2008). Often, these approaches clearly show that they were developed in a school context and are sometimes not easily applicable in the early childhood education context (e.g., mathematical errors). Another instrument in this tradition is CLASS (Pianta, La Paro & Hamre, 2008). Its quality dimensions span organizational support (e.g., time and attention management), instructional support (e.g., student's deep engagement with learning materials), and emotional support (e.g., positive teacher-child interactions). The CLASS instrument can be applied across many contexts and proved to be suited for the use in the context of early childhood education, but its information density with respect to the quality of domain-specific learning processes is limited.

Other approaches to model the quality of educational processes are based on the normative ideal of *adaptive teaching*, which has strong links to ATI research (Parsons et al., 2018). Adaptive teaching is described as the effort to align a learning situation with a learner's needs and educational goals. Corno and Snow (1986) introduced a differentiation of adaptation on macro- and micro-level: Whereas adaptation on macro-level refers to the antecedent process of adapting learning materials, social setups, etc., expected to fit learners needs, adaptation on micro-level refers to the sustained efforts to adjust the fit to meet changing needs of learners during the learning process.

We decided to use the concept of adaptation and not dimensions of teaching quality or

similar approaches in order to conceptualize the quality of learning support for our purposes. Three reasons support this decision: First, our mathematics-specific focus calls for a conceptualization of quality that is sensitive for mathematical learning, hence we discarded the use of broad, less specific measure like CLASS. Second, the rationale of mathematical learning in early childhood education contexts follows the idea of individual development in play or everyday situations. This resonates with the theoretical grounds of adaptive learning support. Third, characterizations of adaptation on macro- and micro-level span indicators across subdimensions of teaching quality and adaptation was identified as a cross-cutting concept in teaching quality frameworks (Praetorius & Charalambous, 2018). A conceptualization of quality based on the ideas of adaptation seems in this sense to be ideal to capture the specifics of early mathematical learning in a way that aligns with teaching quality frameworks used in school contexts.

Quality of mathematical learning support as used in this paper to specify our understanding of instructional quality is hence determined by quality on macro-adaptive and micro-adaptive level (short macro- and micro-adaptive quality, see bottom half of Figure 1). *Macro-adaptive quality* is high if a learning situation is aligned to the mathematical abilities of the children involved and their development. High-quality plans of learning situations hence consider possible affordances or constraints of contexts, materials, or situations for the targeted children. *Micro-adaptive quality* is high, if the children experience apt mathematical learning situations. A sustained deep engagement with the materials, positive group interactions, a warm learning atmosphere as well as a good group management are indicative for high micro-adaptive quality underpinning the cross-cutting nature of the adaptive teaching framework.

It is obvious that on the grounds of these definitions teachers' competence should positively affect the quality of learning support. In the field of early childhood education, studies could partly establish relations between teachers' knowledge and situation-specific skills as well as relations between situations-specific skills and quality of instruction (Dunekacke et al., 2016; Bruns et al., 2017). Research that concisely addresses the relations between teacher competences and quality of the learning support is still missing. In detail, mastering demands of planning as well as reflecting on mathematical learning situations (RC) should lead to higher quality on the macro-adaptive level. RC would enable teachers to balance diverse needs of children and the affordances of the available resources during preparation. It would facilitate the analytic, rational differentiation between the intertwined factors of the context, materials, situations, and children's mathematical skills, so that a sound diagnostic basis contributes to a good match between the individual child and the further learning opportunities, hence macro-adaptive quality.

AC should enable teachers to achieve high micro-adaptive quality. It should allow the teacher to act in-the-moment, coordinate the complex processes of sustaining attention and motivation, scaffold children's learning processes, and constantly monitor possible indications of (un)successful mathematical learning, to name a few. Given that such micro-

adaptive learning support, especially scaffolding individual learning processes, is facilitated by high macro-adaptive quality (Charalambous, Hill & Mitchell, 2012), it is further hypothesized that a teacher's RC impacts micro-adaptive quality as well. The argument is based on two mechanisms: First, slow thinking processes in advance of time-critical tasks function as a prime for cognition, so that wrong expectations were found to interfere with performance (Evans, 2008). For example, if a teacher has overestimated a child's skills to deal with a mathematical situation in advance, it can be a challenging task to notice and tackle problems resulting from the (unexpected) mismatch on-the-fly. Second, it can be observed that untypical time-critical situations often result in erroneous decision-making, as fast thinking processes rely on familiar situations. In other contexts, it has been found that slow thinking processes can partly compensate for such difficulties, for example by deliberate mental simulation of possible actions and their consequences (Evans, 2008). For example, if a teacher does not have the required AC to intuitively support a mathematically struggling child, he or she may – in the case of high RC – be able to come up with possible scaffolds in the moment, mentally evaluate their effectiveness against each other and thus compensate for missing AC. However, compensating for missing AC in this way is effortful and limited by time constraints in pedagogical interaction. The hypothesized effects of teacher competences on teaching quality are also expressed in Figure 1. It should be noted that, for simplicity, the figure shows only the hypothesized effects of teacher competence on the quality of teaching. It does not reflect possible loopback effects from the level of teaching on the development of teacher competence over the course of professional life which are beyond the scope of this paper.

