International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Effects of biology teachers' professional knowledge and cognitive activation on students' achievement

Christian Förtsch, Sonja Werner, Lena von Kotzebue & Birgit J. Neuhaus

To cite this article: Christian Förtsch, Sonja Werner, Lena von Kotzebue & Birgit J. Neuhaus (2016): Effects of biology teachers' professional knowledge and cognitive activation on students' achievement, International Journal of Science Education, DOI: 10.1080/09500693.2016.1257170

To link to this article: http://dx.doi.org/10.1080/09500693.2016.1257170

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
http://dx.doi.org/10.1080/09500693.2016.1257170

Routledge
Taylor & Francis Group

Effects of biology teachers' professional knowledge and cognitive activation on students' achievement

Christian Förtsch, Sonja Werner, Lena von Kotzebue and Birgit J. Neuhaus

Biology Education, Department I, Faculty of Biology, Ludwig-Maximilians-Universität München, Munich, Germany

ABSTRACT
This study examined the effects of teachers' biology-specific dimensions of professional knowledge – pedagogical content knowledge (PCK) and content knowledge (CK) – and cognitively activating biology instruction, as a feature of instructional quality, on students' learning. The sample comprised 39 German secondary school teachers whose lessons on the topic neurobiology were videotaped twice. Teachers' instruction was coded with regard to cognitive activation using a rating manual. Multilevel path analysis results showed a positive significant effect of cognitive activation on students' learning and an indirect effect of teachers' PCK on students' learning mediated through cognitive activation. These findings highlight the importance of PCK in preservice biology teachers' education. Items of the rating manual may be used to provide exemplars of concrete teaching situations during university seminars for preservice teacher education or professional development initiatives for in-service teachers.

ARTICLE HISTORY
Received 22 February 2016
Accepted 2 November 2016

KEYWORDS
Cognitive activation; pedagogical content knowledge; students' achievement; quantitative research; video study; biology education

Theoretical models of instructional quality (e.g. Helmke, 2014; Kunter et al., 2013a) describe teachers' professional knowledge as a common influence factor on instructional quality, and consequently indirectly on students' achievement. In empirical education research, two research fields prevail, which are based on different paradigms: teacher professionalism research and teaching effectiveness research. Teacher professionalism research is based on the expert paradigm and focuses on identifying several aspects of teacher quality (Bromme, 1997). Baumert and Kunter (2013b) developed a concept of professional competence, which can be applied to teachers' professionalism (Kunter et al., 2013b). This concept includes cognitive as well as non-cognitive aspects. In recent research, the focus was on describing the structure of cognitive aspects, capturing their different dimensions, and relating these dimensions to each other (cf. Gess-Newsome, 1999; Jüttner, Boone, Park, & Neuhaus, 2013; Lee & Luft, 2008; Park & Oliver, 2008; Sczudlek et al., 2016). Research on teaching effectiveness is based on the process–mediation–product paradigm (Brophy, 2000; Shuell, 2001). In contrast to teacher professionalism research, the focus is on effective teaching, which provides opportunities to learn. These opportunities can be utilised by students and, therefore, may lead to students' learning, which again results in higher achievement (Praetorius, Lenske, & Helmke, 2012). Effective

CONTACT Christian Förtsch christian.foertsch@bio.lmu.de

teaching can be described by several features of instructional quality, which influence students' achievement (Brunner et al., 2006; Gröschner, Seidel, & Shavelson, 2013; Hattie, 2009; Slavin, 1994). Several meta-studies summarise effective, mainly general, features of instructional quality (Fraser, Walberg, Welch, & Hattie, 1987; Hattie, 2009; Seidel & Shavelson, 2007). Especially in German-speaking countries, three basic dimensions of instructional quality are often differentiated. Within each basic dimension, single general features of instructional quality are organised into the basic dimensions cognitive activation, classroom management, and supportive climate (e.g. Klieme, Lipowsky, Rakoczy, & Ratzka, 2006). Therefore, the basic dimensions of instructional quality can be described as a cluster of variables, which comprise several content-related features of instructional quality (Klieme, Schümer, & Knoll, 2001). However, in this context, Seidel and Shavelson (2007) and Neuhaus (2007) emphasised the importance of domain specificity for fostering students' achievement. Until now, there are only a few studies which combine both research fields by relating teachers' professional knowledge to their instruction and students' achievement (cf. *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* [COACTIV], Kunter et al., 2013a; *Quality of Instruction in Physics* [QuIP], Fischer, Labudde, Neumann, & Viiri, 2014; *Professional Knowledge of Teachers in Science* [ProwiN, German acronym], Tepner et al., 2012). Therefore, more of such studies are necessary (Gess-Newsome, 2013). The present study addresses this issue by measuring teachers' professional knowledge in biology and relating it to cognitively activating instruction as well as students' achievement.

After introducing the theoretical aspects of teachers' professional knowledge and present conceptualisations of cognitive activation, we address empirical findings concerning the effects of professional knowledge on cognitive activation as well as indirect effects on students' achievement.

Teachers' professional knowledge

Teachers' professional competence includes cognitive aspects, such as professional knowledge, as well as non-cognitive aspects, such as motivational orientations, self-regulation, and beliefs (Baumert & Kunter, 2013b). In recent years, the focus of empirical research has been on analysing teachers' professional knowledge (cf. Fischer et al., 2014; Großschedl, Mahler, Kleickmann, & Harms, 2014; Kunter et al., 2013a; Tepner et al., 2012). Teachers' professional knowledge is divided into several dimensions, whereas three dimensions are established in research: Pedagogical knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK) (cf. Abell, 2007; Kunter et al., 2013a). PK is described as interdisciplinary knowledge about several teaching methods, learning strategies, and classroom management (Voss & Kunter, 2013). In contrast, CK and PCK are domain-specific knowledge dimensions. CK is knowledge about the subject matter and its conceptual understanding (cf. Shulman, 1986). Shulman (1987) described PCK as the melding of CK and PK. Based on this description, researchers often only focused on CK or PK. However, neither PK nor CK is sufficient for successful teaching and therefore, PCK is necessary as an own unique knowledge dimension (Baumert et al., 2010; Gess-Newsome, 2013). Especially for biology teachers' professional knowledge, Großschedl et al. (2014) and Großschedl, Harms, Kleickmann, and Glowinski (2015) indicated that PCK and CK are two separable knowledge dimensions. These two dimensions, however, are highly correlated, from what Großschedl et al. (2015) concluded that CK may be relevant for developing PCK. Additionally while focusing on teachers' CK, it was often measured using distal indicators such as the number of professional development initiatives the teachers had participated in, rather than directly assessing the knowledge through validated test items (Baumert et al., 2010; Hill, Ball, Blunk, Goffney, & Rowan, 2007; Jüttner et al., 2013). In this study, PCK was defined as the knowledge that teachers need to make content accessible to their students (Fischer, Borowski, & Tepner, 2012; Shulman, 1986). Although PCK has been conceptualised in very different ways, at least two components of PCK based on Shulman's (1987) definition have been used consistently: knowledge of instructional strategies and representations as well as knowledge of students' misconceptions (cf. van Driel, Verloop, & de Vos, 1998; Lee & Luft, 2008; Park & Oliver, 2008; Schmelzing et al., 2013). These two components of PCK were also used in different large-scale studies, dealing with measuring teachers' PCK in mathematics and science (COACTIV: Kunter et al., 2013a; *Measuring the professional knowledge of preservice mathematics and science teachers* [KiL, German acronym]: Großschedl et al., 2015; ProwiN: Tepner et al., 2012). Within the mathematical study COACTIV, PCK included the component knowledge of instructional strategies with the two facets knowledge of explanations and knowledge of mathematics tasks, as well as the component knowledge of typical mathematical student errors. Additionally, the component knowledge of curriculum was added (Krauss et al., 2013). Using this conceptualisation of PCK in mathematics, researchers showed that teachers use multiple dimensions of knowledge during instruction and that CK and PCK are separate dimensions of teachers' professional knowledge, which are correlated positively (Bednarz & Proulx, 2009; Krauss et al., 2008). Additionally, results of studies in mathematics indicated that mainly teachers' PCK is a predictor for instructional quality and therefore for students' learning. Teachers' CK, for example, only effected the curricular alignment of tasks (cf. Baumert et al., 2010). Within the KiL project, PCK of biology, chemistry, physics, and mathematics teachers was described with four components. Besides the two components initially described by Shulman (1987), curriculum knowledge and knowledge of assessment were added as additional components of PCK (Großschedl et al., 2015). The project ProwiN defined PCK for biology, chemistry, and physics teachers with the two consistently used components: knowledge of students' errors and knowledge of instructional strategies (Tepner et al., 2012). Within the component knowledge of instructional strategies, they focused on two facets, which are major instructional strategies in all of the three science subjects: model use and use of experiments (cf. Jüttner et al., 2013; Tepner et al., 2012). Even though PCK is operationalised similarly in the different science subjects, the component knowledge of student errors is specific for every subject, as students' errors are based on the specific content which is taught. Additionally, instructional strategies have to be adapted to the specific content of a subject. Therefore, we assume that a sophisticated PCK would result in a teacher having different knowledge of students' errors and instructional strategies in each area of science.

Cognitive activation as a feature of instructional quality

From the view of constructivism, learners construct new organised knowledge actively (Loyens & Gijbels, 2008; Mayer, 2004, 2009). The ICAP (Interactive, Constructive, Active, Passive) framework of Chi and Wylie (2014) proposes different modes of active learning based on students' engagement behaviours. The passive mode is defined as learners who just receive information. Active students are behaviourally acting. For the constructive mode, learners have to generate additional outputs, which are 'beyond what was provided in the learning materials' (Chi & Wylie, 2014, p. 222). In order to be interactive, a discourse or a dialogue between the least constructive learners has to take place. Chi and Wylie (2014) described these four modes hierarchically, with interactive being the highest mode. Additionally, they assumed that higher modes lead to an increase in learning and a deeper understanding of the content. Mayer (2004) already formulated a critique of being only behaviourally active and demanded higher modes. This is in line with one aim of National Education Standards (e.g. KMK, 2005; NRC, 2012), which is to focus on learning core ideas and general principles, resulting in conceptual understanding. In order to reach this goal, we assume that instruction should stimulate students to act at least on the constructive mode. As students have to do more cognitive analysis, a conceptual and deeper understanding of the content will be reached (Chi & Wylie, 2014; Craik & Lockhart, 1972; Klieme, Pauli, & Reusser, 2009; Mayer, 2004). Therefore, in our present study, we focus on teaching practices which stimulate students to do more cognitive analysis in order to gain conceptual understanding. This includes teaching practices which lead to behaviours of students at least on the constructive level of the ICAP framework.

The instructional quality feature cognitive activation deals with this issue of promoting conceptual understanding (cf. Lipowsky et al., 2009). Within the model of instructional quality, Klieme et al. (2006) describe cognitive activation as one of the three basic dimensions of instructional quality. The other two basic dimensions, namely classroom management and supportive climate, are characterised as general features of instructional quality. Cognitive activation, however, is considered as a more domain-specific one. Cognitive activation can only be assessed by taking into account a specific content (Klieme et al., 2009). Praetorius, Pauli, Reusser, Rakoczy, and Klieme (2014) showed in a study that the observation of cognitive activation depends on the stage of the unit and the content taught. However, in recent empirical educational research, cognitive activation has been defined in various different ways, which may be due to its domain specificity. The COACTIV project in mathematics (Kunter et al., 2013a) and the QuIP project in physics (Fischer et al., 2014) only focused on tasks when defining cognitive activation. Within COACTIV, cognitive activation was defined as all learning situations that cause students to conceptually participate in the learning task (Kunter et al., 2007). Although the QuIP project also focused on tasks, they defined cognitive activation differently. They coded the level of complexity of teacher tasks and students' answers and calculated the fit between these two levels as a measure for cognitive activation (Ergönenc, Neumann, & Fischer, 2014). Neumann, Kauertz, and Fischer (2012) extended the definition of cognitive activation; however, their definition is very general. They included all features of instructional quality, which led to students' cognitive activation (e.g. cognitive level of tasks) and to students' engagement. Further definitions provide additional information on what such features of instructional quality could be. Allen et al. (2013) described the CLASS-S (*Classroom Assessment Scoring System – Secondary*), an observation protocol aiming to describe effective instruction using three domains, which are similar to the three basic dimensions of instructional quality (Klieme et al., 2006). Their domain instructional support is equivalent to cognitive activation and assumed to foster students' deep understanding of the content. They described it as the reflection of 'teachers' content understanding, focus on analysis and problem solving, and quality of feedback' (Allen et al., 2013, p. 78). Within the project *Instructional Quality and Mathematical Understanding in Different Cultures*, Klieme et al. (2006) and Lipowsky et al. (2009) described three key features of cognitive activation for mathematics instruction, which foster conceptual understanding: students' cognitive level, conceptual instruction, and thoughtful discourse.

Förtsch, Dorfner, Werner, von Kotzebue, and Neuhaus (2016) showed that it is possible to transfer these three key elements of cognitive activation to biology instruction. The first two key elements (students' cognitive level and conceptual instruction) can be operationalised for biology instruction using the competence model of the project *Evaluation of the National Educational Standards for Natural Sciences at the Lower Secondary Level* [ESNaS, German acronym] (e.g. Kremer et al., 2012; Neumann, Fischer, & Kauertz, 2010). The competence model of ESNaS hierarchically describes levels of expected competences, which were defined in the German National Education Standards inter alia with two dimensions. The dimension cognitive processes describes four cognitive levels, which students need for solving a given task: reproduction, selection, organisation, and integration. The other dimension complexity deals with the number of knowledge elements and their connections, which are needed to solve a task, and is described with mainly three levels: fact(s), relation(s), and generic concepts (Kremer et al., 2012). Both dimensions have already been analysed separately in biology lessons. Jatzwauk, Rumann, and Sandmann (2008) analysed tasks in biology lessons and showed that mainly tasks on a low cognitive level have been used. Wadouh, Liu, Sandmann, and Neuhaus (2014) coded utterances of biology teachers according to their complexity. They found that utterances were mainly on the fact level and, therefore, of low complexity. However, in classes where more relations and concepts were stressed, students showed higher achievement. Förtsch, Dorfner, Baumgartner, Werner, von Kotzebue, and Neuhaus (2016) showed that biology instruction, which includes generic concepts, positively effects students' achievement. The third key element of cognitive activation (thoughtful discourse) is characterised by questions, which stimulate students to 'process and reflect on content, recognise relationships among and implications of its key ideas, think critically about it, and use it in problem solving, decision making or other higher-order applications' (Brophy, 2000, p. 19). The study of Jatzwauk et al. (2008), however, showed that in biology instruction, students are especially asked to give short answers, which do not stimulate them for thoughtful discourse. In summary, it can be assumed that recent biology instruction is mainly on a low level concerning cognitive activation (cf. Förtsch et al., 2016). Based on the definitions of cognitive activation including its three key elements, cognitively activating instruction should lead to students acting on a constructive or interactive level (cf. ICAP framework; Chi & Wylie, 2014).