Research questions

In the previous sections, we explained how the teacher competence model frames cognitive dispositions to master different professional demands, drawing on the research on teacher expertise and the dual processing theory. We also detailed how teacher competence should relate to instructional quality. Although this approach is strongly grounded from a theoretical point of view, it is still emerging and there is a lack of comprehensive empirical studies that examine the modeled relationships empirically. We address this gap and investigate the presented approach to assess teacher competence (in early childhood education) in respect to different sources of validity evidence, namely aspects of validity evidence in respect to internal structure and the relation to other variables (AERA, APA & NCME, 2014).

The research questions are: Are constructs of competence and teacher knowledge empirically separable in an early education context? Is teacher competence predictive for the quality of learning support?

We hypothesize that early childhood teacher competence can be differentiated according to our theoretical model of action-related and reflective competence with both components grounding on professional knowledge on children's mathematical development

(hypothesis 1). According to the assumed cognitive processes, it is hypothesized that reflective competence shows a stronger connection to professional knowledge than action-related competence (hypothesis 2). We further investigate the relation between teachers' competence and the micro- and macro-adaptive quality of mathematical learning support with the hypothesis that macro-adaptive quality is related to reflective competence and micro-adaptive quality is related to action-related as well as reflective competence (hypothesis 3).

Methods

Study design and sample

We conducted a cross-sectional study to investigate the hypotheses. The study is a secondary analysis of data from a larger project. Early childhood teachers were recruited in Germany (GER) and German-speaking Switzerland (CH) via an open project call. The sample comprised $N = 165$ teachers ($n = 84$ GER, $n = 81$ CH, 7 male, with an average of 13.69 years of professional experience ($SD = 10.1$ years, no significant difference between subsamples from GER and CH: $t(66.6) = 1.6$, $p = .12$). As the structure of teacher education differs in GER and CH, the subsamples differed in respect to the rate of academic graduation (47.7 % in CH, 7.4% in GER). The convenience sample cannot be regarded as representative.

Within the context of the larger study, participants took part in two assessment waves, at the beginning of the new school year and after 6 months. At both measurement points, we assessed the quality of the learning support of the teachers during playing mathematically rich board games (e.g., Schmassmann & Moser Opitz, 2007) as well as with one week delay the professional competence. Given practical constraints, the time difference for single participants was up to four weeks. The larger project offered interventions with focus on professional competences between the points of measurement (see Lindmeier et al., 2020). To answer the research questions in this paper related to structural relationships, we use a cross sectional view and the data from both points of measurements as if they were independent measures as assumptions on the structures should hold across different levels of expertise. In addition, as Praetorius et al. (2014) have shown, instructional quality with focus on subject-specific aspects substantially tends to starkly vary between situations, so that it is recommended to sample across situations, in our case using data from two measuring points half a school year apart. Finally, $N = 155$ and $N = 147$ of the teachers took part at the two waves of measurement. Due to missing values, this results in a total of $N = 302$ cases for the cross-sectional analyses.

Instruments: Professional competence and professional knowledge

We collected data for teachers' competence components RC and AC and administered an additional short test on professional knowledge (basic knowledge, BK), specifically pedagogical content knowledge (as declarative, propositional knowledge). The data was collected by a standardized test, partly in a paper-and-pencil format and partly computer-based. In the following, we give a brief description of the test. The development of the test

ZETETIKÉ

instruments is documented in detail in Hepberger et al. (2020).

All instruments focused on the content of numbers and operations. The content of the items was aligned with empirical findings on the development of numerical skills of children between the ages of 3-6 which contributes to the evidence on test content. Krajewski and Schneider (2009) model the early development with the help of three stages as the increasing ability to deal with quantity to number-word linkages. In stage 1 (basic numerical skills), abilities to discriminate between quantities and skills to deal with number words are still disintegrated. Important abilities are beginning subitizing abilities and to know number words in their sequence. These abilities can be understood as precursors for quantity-number linkages which are indicative of stage 2: Children relate quantities to numbers and number words and are able to determine quantities as well as compare numbers in respect to size. In stage 3, children use relations between quantities and relations between numbers in a sophisticated way, for example when (de)composing numbers.

The AC test had 9 items covering different demands to foster childrens' mathematical learning (sample item in Appendix). All AC test items were video-based and administered under time pressure on a computer. Therefore, short videos (max. 2 min) were shown which presented realistic mathematics-related kindergarten situations, each requiring specific teacher intervention. The teachers had to react directly to the children's activities to adaptively create an opportunity to learn mathematics. They answered by talking into a headset to mirror teacher reaction in a natural way. The computer-based test environment was specifically developed for these requirements (see also Knievel et al., 2015; Jeschke et al., 2019). The RC test had 9 items covering different demands of anticipating and planning, as well as diagnosing children's learning (sample item in Appendix). Four RC items were video-based and also administered on the computer, but without time pressure. The remaining items of the test instrument were paper-pencil based and in short answer or complex multiple-choice format. The teachers were asked to analyze the presented realistic situations for the purpose of planning or diagnosing and to write down their conclusions.

The BK test had 9 items covering knowledge on children's mathematical learning in the area of numbers and operations. It was administered paper-pencil based and in short answer or complex multiple-choice format and can be considered a typical professional knowledge test.