Students' cognitive activation, however, is not directly observable. Therefore, teaching practices, which are assumed to be cognitively activating (Lipowsky, 2009) or lead to students' acting at least on the constructive level (cf. ICAP framework; Chi & Wylie, 2014), can be used for describing cognitive activation indirectly. Such teaching practices which are assumed to foster students' cognitive activation are, for example, setting challenging tasks, provoking cognitive conflicts, confronting students with contrary ideas or interpretations, linking with prior knowledge, and promoting discourse among students (Förtsch et al., 2016; Klieme et al., 2001; Lipowsky et al., 2009). These practices provide opportunities to learn and encourage students to become activated, do more cognitive analysis, and be constructive (Chi & Wylie, 2014; Helmke, 2014; Kleickmann, 2012; Klieme et al., 2009; Kunter et al., 2013a; Slavin, 1994). However, before it leads to conceptual knowledge or deeper understanding of the content, students have to utilise these opportunities (Klieme et al., 2009; Reusser, 2009). Therefore, in this study, we define cognitive activation as teaching practices that lead to students' behaviours at least on the constructive level (cf. ICAP framework; Chi & Wylie, 2014) and stimulate students to do more cognitive analysis, aiming to foster conceptual understanding.

Effects of domain-specific professional knowledge on cognitive activation and students' achievement

In recent research on mathematics and science education, there are only few indications about the effects of domain-specific professional knowledge on cognitive activation and students' achievement (cf. Cauet, Liepertz, Borowski, & Fischer, 2015; Ergönenc et al., 2014; Kunter et al., 2013b).

In mathematics, the project COACTIV analysed the effect of teachers' domain-specific professional knowledge – mediated by several instructional quality features, among others cognitive activation – on students' achievement (Kunter et al., 2013b). Kunter et al. (2013b), however, analysed instruction indirectly. They operationalised cognitive activation at the task level as the cognitive potential of learning opportunities and used a classification schema to code the tasks (Jordan et al., 2008; Kunter et al., 2013b). Baumert and Kunter (2013a) identified a high correlation between teachers' PCK and CK. They also showed that their PCK had an effect on instructional quality, namely on cognitive activation. CK did not effect cognitive activation. Analysing their data with multilevel structural equation models, Kunter et al. (2013b) showed an indirect effect of teachers' PCK on students' achievement mediated by cognitive activation.

In physics education research, the project QuIP videotaped physics instruction in three countries – Germany, Finland, and Switzerland – and analysed cognitive activation again at the task level (Ergönenc et al., 2014). They analysed teachers' tasks as well as students' answers in terms of the level of complexity and defined cognitive activation as a fit between the two levels. Their results of two-level path models showed a significant direct effect of teachers' PCK on students' achievement. When cognitive activation was added to the model, it was found that teachers' PCK no longer had a significant effect on students' achievement, but had a positive significant influence on cognitive activation (Ergönenc et al., 2014). However, in this study, cognitive activation did not significantly influence students' achievement and teachers' CK was not included in their analyses.

Within a substudy of the physics part of the project ProwiN, Cauet et al. (2015) analysed cognitive activation of videotaped physics instruction using a rating manual, based on Vogelsang and Reinhold's (2013) study, which focused on the whole lesson. Their results from multilevel analyses showed no significant effect of teachers' PCK, nor that of their CK, but a positive significant effect of cognitive activation, on students' achievement. Cauet et al. (2015) did not find a significant correlation between teachers' domain-specific physics knowledge and cognitive activation, but, compared to COACTIV and QuIP, they did not calculate a model where the effect of teachers' domain-specific knowledge on students' achievement is mediated by cognitive activation.

Aims and hypotheses

Our main goal was to identify teacher and instructional variables that directly and indirectly effect students' achievement. From the perspective of teaching effectiveness research, it can be assumed that cognitive activation – as a feature of instructional quality – positively effects students' achievement, at least in mathematics and physics instruction (Cauet et al., 2015; Kunter et al., 2013b). However, transferability of cognitive activation between different contents is limited and there are hardly any studies showing the effectiveness of cognitive activation in biology instruction (cf. Förtsch et al., 2016; Klieme et al., 2009; Praetorius et al., 2012). Additionally, earlier studies from the fields of mathematics and physics education indicate that teachers' PCK and not their CK effects cognitive activation (Cauet et al., 2015; Ergönenc et al., 2014; Kunter et al., 2013b). However, all of these studies show limitations: instruction was observed indirectly (cf. Kunter et al., 2013b), instruction was observed directly but the focus was only on tasks and not on the whole lesson (cf. Ergönenc et al., 2014), or no mediation model was analysed (cf. Cauet et al., 2015). Until now, there have been no such studies in biology education which examine the effects of different dimensions of teachers' professional knowledge on students' achievement, mediated by features of instructional quality. Considering especially the subject-specificity of PCK and its operationalisation by different components and several subject-specific facets, the results from mathematics and physics seem not automatically transferable to biology education.

Therefore, in a first step of our study, we hypothesised that cognitive activation in biology instruction with the specific topic neurobiology has a positive effect on students' achievement (Hypothesis 1) (based on Cauet et al., 2015; Kunter et al., 2013b; Lipowsky et al., 2009).

Bringing together teaching effectiveness research and research on teacher professionalism, we hypothesised that teachers' PCK has an indirect positive effect on students' achievement, mediated through cognitive activation (Hypothesis 2) (based on Ergönenc et al., 2014; Kunter et al., 2013b).

In contrast, teachers' CK should not have an effect on cognitive activation, and, therefore, no indirect effect on students' achievement mediated by cognitive activation (Hypothesis 3) (based on Cauet et al., 2015; Kunter et al., 2013b).

Methods

Design and sample

This present study was embedded in the cooperative project ProwiN jointly conducted by researchers from several universities in Germany (Tepner et al., 2012). This project is a cross-sectional study and analysed effects of teachers' professional knowledge on several features of instructional quality, and effects of the latter, in turn, on students' outcome. Within the biological part of ProwiN, 43 biology teachers from German secondary schools in the state of Bavaria participated. All teachers have been videotaped for two lessons ($N = 85$ videos). For one teacher, we only could videotape one lesson owing to illness. Both lessons were on the topic neurobiology, which contains about 18 lessons according to the Bavarian biology curriculum (Bavarian State Ministry for Education and Culture [StMUK, German acronym], 2004). All teachers were asked to teach the content reflex arc in the first videotaped lesson. For the second lesson, teachers could choose another content within the topic neurobiology. For both lessons, teachers had no guidelines and were not aware of the focus of the study. All of the participating teachers completed their university studies to become a teacher in German secondary schools for the subject biology and a second science subject, such as chemistry or physics. As teacher education on the university level is dictated by the state of Bavaria, all participating teachers had similar preparations concerning their content and pedagogical education. Additionally after university education, all teachers completed practical training in secondary schools for two years (for teacher education in Germany, cf. Bauer et al., 2011; Cortina & Thames, 2013). Before videotaping, the participating teachers completed a test on their professional knowledge. Participating students ($N = 1138$) completed an achievement test before and after videotaping, as well as a questionnaire on motivational aspects (Werner, 2016; Werner, Förtsch, von Kotzebue, & Neuhaus, 2016). This paper presents data from the biological part of ProwiN and, therefore, only focusses on biology teachers' PCK and CK. Especially the effects of teachers' PCK and CK, and cognitive activation on students' achievement were examined.

In the present analysis, we used data from a subsample of 39 teachers (53% female; teaching experience after the traineeship in years: $M = 6.1$, $SD = 5.7$; age in years: $M = 35.6$, $SD = 8.3$) and 78 videos. The reason for this was that we could not collect students' post-test achievement data of four teachers. These teachers, therefore, were not considered in our analyses. The student subsample consisted of 827 students ($M = 21.2$ students per class; 49.7% female; age in years: $M = 14.3$; $SD = .6$).

Teachers' professional knowledge test

Teachers' PCK and CK were assessed by a shortened version of the professional knowledge test developed by Jüttner et al. (2013). The PCK test included 8 open-ended items and 1 multiple-choice item on three PCK facets: *model use*, *use of experiments*, and *student errors*. These three facets were chosen based on the model of Tepner et al. (2012) and cover two important components of PCK: knowledge of instructional strategies and knowledge of students' errors (cf. van Driel et al., 1998; Park & Oliver, 2008; Shulman, 1986). A detailed description of the development of the test including sample item is published in Jüttner and Neuhaus (2012). Items of the CK test (5 multiple-choice items; 7 open-ended items) were from the biology topics cytology and neurobiology (Werner, 2016). To ensure an objective analysis of the tests, two independent markers used a coding manual to code ten percent of both the PCK and CK tests. A high agreement between the two markers has been shown by the results of two-way random intra-class correlations (ICCs) ($ICC_{(absolute)}$) (PCK test: $ICC_{(absolute)} = .96$, $F_{(116,116)} = 24.30$, $p < .001$, $N = 117$; CK test: $ICC_{(absolute)} = .94$, $F_{(119,119)} = 16.63$, $p < .001$, $N = 120$) (Werner, 2016; Wirtz & Caspar, 2002). Both tests were analysed separately using the partial credit Rasch model (PCM; Bond & Fox, 2007). The PCM is a model within the item response theory. It is used for ordinal data and analyses test instruments or questionnaires that aim to measure a single variable using several items (Bond & Fox, 2007; Boone & Scantlebury, 2006). With the PCM, two equal interval measures – item difficulties and person abilities – are computed simultaneously based on the ordinal raw data. Person ability is conceptualised as a latent variable (Boone & Scantlebury, 2006). For evaluating if the raw data fits the PCM, different fit values for items and persons are used. According to Boone, Staver, and Yale (2014), item Outfit-MNSQ (mean-square) values smaller than 1.5 indicate a productive measurement. Additionally, reliabilities for items and persons are calculated separately. High values of item reliability mean that the sample is big enough to measure the latent variable precisely and that the item difficulties range wide enough. The person reliability describes the internal consistency of the measure. A value of .50, for example, discriminates the sample in 1 or 2 levels. Higher values discriminate in more levels (Boone et al., 2014).

The results for both the PCK and CK tests showed satisfactory fit values (all Outfit-MNSQ < 1.5: for person and item reliabilities, see Table 1) (Bond & Fox, 2007; Boone et al., 2014).

Additional information on the PCK and CK tests is reported in Jüttner and Neuhaus (2013a; 2013b), Jüttner et al. (2013), Werner (2016), Werner, Förtsch, von Kotzebue, and Neuhaus (2016).

Rating manual of cognitive activation

We used the rating manual of Förtsch et al. (2016) to analyse videotaped instruction in terms of cognitive activation. The rating manual comprised 37 items, which were grouped into seven content-related sections: (A) *Supporting Knowledge Linking* (5 items), (B) *Exploration of Students' Pre-Knowledge and Conceptions* (5 items), (C) *Exploration of Students' Way of Thinking* (6 items), (D) *Dealing with Students' Conceptions* (6 items), (E) *Teacher as a Mediator* (7 items), (F) *Teachers' Receptive Understanding of Teaching* (3 items; inversely recoded), and (G) *Challenging Learning Opportunities* (5 items) (Förtsch et al., 2016). Every section represents another facet of cognitive activation, and all of the three key elements of cognitive activation as described by Klieme et al. (2006) and Lipowsky et al. (2009) were covered. Sections C, F, and G are related to students' cognitive level; sections A, B, and D are related to conceptual instruction; and sections C and E are related to thoughtful discourse. Additionally, the items of the rating manual were adapted to the subject biology (Förtsch et al., 2016). The items were rated on a three-point Likert scale with the rating options 'not observed' (score = 1), 'partly observed' (score = 2), and 'observed' (score = 3). Each item represented a specific teaching practice that covers a brief moment in instruction, in which students could be cognitively activated. For each rating option, examples were given of how this practice could look like in instruction. The raters watched the whole lesson and were able to take notes. Additionally, they were able to stop the video and rate single items when they observed a specific teaching practice. At the end of a lesson, they made sure that they chose one rating option for each item.

Table 1. Person and item reliabilities of the PCK and CK tests (cf. Werner, 2016; Werner, Förtsch, von Kotzebue, & Neuhaus, 2016)

	Person reliability	Item reliability
PCK	.53	.96
CK	.73	.99

To ensure the quality of the coding, two raters were trained with another sample of videos until an objective measurement was reached (cf. Förtsch et al., 2016). Afterwards these two independent raters coded 10% of the videos. Cohen's kappa value for this instrument was found to be .88, which indicated a substantial inter-rater agreement (Landis & Koch, 1977). In order to obtain equal interval measures and to investigate if the items defined a single construct, the single ratings were analysed with a unidimensional PCM (Bond & Fox, 2007). All items' fit values were in the required ranges: all Outfit-MNSQ < 1.5. The person reliability of .95 and the item reliability of .97 also showed good values (Bond & Fox, 2007; Boone et al., 2014). Afterwards the person abilities for *cognitive activation* for the two videos of each teacher were averaged to obtain one mean value for the construct *cognitive activation* for every teacher. We calculated the mean of teachers' ability in order to assess teachers' ability for cognitively activating teaching of the topic neurobiology, as the students' achievement test was also on the whole topic neurobiology, and therefore a better fit between instruction and students' outcome can be reached. Additional information on the rating manual is reported in Förtsch et al. (2016).