Trained raters coded the open answers according to a code manual. It provided specific criteria for categorizing similar answers in respect to how the answers address the represented mathematical learning demands for each item, for example, whether the answer addresses the skills or struggles shown by the leaner (see the scoring rationales for sample items in the Appendix). The codes were then scored as 1 (full credit), 0.5 (partial credit), or 0 (no credit) depending on the degree they were seen to be indicative of RC or AC. The scoring scheme was discussed and validated by the group of authors, all experts in the field of early education and/or mathematics education working at five different institutions in Germany and Switzerland during the test development (Hepberger et al., 2020).

To estimate the interrater-agreement, two independent trained raters rated 25% of the sample's open answers. Cohen's κ ranged from .60-.96 for the AC items and .58-92 for RC items. The manifest reliabilities of the scales in this study were α = .66 (AC), α = .57 (RC), and α = .62 (BK). A former version of the instrument showed WLE-reliabilities of .73 (AC, 7 items), .77 (RC, 10 items), and .61 (BK, 7 items) in a recent study with different participants of the same target group (Hepberger et al. 2020). The reliabilities can be considered as marginally sufficient, given the complex constructs and the breadth of topics related to children's numerical development in that age.

The scale descriptives and correlations are given in Table 1. The means show that the range of abilities is covered by the test with moderate manifest correlations between scales.

Tabel 1 – Scale descriptives teacher test

Construct	# items	N	M	SD	Min	Max	Correlation to BK	Correlation to RC
AC	9	291	4.58	1.76	0.0	9.0	.34***	.39***
RC	9	258	5.49	1.54	0.0	9.0	.37***	
BK	7	232	4.98	1.98	0.0	9.0		

*** p < .001, **p < .01, * p < .05

Instruments: Quality of learning support

We further applied a measure of quality on mathematical learning situations that were prepared and carried out by the teachers in their own institutions. A board game provided the context of the learning situations. At the first measurement point, teachers were asked to play a specific board game (Schmassmann & Moser Opitz, 2007) which they were given in advance to get familiar with. At the second measurement point, they could choose games from a set that was introduced in the larger project context (see Lindmeier et al., 2020). The teachers selected pre-schoolers of their group and gaming situations of 15 minutes were videotaped. The quality rating is reported in detail in Meier-Wyder et al. (submitted), including an in-depth discussion of its development based on existing measures (Pauli & Reusser, 2011; Pianta et al., 2008; Wullschleger, 2017; Hill et al., 2008), so that the following paragraphs summarize only the necessary information for the use within the scope of this paper. The measure allows distinguishing two dimensions of quality referring to (1) quality aspects with relation to the macro-adaptive level of learning support (4 indicators) and (2) quality aspects with relation to the micro-adaptive level of learning support (4 indicators).

The videotaped gaming situations were used to rate the micro-adaptive dimension. A sample indicator for the micro-adaptive level was *stimulating learning processes*. The quality rating based on the criterion whether the math-related moves of the teachers were related to the children's moves and whether teachers' supportive actions, e.g., explanations or questions, were adapted to the children's mathematical skill level. It hence addressed some aspects of the CLASS-dimension instructional support with a focus on adaptivity. Three more indicators for the micro-adaptive dimension were obtained from the videos through observer rating: *emotional warmth, group management* (both simplified versions of the respective CLASS-dimensions, see Meier-Wyder et al., submitted), and an adapted dimension *use of*

mathematical language (Hill et al., 2008; Klibanoff et al., 2006) to account for subject-specific aspects if micro-adaption.

In contrast to micro-adaptive quality, it is difficult to objectively rate macro-adaptive quality by observation, as it depends on subjective factors like the teacher's information about the target children or the teacher's intended aims. We used a short structured post-interview to infer whether the observed gaming situation was adaptively planned and purposefully used to foster mathematical learning. For example, we asked for the rationale underlying the selection of children, the rationale for (un)expected important moves when playing the game, or observations of future learning support needs. A sample indicator for the macro-adaptive level was *planning of the learning situation*. The quality rating was based on the criterion whether the teachers used information about the children's mathematical ability to purposefully plan the game situation. Three more indicators were obtained from the interviews (*diagnosis of the children's skills*, *reflection of mathematical learning support*, *planning of further learning support*) in order to infer macro-adaptive quality with a larger scope.

The ratings were applied at each measurement point to two sessions of board gaming and related interviews. The teacher played in each of the two sessions with a different small group of 2-3 children which they selected themselves. Gaming videos were divided into 5-minute sampling intervals, the interviews were not split up. Trained raters rated the data in respect to all 8 indicators on a criterial scale ranging from 0 (not observed), 1 (partly observed), 2 (predominantly observed), to 3 (clearly observed). We averaged the ratings across the two situations and in case of the micro-adaptive quality across sampling intervals. Two independent persons double-rated 15% of the data from the first wave. We applied generalizability theory to estimate the influence of the variance sources *situations* and *raters* (Meier-Wyder et al., submitted). The G-coefficients for the macro-adaptive level of learning support ranged from .81 to .90 and for the micro-adaptive learning support from .81 to .95, indicating high interrater-agreement.

The intended scales of the micro- and macro-adaptive quality had reliabilities of $\alpha = .76$ (micro) and $\alpha = .65$ (macro) in our study, so that the reliability of the instrument was good for the micro-adaptive rating and sufficient for the macro-adaptive rating. The scale descriptives are given in Table 2. The quality measures had a theoretical range of 0-12 points, so the instrument could well cover the range of observed quality with a moderate correlation between the two aspects.

Table 2: Scale descriptives quality ratings and manifest correlations between quality aspects and to teacher test scores

Facet	Indicators	M	SD	Min	Max	Manifest correlation to			
						Micro	AC	RC	BK
Macro	4	5.66	2.35	1.00	12.00	0.44***	.25*	.28**	.28**
Micro	4	6.32	1.90	1.33	11.67	–	.11	.11	.22**

*** $p < .001$, ** $p < .01$, * $p < .05$

Data analysis plan

Evidence based on internal structure of teacher measures: Using confirmatory factor analyses (CFA) we investigate whether the theoretical differentiation between the constructs of competence (AC, RC) and professional knowledge (BK) is supported by our data (hypothesis 1). We therefore apply different confirmatory factor models (CFA) and compare their fit. The theoretically derived three-dimensional model (3-dim model AC vs. RC vs. BK, see Figure 2) is contrasted with an alternative one-dimensional model (1-dim model AC+RC+BK). Further, we contrast a two-dimensional model because, according to the dual processing theory, RC and professional knowledge relate to slow cognitive processes, whereas AC relates to fast cognitive processes. If the data mirrors the two types of cognitive processes according to the dual processing theory, RC and BK would be hard to separate empirically and a 2-dimensional model would fit the data (AC vs. RC+BK). The contrasting of the latter models and the comparisons of the relations between AC and BK as well as RC and BK also inform whether RC is indeed more strongly related to BK than AC to BK in line with hypothesis 2.
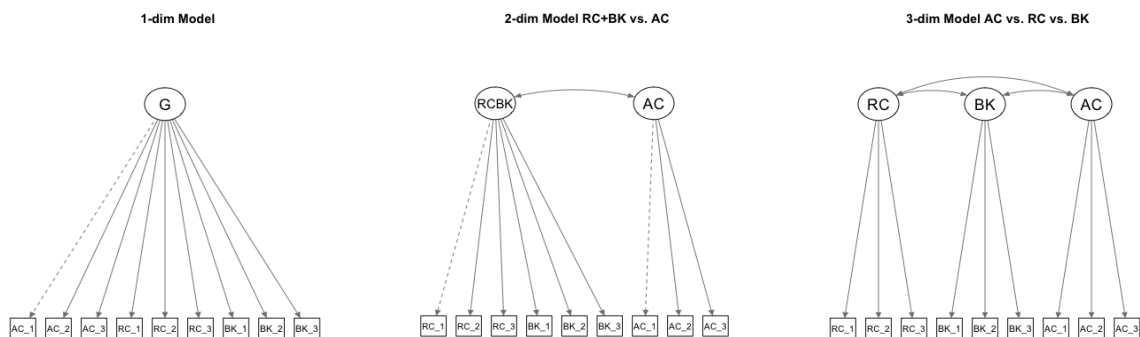

Figure 2: Specified CFA models

Evidence based on the relation to other variables: In the subsequent analyses, we model the hypothesized relations between the components of teacher competence and instructional quality with the means of structural equation modelling (SEM). Therefore, we first specify two models (Figure 3) to separately predict the macro-adaptive and micro-adaptive quality through teacher competence. To acknowledge the common grounds of AC and RC in professional knowledge and in order to investigate in depth the differential impact of AC and RC on micro- and macro-level, we additionally fit a comprehensive path model (SEM 3, Figure 4) that allows mirroring the theoretical relations and simultaneously estimate the hypothesized effects of RC on macro- as well as micro-adaptive and the effect of AC on micro-adaptive quality (hypothesis 3).

Due to restrictions given the sample size, we use item-to-construct balance parceling for the analyses (Little et al., 2002). Therefore, the test items were grouped as parcels of 3 items. We use partly manifest scale scores in addition to parceling to estimate the complex

models SEM 3, so that no model requires more than 24 parameter estimates (ratio of observations per estimated parameter maximal 13:1). CFA and SEM in this paper was computed with the software R (version 4.0.2) and the library lavaan (version 0.6-7, Rosseel, 2012), using the maximum likelihood estimator.

Results

Internal structure of teacher measures

Three nested CFA models were contrasted to investigate the internal structure. The absolute fit-indices chi^2/df, CFI, TLI, and RMSEA (Table 3) of the specified 3-dim model satisfy common cut-off criteria. The less reliable chi^2 fit test criterion is missed (Hu, & Bentler, 1998). The 1- and 2-dim model do not satisfy the cut-off criteria. The comparison of the fit-indices and a chi^2-difference test of the model fits also show that the 3-dim model fits the data significantly better than the 1-dim (test of difference: $chi^2(3) = 58.84$, p < .001) and 2-dim model ($chi^2(2) = 29.47$, p < .001). Hence, the 3-dim model is best suited to model the data what supports our hypothesis 1. The data does support the assumption that AC, RC and knowledge (BK) are empirically separable, yet related constructs.