Students' data

Students' achievement was measured before and after videotaping using a paper-and-pencil knowledge test on the topic neurobiology with a focus on the content reflex arc. Both tests included multiple-choice and open partial credit items (18 pretest items; 22 post-test items). The achievement test included items where students had to do more cognitive analysis (be constructive; cf. ICAP framework) and need conceptual understanding in order to solve the item. An example item would be *Explain how the human nervous system is adapted to humans' way of living. Give two examples.* Students completed the pretest in the lesson before the teachers started to teach the topic neurobiology, which contains about 18 lessons according to the Bavarian biology curriculum (StMUK, 2004). When the teacher finished the topic neurobiology, in the next lesson, students completed the post-test. For verifying the objectivity of the coding, two independent test markers coded 10% of each of the pre- and post-test items. Results from two-way random ICCs ($ICC_{(\text{absolute})}$) showed a high agreement between the two markers (pretest: $ICC_{(\text{absolute})} = .99$, $F_{(1277,1277)} = 77.92$, $p < .001$, $N = 1278$; post-test: $ICC_{(\text{absolute})} = .98$, $F_{(2477,2477)} = 56.50$, $p < .001$, $N = 2478$) (Werner, Förtsch, von Kotzebue, & Neuhaus, 2016; Werner, 2016; Wirtz & Caspar, 2002). Both the knowledge pretest and post-test were analysed using the PCM and item thresholds of the post-test were anchored to the item thresholds of the pretest (Bond & Fox, 2007). All items showed good fit values: all Outfit-MNSQ $<$ 1.3. The person reliabilities of 0.63 (pretest) and 0.78 (post-test) and item reliabilities of 1.0 (pretest and post-test) were also satisfactory (Bond & Fox, 2007; Boone et al., 2014; Werner, 2016; Werner, Förtsch, von Kotzebue, & Neuhaus, 2016). Additionally, students completed a questionnaire on motivational aspects (cf. Wild, Gerber, Exeler, & Remy, 2001), including *willingness to make an effort* (3 items; Cronbach's alpha = .72). The questionnaire had a four-point Likert scale with response options ranging from 'strongly disagree' (score = 1) to 'strongly agree' (score = 4). These variables were used as control variables, as it is assumed that motivational variables, among others, *willingness to make an effort* effect students' achievement (e.g. Schmidt-Atzert, 2006). Further information on students' data is reported in Werner (2016) and Werner, Förtsch, von Kotzebue, and Neuhaus (2016).

Analyses

The data had a hierarchical structure. Teachers' professional knowledge, PCK and CK, and *cognitive activation* in their teaching were measured on the class level. Students' *willingness to make an effort* and *achievement in pre- and post-test* were measured on the individual student level. Students' *achievement in post-test* was conceptualised as an outcome variable at the student and class levels. For the variables teachers' PCK and CK, *cognitive activation*, and students' *achievement in pre- and post-test*, we conducted Rasch analyses (PCM; Bond & Fox, 2007) using the software Winsteps 3.81 (Linacre, 2015). The equal interval person abilities were used for all following multilevel analyses. The multilevel analyses were conducted with the program Mplus 7.3 (Muthén & Muthén, 2012). As not all variables were normal distributed, we used the weighted least squares means and variance-adjusted estimator, which leads to similar results as the maximum likelihood estimator (Beauducel & Herzberg, 2006). Concerning the model fit, we used the comparative fit index (*CFI*), the root-mean-square error of approximation (*RMSEA*), and the standardised root-mean-square residual (*SRMR*) – separately for the between- and within-class covariance matrices ($SRMR_{\text{between}}$, $SRMS_{\text{within}}$). According to Hu and Bentler (1998), a good model fit is indicated by *CFI* values above .90, *RMSEA* values below .05, and *SRMS* values below .08. The results of the multilevel analyses were shown as standardised values, so that one unit change represents a standard deviation change in the original (z-standardised). In all multilevel analyses, the variables PCK, CK, and *cognitive activation* were grand-mean centred.

Concerning our research aims, we were interested in the effects on the class level. Therefore, we calculated different multilevel path models. For all analyses, we modelled data on the class and student levels simultaneously. Additionally, we included students' *achievement in pretest* and students' *willingness to make an effort* as control variables on the student level, to interpret differences in students' *achievement in post-test* as effects of class-level variables. Apart from *achievement in pretest*, we used *willingness to make an effort* as a control variable as effort and previous knowledge are known to effect achievement (e.g. Schmidt-Atzert, 2006).

We started with calculating the ICC, where students' achievement in the post-test was modelled on both levels without any predictors, to assess the amount of between-class variance of students' achievement in the post-test (null model).

For testing Hypothesis 1, we calculated a multilevel path model (M1), where cognitive activation effects students' achievement in the post-test.

For testing Hypotheses 2 and 3, we calculated several multilevel path models to analyse an upper level mediation (2 → 2 → 1 mediation; cf. Bauer, Preacher, & Gil, 2006). According to the guidelines described by Baron and Kenny (1986) for single mediation and Preacher and Hayes (2008) and MacKinnon (2008) for upper level mediations, we started with testing the total effects of the independent variables on the dependent variable (path c). These total effects, however, do not have to become significant for mediation to occur (Preacher & Hayes, 2008). In the next step, we tested for effects of the independent variables on the mediator (path a), effects of the mediator on the dependent variable (path b), and direct effects of the independent variables on the dependent variable (path c').

Models M2a, M2b, and M2c analyse the total effects of teachers' PCK and CK on students' achievement in the post-test (path c). In the next step, we calculated model M3, where teachers' PCK and CK effected *cognitive activation* (path a) and *cognitive activation* effected students' achievement in the post-test (path b). Additionally, we added direct effects of teachers' PCK and CK on students' achievement in the post-test (path c').

Results

An overview of all variables on the class and student levels, including means and standard deviations, and intercorrelations are shown in Table 2. For the variables PCK, CK, *cognitive activation*, and students' *achievement in pre- and post-test*, person abilities from the PCM analyses are used. For interpreting person abilities, the corresponding item difficulties have to be considered (Boone et al., 2014). Therefore, negative mean values for PCK, and students' *achievement in pre- and post-test* indicate that the used test was rather too difficult for the persons in our sample. The negative mean value of *cognitive activation* results from the fact that many items of the rating manual were coded as not or only partly observed.

The results from the null model showed an ICC of .192, meaning that 19.2% of the variance in students' achievement in the post-test can be explained by between-class differences.

Effects of cognitive activation on students' achievement in the post-test

To test our first hypothesis, we calculated a two-level path model, where *cognitive activation* predicted students' *achievement in post-test* on the class level. We included the control variables students' *achievement in pretest* and *willingness to make an effort* as predictors for students' *achievement in post-test* on the student level. The model showed a significant effect of *cognitive activation* (β = .39, *SE* = .15, *p* = .012) on students' *achievement*

Table 2. Mean scores and standard deviations and intercorrelations of all class-level and student-level variables

	N	*M*	*SD*	1	2	3	4	5
Class level								
Teacher variables								
1. PCK[a]	39	−0.37	0.52	–				
2. CK[a]	39	0.92	0.45	.15	–			
Instructional variable								
3. Cognitive activation[a]	39	−1.61	1.46	.23	−.13	–		
Student level								
4. Achievement in post-test[a]	827	0.05	0.46	.08*	−.09**	.22***	–	
5. Achievement in pretest[a]	827	−0.43	0.61	−.09**	−.06	.08*	.32***	–
6. Willingness to make an effort[b]	827	3.45	0.52	.01	−.02	.07*	.22***	.10**

[a]Person abilities of variables scaled according to the PCM in Logits
[b]Likert scale ranking from (1) 'strongly disagree' to (4) 'strongly agree'.
*$p < .05$.
**$p < .01$.
***$p < .001$.

in post-test. *Cognitive activation* explained 15% of the variance in students' *achievement in post-test* on the class level ($R^2 = .15$) and the model showed good fit values, $\chi^2(3) = 156.19$, $p < .001$; $CFI = 1.000$, $RMSEA = .000$, $SRMR_{\text{within}} = .000$, $SRMR_{\text{between}} = .000$. These results indicate that a higher amount of cognitive activation in biology instruction leads to an increase in students' achievement.

Effects of teachers' domain-specific professional knowledge on students' achievement mediated by cognitive activation

Addressing our Hypotheses 2 and 3, we first analysed the total effects of teachers' domain-specific knowledge dimensions on students' *achievement in post-test*. Therefore, we calculated two models (see Table 3, M2a–2b), where PCK and CK were tested separately. For each model, we included the control variables students' *achievement in pretest* and *willingness to make an effort* as predictors for students' *achievement in post-test* on the student level. We could not show a total significant effect of teachers' PCK (M2a, $\beta = .23$, $SE = .16$, $p = .158$, $R^2 = .05$) and CK (M2b, $\beta = -.18$, $SE = .28$, $p = .528$; $R^2 = .03$) on students' achievement. In model 2c (see Table 3), we calculated a multiple regression model, where PCK and CK simultaneously predicted students' *achievement in post-test*. Again, none of the knowledge dimensions showed a positive significant effect on students' achievement. The fit values for model M2c showed good values, $\chi^2(4) = 153.75$, $p < .001$; $CFI = 1.000$, $RMSEA = .000$, $SRMR_{\text{within}} = .000$, $SRMR_{\text{between}} = .000$.

Next, we calculated model M3, where instructional quality mediated the effect of teachers' domain-specific knowledge on students' achievement. In this mediation model, PCK and CK predicted *cognitive activation*, which in turn predicted students' *achievement in post-test*. Additionally, direct effects of PCK and CK were added to the model. We again controlled for students' *achievement in pretest* and *willingness to make an effort* at the student level. The mediation model showed good fit values, $\chi^2(8) = 165.50$, $p < .000$; $CFI = 1.000$, $RMSEA = .000$, $SRMR_{\text{within}} = .000$, $SRMR_{\text{between}} = .000$. The results of the upper level mediation model are shown in Figure 1. *Cognitive activation* had a positive significant effect on students' achievement (path b in the mediation model) and was, itself, positively and significantly effected by teachers' PCK, but not by their CK (path a in the mediation model). Teachers' domain-specific knowledge explained 8% of the variance in *cognitive activation*. Additionally, PCK and CK did not significantly correlate with each other ($r = .15$, $SE = .10$, $p = .161$). Consequently, teachers' domain-specific

Table 3. Predicting class-level total effects of teachers' professional knowledge on students' achievement in post-test

Model	Achievement in post-test[a]		
	B	*SE*	R^2
2a			.05
PCK	.23	.16	
2b			.03
CK	−.18	.28	
2c			.10
PCK	.27	.16	
CK	−.22	.25	

[a]The regression models included achievement in the pretest and willingness to make an effort at the student level as control variables.

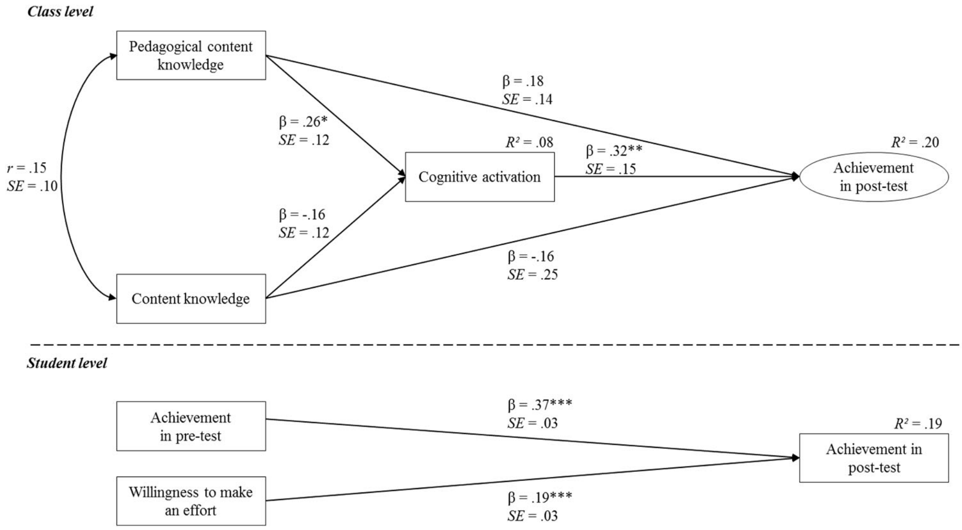

Figure 1. Upper level mediation model (M3). Cognitive activation mediates the effect of teachers' PCK on students' achievement in the post-test. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

knowledge – but only their PCK and not their CK – had an indirect effect, mediated by *cognitive activation*, on students' achievement. Both CK and PCK did not show a direct effect on students' *achievement in post-test* (path c'). Altogether, all variables on the class level explained 20% of variance in students' *achievement in post-test*.

In other words, in those classes where teachers' PCK was high, instruction was high-cognitively activating, which led to students performing better in the achievement post-test. Teachers' CK, however, did not affect the amount of cognitive activation in biology instruction.

On the student level, the two control variables explained 19% of the variance in students' achievement in the post-test, and both had a positive significant effect on students' achievement in the post-test.

Discussion

In our study, we tried to analyse the effects of teachers' domain-specific knowledge and cognitive activation – a feature of instructional quality – on students' achievement. For this purpose, PCK and CK of German biology teachers of secondary schools in the state of Bavaria were measured and their instruction on the topic neurobiology was videotaped. To measure teachers' PCK and CK on the topic neurobiology, we used objective, reliable, and valid paper-and-pencil tests (Jüttner & Neuhaus, 2013a; 2013b; Werner, 2016), which enabled us to measure teachers' knowledge directly. The cognitive activation rating manual also showed a good objectivity and reliability. As we analysed cognitive activation not only by focusing on single aspects – such as tasks (cf. Ergönenc et al., 2014; Kunter et al., 2013b) – but also by rating the whole videotaped biology lesson, we assumed that we could validly measure the construct of cognitive activation (cf. Waldis, Grob, Pauli, & Reusser, 2010).

In the first step, we analysed the effect of the instructional quality feature cognitive activation on students' achievement. Our results showed that cognitive activation had a positive significant effect on students' achievement. This result can be interpreted in a way that cognitively activating biology instruction is advisable for fostering students' achievement. To achieve this positive effect, teachers should focus on supporting knowledge linking by linking current content to already taught content or to future content. Furthermore, teachers should focus on students' conceptions and the active dealing with them. In a concrete teaching situation, this would mean that teachers ask their students about their prior knowledge without targeting a specific answer. This could be made at the beginning of a lesson by asking students to formulate hypotheses to an open question. Additionally, students' conceptions should also be used actively in further instruction. Therefore, in the course of the instruction, teachers can refer back to students' conceptions and discuss with them if they are right or not. A third point on which teachers should focus is the learning opportunities they provide in their instruction. For example, challenging instructional tasks, where content is not only reproduced, should be implemented. Especially, open-ended instructional tasks should be implemented more often and students should be asked to give reasons for their answers (cf. Förtsch et al., 2016).