Table 3: Fit indices of CFA models

	df	chi2 (p)	Chi2/df	CFI	TLI	RMSEA	CI RMSEA	AIC	BIC	adjBIC
1-dim	27	97.58 (.00)	3.6	0.78	0.71	0.11	0.09 - 0.14	4141	4201	4144
2-dim	26	68.21 (.00)	2,6	0.87	0.82	0.09	0.06 - 0.11	4113	4177	4117
3-dim	24	38.74 (.03)	1.6	0.95	0.93	0.05	0.03 - 0.08	4088	4158	4092

RMSEA root mean square error of approximation, CFI comparative fit index, TLI Tucker-Lewis index, AIC Akaike information criterion, BIC Bayesian information criterion, adjBIC sample size adjusted BIC information criterion

The standardized factor loadings in the 3-dim model range from λ = .41 - .77 with all parcels contributing significantly ($p < .001$) to the latent factors. The estimated correlations of the latent factors indicate moderate to strong relations (AC~~BK = .45, RC~~BK = .57, AC~~RC = .71, all $p < .001$). The estimate of the correlation between RC and BK is higher than the one between AC and BK, but the confidence intervals show a considerable overlap so that the difference is not statistically significant. Hence, the in-depth analysis did not support our hypothesis 2. The data does not provide concise evidence that RC is more strongly connected to knowledge (BK) than AC.

Relation between teacher measures and quality of learning support

The manifest correlations (Table 2) between the teacher test scores and the ratings of macro- and micro-adaptive quality indicate that AC and RC are not correlated to micro-adaptive quality and only weakly correlated to macro-adaptive quality. BK shows positive weak to moderate relations with the two aspects of quality, indicating a similar predictive power of the BK for instructional quality as in other studies (e.g., Santagata & Lee, 2019; Hill et al., 2008).

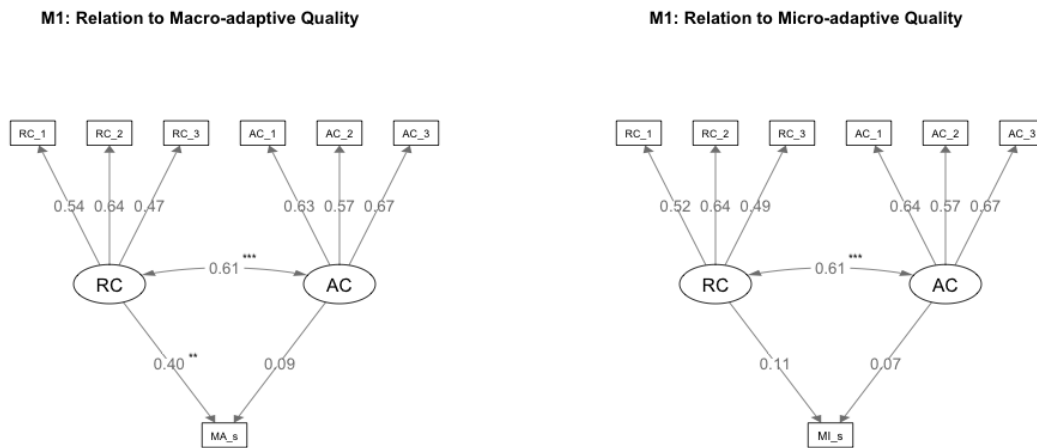**M1: Relation to Macro-adaptive Quality**   **M1: Relation to Micro-adaptive Quality**

Figure 3: SEM models (SEM Macro, SEM Micro) with standardized parameter estimates
(*** p < .001, ** p < .01, * p < .05)

Figure 3 shows the specified models (SEM Macro and SEM Micro) to investigate the relation of the latent constructs of competence to the aspects of quality of learning support separately. The fit indices for both models (Table 4) indicate a satisfactory fit of the models to the data.

Table 4: Fit indices of SEM models

	df	chi2 (p)	Chi2/df	CFI	TLI	RMSEA	CI RMSEA
SEM Macro	12	7.87 (.90)	0.4	1.00	1.02	0.00	0.00-0.04
SEM Micro	12	12.10 (.43)	1.0	1.00	0.99	0.01	0.00-0.06
SEM 3	11	30.88 (.01)	2.8	0.94	0.87	0.08	0.05-0.11

RMSEA root mean square error of approximation, CFI comparative fit index, TLI Tucker-Lewis index, AIC Akaike information criterion, BIC Bayesian information criterion, adjBIC sample size adjusted BIC information criterion

The investigation of the parameter estimates (Figure 4) confirms that AC shows no significant effect on either aspect of quality, although AC and RC correlate moderately. RC, in contrast, has a positive effect on macro-adaptive quality. The models could explain 21% (macro) and 3% (micro) of the variance in quality ratings. The analyses hence confirm one of the three hypothesized relations, namely that RC is positively related to quality of the macro-adaptive learning support (hypothesis 3).