However, the mean person ability of all participating teachers to teach cognitively activating is negative. It can, therefore, be concluded that many of the items of the rating manual were rated as not or only partly observed. Accordingly, German biology instruction may be described on a rather low level of cognitive activation. Nonetheless, higher cognitively activating instruction seems to be effective in terms of fostering students' achievement. The effectiveness of cognitive activation found in this study is in line with that in empirical studies from the field of mathematics and physics (e.g. Cauet et al., 2015; Kunter et al., 2013b; Lipowsky et al., 2009) and strengthens the assumption that cognitive activation is transferable to biology teaching (cf. Förtsch et al., 2016). With our results in this study, we could confirm our Hypothesis 1, the effectiveness of cognitive activation in biology lessons. Consequently, higher cognitively activating biology instruction leads to higher students' learning and achievement. Our definition of cognitive activation, however, was limited to teaching practices that lead to higher level thinking in terms of the constructive level of the ICAP framework (Chi & Wylie, 2014). This definition is appropriate if the aim of these teaching practices is a conceptual understanding. In contrast, when the aim of instruction is to learn facts, teaching practices that lead to lower cognitive processes may be sufficient. According to the ICAP framework of Chi and Wylie (2014), these processes are called *passive* or *active*. Since studies from biology indicate that tasks are mainly on a low cognitive level (Förtsch, Werner, von Kotzebue, & Neuhaus, 2016; Jatzwauk et al., 2008), actual classroom teaching may more address learning facts. Therefore, the role of cognitive activation might be less when learning lower level objectives as facts. However, several National Education Standards demand conceptual understanding. For example in Germany and in the USA, core ideas were introduced to structure biological content and to promote conceptual understanding (KMK], 2005; NRC, 2012). Accordingly, cognitive activation as examined in our study could help to prevent pure learning facts and, therefore, meet these demands of National Education Standards. However, cognitive activating teaching practices do not necessarily result in cognitive processes for the desired learning outcomes. In accordance with the process–mediation–product paradigm, teaching practices only provide learning opportunities for students. Students, however, are self-responsible for utilising these opportunities (e.g. Klieme et al., 2009; Reusser, 2009). Therefore, cognitively activating teaching practices, which foster higher level thinking, do not guarantee that students learn concepts or general principles. In this context, another question remains. Our conceptualisation of cognitive activation includes several dimensions, for example knowledge linking, dealing with students' conceptions, or challenging learning opportunities. The importance of these single dimensions for students' learning remains unclear. Studies which focused on single dimensions showed partially no effects on students' achievement (e.g. Ewerhardy, Kleickmann, & Möller, 2012; Houtveen, van de Grift, & Creemers, 2004; Widodo & Duit, 2004). This is why further studies, which focus on analysing single dimensions of cognitive activation, are needed to understand the effects on students' learning in a more detailed manner. With this, it could be also possible to clarify if single dimensions or rather an interplay between several dimensions is crucial for learning (cf. Ewerhardy et al., 2012).

In the second step, we calculated a multilevel mediation model, where cognitive activation mediates the effect of teachers' PCK and CK on students' achievement. Additionally, we controlled for students' prior knowledge on the student level – achievement in the pretest – and their willingness to make an effort. In contrast to earlier studies in mathematics or physics (e.g. Kunter et al., 2013b), the results of our study showed that PCK and CK did not correlate significantly. Therefore, we assume that PCK and CK, as gathered in our study, are not related. Thus, we were able to analyse the effect of each subject-specific dimension of teachers' professional knowledge on cognitive activation independently. Our results showed that PCK had a positive significant effect on cognitive activation, which, in turn, still had a significant effect on students' achievement. Teachers' CK, however, did not significantly influence cognitive activation, meaning that for cognitively activating teaching mainly PCK is needed. Based on these results, we could also confirm our Hypotheses 2 and 3. A higher PCK of teachers, therefore, may lead to more cognitively activating instruction, which in turn may lead to higher students' achievement. This indirect effect of PCK on students' achievement, mediated by cognitive activation, is consistent with findings from the COACTIV project in mathematics (Kunter et al., 2013b). Inconsistent findings from physics education studies may be the results of a different conceptualisation of cognitive activation (cf. Ergönenc et al., 2014) or methodological restrictions (cf. Cauet et al., 2015). Ergönenc et al. (2014) defined cognitive activation as the fit between teachers' tasks and their student answers and did not find significant effects on students' achievement. We conceptualised cognitive activation, similar to the COACTIV project (e.g. Kunter et al., 2013b), as learning opportunities offered by the teacher during the whole lesson. Cauet et al. (2015) measured cognitive activation in physics lessons using a rating manual similar to the one we used. They also found positive significant effects of cognitive activation on students' achievement. However, Cauet et al. (2015) could not show correlations between teachers' domain-specific knowledge and cognitive activation. This may be the result of a small sample size ($N = 23$ teachers) or the fact that they did not calculate a multilevel mediation model including teachers' PCK and CK, cognitive activation, and students' variables. Additionally, the conceptualisation of PCK in physics may focus on other facets than we did in biology. Teachers' PCK in a specific subject depends on the content of the subject, as PCK is defined as knowledge which teachers need to make the content accessible to students

(Shulman, 1987). Therefore, physics teachers need to have knowledge about different students' errors than biology teachers. Despite the similarity of the facet knowledge of instructional strategies in science subjects, these strategies have to be adapted to specific contents. These different conceptualisations of PCK could lead to different results. However, the positive effect of teachers' PCK on cognitive activation seems understandable, as the main facets of PCK are knowledge about teaching strategies and representations as well as knowledge about students' misconceptions (cf. van Driel et al., 1998; Park & Oliver, 2008; Shulman, 1986; Tepner et al., 2012). Our conceptualisation of cognitive activation includes mainly different teaching strategies – for example, offering challenging learning opportunities – and dealing with students' conceptions. The result that teachers' CK did not show effects on cognitive activation in their teaching should not be interpreted to mean that their CK is unimportant for cognitively activating instruction on the topic neurobiology or more general biology. We consider PCK and CK as two constructs that are needed for professional actions in effective classroom teaching (cf. Großschedl et al., 2014; Krauss et al., 2008; Riese & Reinhold, 2010). Baumert and Kunter (2013b) argue that CK may be important for developing PCK and thereby offering cognitively activating instruction. Kahan, Cooper, and Bethea (2003) stated that teachers need CK to discern teachable moments and catch them, but it is not a sufficient condition for effective teaching – contrary to PCK (Ball, Lubienski, & Mewborn, 2001). Sczudlek et al. (2016) showed for biology teachers that their CK is a prerequisite for PCK. The professional knowledge test of our study, however, was specifically designed for measuring PCK and CK empirically independent of each other in order to analyse the effects of the dimensions of teachers' professional knowledge independently (cf. Werner, 2016; Werner, Förtsch, von Kotzebue, and Neuhaus, 2016).

In summary, we could show the effectiveness of cognitive activation and that higher PCK of a teacher fostered instructional quality. Therefore, we identified two main aspects that directly or indirectly lead to higher students' achievement. However, our study showed some limitations concerning the generalisability. On the one hand, teachers were not selected randomly, but participated voluntarily in the study. On the other hand, our results only apply to secondary schools in the German state of Bavaria. Further research in different states and school types is needed to draw final conclusions. In addition, it should be noted that this study referred to one specific topic in biology instruction: neurobiology, with a focus on one content, reflex arc. We concentrated on only one topic because this allowed us to closely coordinate all our instruments. Replicating our results in other biological topics would be preferable.