Comprehensive SEM Model Differential Effects

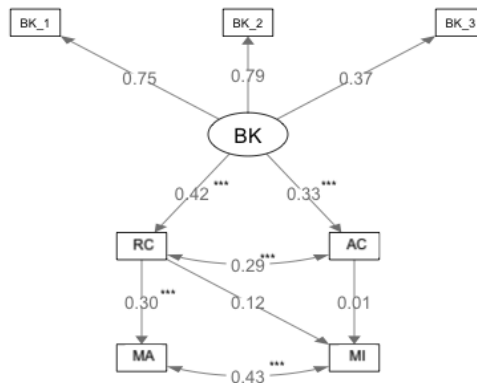

Figure 4: Comprehensive SEM model (SEM3) with standardized parameter estimates
(***p < .001, **p < .01, *p < .05)

Important aspects of the theory (RC and AC rooted in BK, differential effect on and relation between quality aspects) were not reflected in the previous SEM models. The final model (SEM 3, Figure 4) hence aims to mirror these theoretical relations. Sample size restricted the possibilities to model latent factors so that we choose to work with manifest RC and AC scores. The estimated model fit scores partly just fall short of the cut-off criteria (Table 4), so that we still decided to report the parameter estimates to give an informative impression of the data structure (Figure 4). They support the hypothesized positive effect of RC on aspects of macro-adaptive quality (path coefficient .30, $p < .001$) but not micro-adaptive quality (path coefficient .12, n.s.). As the former analyses suggest, AC is not related to quality of the learning support in our data when controlling for BK with a parameter estimate of .01 (n.s.). In summary, although the comprehensive model should be interpreted with caution, this supports the earlier results of the two separate SEMs.

Discussion

This study adds to existing research that uses competence models as an alternative and extension to professional knowledge constructs to model teacher expertise. Teacher competences are hereby seen as complex cognitive dispositions with a close relation to professional demands. In our approach, we apply dual processing theories to distinguish different professional demands of teachers. This reduces the complexity to two fundamentally different contextual affordances, fast and slow demands, mirrored in the constructs of action-related and reflective competence. The study hence develops a specific understanding of competence in line with existing theories. It can be debated, however, to what extent these competence constructs, especially RC, overlap with more situated understandings of PCK.

For example, diagnosing student errors was used as a criterion task for RC in this study and for PCK in other studies (e.g., Hill et al., 2008). AC with its clear reference to in-instructional (fast) cognitive processes clearly exceeds the usual understandings of PCK. Subsequent in-depth analyses would further our understanding of commonalities and differences, but are still hampered by the many – and often elusive – understandings of PCK (Depaepe et al., 2013). It would be interesting to see more researchers incorporating a differentiation according to types of professional demands to gain more clarity in respect to the comparability of constructs.

We argued that the assessment approach chosen shows a high coherence with the underlying constructs and led to the development of innovative methods with a video-, computer-based, and speeded test instrument for AC. The measurement approach was applicable in the early childhood education context, but of course in comparison to paper-pencil tests, the instruments can be seen as costly and show partly lower reliabilities. It is a non-trivial problem to develop video-vignettes that are realistic, short enough, and represent a situation for which clear, research-based reference frames for the scoring are at hand. Still, as the interrater reliabilities in our study show, it is difficult to objectively rate the open answers in oral language. The well-researched developmental model for children's learning of early numbers and operations facilitated the consensus of experts on adequate teacher answers in our study. A precise understanding of the conditions of learning in a domain of interest and a well-established consensus on best practices are important prerequisites for the development of teacher competence tests and at the same time often not available. Hence, our study is limited to the central content of numbers and operations where the research base is sound compared to other content areas, like shape and space, and, yet, there is still room for improving the psychometric quality.

The evidence regarding internal structure indicates that the AC and RC assessment approach captures different constructs. It supports the hypothesis based on the dual processing theory that teachers might indeed use different cognitive resources and different cognitive processes with different (fast and slow) demands, although more sophisticated hypotheses about a closer relationship between RC and BK than AC and BK could not be confirmed by the data. The analyses concerning internal structures are limited in several aspects. First, our instruments' design does not allow us to consider potential mode effects of the speeded (AC) vs. non-speeded (RC) test, but research employing the same assessment approach with primary teachers could repel a mode effect (Knievel et al. 2015). Second, it cannot explain the observed differences beyond describing their relation to professional knowledge. Further research should widen the scope and consider the effect of other factors found to be relevant for teacher expertise, e.g., motivational-affective characteristics and their potentially differential effect on RC and AC. Nevertheless, the study makes an important contribution because research, as well as education and training of teachers, usually neglected these kinds of differences so far despite their relevance.

The empirical evidence in respect to the relation between teacher competence and the

ZETETIKÉ

quality of the learning support does not conclusively support the theoretical assumptions about relations between constructs. AC shows no effect in our study. RC seems in line with our hypotheses relevant for the quality of learning support on the macro-level, but contrary to our expectations not additionally on the micro-level. When interpreting these findings, it is important to consider several limitations beyond the complexities associated with the new measurement approach. First, the sample size for the most complex model is only marginally sufficient, so that it should be read as a heuristic analysis and the findings have to be replicated in larger samples. Second, the education of early childhood teachers in our context is not uniform. Teachers most likely did not study aspects of mathematical learning to the same extent. Our sample size does not allow to differentiate between subgroups of teachers, e.g., according to graduation or country, but related variance might impair our analyses. Third, the cross-sectional analysis was conducted with collapsed data from two measurement points of teachers taking part in an intervention. This diversifies our data in respect to different levels of expertise and different assessment contexts, so that the question of internal and external relationships can be investigated more broadly. But of course, we cannot exclude the possibility that this procedure may affect the results. To get an impression of the stability of our findings for subgroups, we experimentally fitted the specified SEM models to the data of both measurement points separately. Because of the sample sizes, the results should be interpreted with caution, but the findings are in line with those presented above.

Fourth, the quality ratings used in our study rely on situations of playing board games in a small group setting, partly with a specific game. Compared to our competence test instrument which covered different play- and everyday-based situations, the gaming context can be understood as being more specific, which might represent a slight mismatch of test content. On the other hand, our quality ratings had a strong focus on child-teacher interaction, so that the conceptual base (adaptive teaching) can be considered coherent with the teacher competence construct. Our study design does not allow investigating whether the mismatch in respect to content actually hinders fully supporting the theoretical relations. The fact that the teacher knowledge measure (BK) predicts instructional quality in our study with similar strength as in other studies supports the need to look more closely on the alignment of the competence measures regarding content. But surely, further studies are needed to disentangle these aspects.

Our study shows that testing for teacher competence is possible but still difficult. Despite of coherent frameworks, we were only partially successful to empirically trace theoretical relationships with data from a field study. However, in comparison to research on teacher knowledge, the research on teacher competence is still emerging as a whole and specifically in the field of early education. Findings might still be tightly connected to conceptual and methodological decisions in singular studies. It will be important that future research replicates or contrasts findings from different perspectives on teacher expertise and teaching quality.

ZETETIKÉ

Acknowledgements

References

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American educational research journal*, *47*(1), 133-180.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM, 44*(3), 223-247

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

Blömeke, S., Jenßen, L., Grassmann, M., Dunekacke, S., & Wedekind, H. (2017). Process mediates structure: The relation between preschool teacher education and preschool teachers' knowledge. *Journal of Educational Psychology*, *109*(3), 338.

Bruns, J., Eichen, L., & Gasteiger, H. (2017). Mathematics-Related Competence of Early Childhood Teachers Visiting a Continuous Professional Development Course: An Intervention Study. *Mathematics Teacher Education and Development*, *19*(3), 76-93.

Charalambous, C. Y., Hill, H. C., & Mitchell, R. N. (2012). Two negatives don't always make a positive: Exploring how limitations in teacher knowledge and the curriculum contribute to instructional quality. *Journal of Curriculum Studies*, *44*(4), 489-513.

Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. *Handbook of research on teaching*, *3*(605-629).

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., … Barbarin, O. (2009). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. *Early Education and Development*, *20*(2), 346–372.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and teacher education*, *34*, 12-25.

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

Early, D., Maxwell, K., Burchinal, M., Alva, S., Bender, R., Bryant, D., … Zill, N. (2007). Teacher education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78* (2), 558-580.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, *59*, 255-278.

Gasteiger, H. (2012). Fostering early mathematical competencies in natural learning situations – foundation and challenges of a competence-oriented concept of mathematics education in kindergarten. *JMD, 33*(2), 181-201.

Gasteiger, H. (2015). Early mathematics in play situations: Continuity of learning. In *Mathematics and Transition to School* (pp. 255-271). Springer, Singapore.

Gasteiger, H., & Benz, C. (2018). Enhancing and analysing kindergarten teachers' professional knowledge for early mathematics education. *Journal of Mathematical Behavior.*

Hepberger, B., Moser Opitz, E., Heinze, A., & Lindmeier, A. (2020). Entwicklung und Validierung eines Tests zur Erfassung der mathematikspezifischen professionellen Kompetenzen von frühpädagogischen Fachkräften der Elementarstufe. *Psychologie in Unterricht und Erziehung, 67*(2), 81–94. doi:0.2378/peu2019.art24d

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and instruction*, *26*(4), 430-511.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*(4), 424.

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

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. *AERJ*, *49*(3), 568-589.

Klibanoff, R. S., Levine, S. C., Huttenlocher, J., Vasilyeva, M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk". *Developmental psychology, 42* (1), 59-69.

Knievel, I., Lindmeier, A. M., & Heinze, A. (2015). Beyond knowledge. Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education, 13* (2), 309-329.

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modeling and assessment. *Zeitschrift für Psychologie, 216*, 61–73.

Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction, 19*(6), 513-526.

Lindmeier, A., Seemann, S., Kuratli-Geeler, S., Wullschleger, A., Dunekacke, S., Leuchter, M., ... & Heinze, A. (2020). Modelling early childhood teachers' mathematics-specific professional competence and its differential growth through professional development– an aspect of structural validity. *Research in Mathematics Education*, *22*(2), 168-187.

Lindmeier, A. (2011). *Modeling and measuring knowledge and competences of teachers*. Münster, Waxmann.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural equation modeling*, *9*(2), 151-173.

Manizade, A. G., & Mason, M. M. (2011). Using Delphi methodology to design assessments of teachers' pedagogical content knowledge. *Educational Studies in Mathematics*, *76*(2), 183-207.

McMullan, M., Endacott, R., Gray, M. A., Jasper, M., Miller, C. M., Scholes, J., & Webb, C. (2003). Portfolios and assessment of competence: a review of the literature. *Journal of advanced nursing*, *41*(3), 283-294.

Meier-Wyder, A., Wullschleger, A., Lindmeier, A., Leuchter, M., Heinze, A., Vogt, F., & Moser Opitz, E. (submitted). Qualitätsmessung von Lernunterstützung in mathematischen Lernsituationen im Kindergarten – Ein Instrument zur Analyse von Videoaufnahmen.

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., … Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, *88*(2), 205–242.

Pauli, C., & Reusser, K. (2011). Expertise in Swiss mathematics instruction. In *Expertise in Mathematics Instruction* (pp. 85-107). Springer, Boston, MA.

Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In *Mathematical knowledge in teaching* (pp. 9-25). Springer, Dordrecht.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS) manual, pre-K*. Baltimore, MD: Paul H. Brookes Pub. Co.

Praetorius, A. K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, *31*, 2-12.

Praetorius, A.-K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: looking back and looking forward. *ZDM*, *50*(3), 535–553.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning?. *Educational researcher*, *29*(1), 4-15.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Santagata, R., & Lee, J. (2019). Mathematical knowledge for teaching and the mathematical quality of instruction: a study of novice elementary school teachers. *Journal of Mathematics Teacher Education*, 1-28.

Schmassmann, M., & Moser Opitz, E. (2007). Goldstückspiel. In Schmassmann, M., & Moser Opitz, E., *Heilpädagogischer Kommentar zum Schweizer Zahlenbuch 1*. Zug: Klett.

Schön, D.A. (1983/2002). *The Reflective Practitioner: How professionals think in action.* Aldershot: Ashgate.

Shavelson, R. J. (2010). On the measurement of competency. *Empirical research in vocational education and training*, *2*(1), 41-63.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Stebler, R., Vogt, F., Wolf, I., Hauser, B., & Rechtsteiner, K. (2013). Play-based mathematics in kindergarten. A video analysis of children's mathematical behaviour while playing a board game in small groups. *JMD, 34*(2), 149-175.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychological Review, 22*, 271–296.

Walsh, G., Sproule, L., McGuinness, C., Trew, K., & Ingram, G. (2010). Developmentally appropriate practice and play-based pedagogy in early years education. *Retrieved December*, *18*, 2016.

Wullschleger, A. (2017). *Individuell-adaptive Lernunterstützung bei der spielintegrierten Förderung von Mengen-Zahlen-Kompetenzen*. Münster: Waxmann.

Digital Appendix

AC Sample Item "set the table"

Summary: A child (Bruno, approx. 5 years old) sets the table for himself and four puppets. He starts with a wearisome procedure, showing counting skills up to 5. Teachers are instructed to support the child.

Description of the video (0:31 m:s): The video shows the table with four puppets seated and an empty place. The child counts how many places have to be set. 'One-two-three-four-five'. The tableware is in a chest a few steps away. He takes 3 plates and starts setting the table. 'One, two, three. Some are missing!'. He fetches several plates, puts two on the empty places, 'one, two plates', and returns the rest of the plates. He fetches two cups and places them. 'One, two'. He hesitates, looks at the other places appearing frustrated, and turns towards the camera stating 'Many cups are missing!'

Prompt: You encountered this situation and recognized a mathematical learning opportunity. Please give Bruno a stimulating prompt to foster his handling of the situation.

Scoring rationale: Given the exhibited mathematical skills (counting up to 5) and the chosen procedure (taking tableware, filling empty places without counting), we expect teachers to scaffold the child's usage of numbers, e.g., supporting numerical part-whole relations. Accordingly, answers that prompt determining the number of missing items <u>and</u> relate them to the number of set items were seen as indicative of high AC (1 credit, sample answer "Try counting who still needs cups. Then you know how many you need, besides the two already set, to match the 5"). Prompts that focus on a solution strategy within the range of the exhibited skills but not transcending them were seen as indicative of AC, but not completely reaching the aim of providing apt support for mathematical learning (0.5 credits, "If you count who still needs cups, you know how many you need"). Prompts that aim at Bruno repeating his actions ("Just keep going, you can just count how many you set down"), at mathematical strategies out of reach for the child ("You just have to calculate 5 minus 2") or focus on non-mathematical aspects ("You still need cutlery") were seen as not indicative of AC (0 credits).

RC Sample Item „counting up to 40"

Summary: A girl (6 years old) shows characteristic difficulties when counting. The teachers are asked to diagnose the child's skills.

Description of the video (1:16 m:s): A child is seated in front of the camera and counts (up to 40). When crossing 10s, she falters and cannot correctly name the multiples of 10. Off screen, a teacher scaffolds the child when she hesitates at a multiple of 10 by giving the correct name and encouraging: 'What comes next?' The child then autonomously counts on up to the next multiple of 10, but leaves out numbers with repeating digits (22, 33).

Prompt: The child shows two systematic difficulties. What are they? Please describe the difficulties in bullet points.

Scoring rationale: Each correctly identified mathematics-specific difficulty was scored with 0.5 credits, so that a maximum of 1 credit could be reached. Answers were credited 0 if they identified difficulties not shown by the child ("She repeats several numbers"), were not pointing to a math-related issue ("The child struggles with pronunciation", "The child should concentrate better") or were unspecific ("Counting is difficult for her").