Another question which arises is, what also determines cognitive activation? The domain-specific knowledge dimension PCK explains 8% of the variance in cognitive activation. In our study, we included only two of the three dimensions of teachers' professional knowledge. Therefore, PK – also measured in the framework of ProwiN (Lenske, Thillmann, Wirth, Dicke, & Leutner, 2015) – should be considered in further analyses to investigate the effect of an interdisciplinary knowledge dimension on features of instructional quality. A study of Vogelsang and Reinhold (2013) already indicated that PK may be a predictor of cognitive activation. Additionally, further variables of teachers' professional competence should be taken into account. Besides cognitive aspects of teachers' professional competence, non-cognitive aspects such as beliefs (cf. Baumert & Kunter, 2013b) may be included in future research to capture the influence of teachers' professional competence on biology-specific instructional quality as completely as possible. Additionally, the effects of teacher variables on various other biology-specific features of instructional quality (e.g. Tepner et al., 2012) have to be analysed. One such feature is, for example, *teaching based on core ideas* (Förtsch, Dorfner, Baumgartner, Werner, von Kotzebue, & Neuhaus, 2016; Nachreiner, Spangler, & Neuhaus, 2015), which aims at fostering students' conceptual knowledge as defined in the German National Education Standards in biology (cf. KMK, 2005). *Teaching based on core ideas* considers biology-specific features of instructional quality – including cognitive activation – and general psychological theories (Förtsch, Dorfner, Baumgartner, Werner, von Kotzebue, & Neuhaus, 2016).

The used rating manual for cognitive activation can contribute not only to further research, but also for teacher education. Our results in this study imply that, besides CK, fostering PCK is an important part of preservice biology teachers' education at the university level. Großschedl et al. (2015) showed that the PCK of preservice teachers is positively correlated with their learning opportunities in terms of the period of university studies. Therefore, it seems important that preservice teachers first learn constructs such as cognitive activation theoretically and then apply this feature of instructional quality in practice in order to teach effectively. As this study showed, biology instruction is on a rather low cognitively activating level. But we also showed that higher levels of cognitive activation foster students' achievement. Therefore, the items of our rating manual can provide exemplars of concrete teaching situations which cognitively activate students in their learning. As we described concrete teaching situations for each rating option and therefore also for cognitively activating teaching practices, the items of the rating manual may be helpful for teacher education. Using the ratings from this study can help to identify suitable videos of lessons, which can contribute to a change in teachers' instructional practice, when integrating these videos into teacher education. Additionally, the rating manual can serve as an observation instrument for describing and evaluating biology instruction. Within the project *UNI-Klassen*, classrooms are equipped with video cameras in order to broadcast instruction into an observation room. Preservice teachers, as participants of *UNI-Klassen* seminars at the university thus have the opportunity to use the rating manual to evaluate teaching approaches of fellow students concerning the instructional quality feature cognitive activation. Results from different projects have already indicated that professional development initiatives foster cognitively activating lessons of in-service teachers (e.g. Shayer, 1999; Shayer & Adhami, 2007; Stein & Lane, 1996). Consequently, our rating manual may be also used during professional development initiatives for explaining to in-service biology teachers how to use cognitive activation in their teaching and thus foster their students' learning.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung – BMBF) [Grant number 01JH0904].

References

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1149). Mahwah, NJ: Lawrence Erlbaum Associates.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observation of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system-secondary. *School Psychology Review*, *42*(1), 76–98.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433–456). Washington, DC: American Educational Research Association.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bauer, J., Diercks, U., Retelsdorf, J., Kauper, T., Zimmermann, F., Köller, O., … Prenzel, M. (2011). Spannungsfeld Polyvalenz in der Lehrerbildung [Polyvalence of teacher training programmes]. *Zeitschrift für Erziehungswissenschaft*, *14*(4), 629–649. doi:10.1007/s11618-011-0239-7

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*(2), 142–163. doi:10.1037/1082-989X.11.2.142

Baumert, J., & Kunter, M. (2013a). The effect of content knowledge and pedagogical content knowledge on instructional quality and student achievement. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 175–206). New York, NY: Springer.

Baumert, J., & Kunter, M. (2013b). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–48). New York, NY: Springer.

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. doi:10.3102/0002831209345157

Bavarian State Ministry for Education and Culture [StMUK] (Ed.) (2004). *Lehrplan für das Gymnasium in Bayern* [*Curriculum for the Bavarian Gymnasium*]. Wolnzach: Kastner.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. doi:10.1207/s15328007sem1302_2

Bednarz, N., & Proulx, J. (2009). Knowing and using mathematics in teaching: Conceptual and epistemological clarifications. *For the Learning of Mathematics*, *29*(3), 11–17.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253–269. doi:10.1002/sce.20106

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.

Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers [Competencies, functions, and teaching of the teacher]. In F. E. Weinert (Ed.), *Psychologie des Unterrichts und der Schule* [Psychology of instruction and school] (pp. 177–212). Göttingen: Hogrefe.

Brophy, J. E. (2000). *Teaching*. Brussels: International Academy of Education/International Bureau of Education (IAE).

Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W., … Tsai, Y.-M. (2006). Die professionelle Kompetenz von Mathematiklehrkäften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht: Eine Zwischenbilanz des COACTIV-Projekts [Mathematics teachers' professional competence: Conceptualization, measuring, and importance for teaching: Interim results of the COACTIV project]. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* [Studies on educational quality: Final report of the DFG priority program] (pp. 54–82). Münster: Waxmann.

Cauet, E., Liepertz, S., Borowski, A., & Fischer, H. E. (2015). Does it matter what we measure?: Domain-specific professional knowledge of physics teachers. *Schweizerische Zeitschrift für Bildungswissenschaften*, *37*(3), 462–479.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. doi:10.1080/00461520.2014.965823

Conference of the Ministers of Education [KMK] (2005). *Beschlüsse der Kultusministerkonferenz Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Jahrgangsstufe 10)* [Resolution of the standing conference of the ministers of education and cultural affairs of the Länder in the federal republic of Germany education standards for the subject biology (grade 10)]. München: Luchterhand.

Cortina, K. S., & Thames, M. H. (2013). Teacher education in Germany. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (pp. 49–62). New York, NY: Springer.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, *35*(6), 673–695.

Ergönenc, J., Neumann, K., & Fischer, H. E. (2014). The impact of pedagogical content knowledge on cognitive activation and students learning. In H. E. Fischer, P. Labudde, K. Neumann, & J. Viiri (Eds.), *Quality of instruction in physics* (pp. 145–160). Münster: Waxmann.

Ewerhardy, A., Kleickmann, T., & Möller, K. (2012). Fördert ein konstruktivistisch orientierter naturwissenschaftlicher Sachunterricht mit strukturierenden Anteilen das konzeptuelle Verständnis bei den Lernenden? [Does constructivist teaching with structuring elements in primary science instruction foster students' conceptual understanding?] *Zeitschrift für Grundschulforschung. Bildung im Elementar- und Primarbereich*, *5*(1), 76–88.

Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional knowledge of science teachers. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Springer international handbooks of education: Vol. 24,2. Second international handbook of science education* (pp. 435–448). Berlin: Springer.

Fischer, H. E., Labudde, P., Neumann, K., & Viiri, J. (Eds.). (2014). *Quality of instruction in physics*. Münster: Waxmann.

Förtsch, C., Dorfner, T., Baumgartner, J., Werner, S., von Kotzebue, L., & Neuhaus, B. J. (2016). *Fostering students' conceptual knowledge in biology in the context of German National Education Standards*. Manuscript submitted for publication.

Förtsch, C., Werner, S., Dorfner, T., von Kotzebue, L., & Neuhaus, B. J. (2016). Effects of cognitive activation in biology lessons on students' situational interest and achievement. *Research in Science Education*. doi:10.1007/s11165-016-9517-y

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, *11*(2), 147–252.

Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 3–17). Dordrecht: Kluwer Academic.

Gess-Newsome, J. (2013). Pedagogical content knowledge. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 257–259). New York, NY: Routledge.

Großschedl, J., Harms, U., Kleickmann, T., & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: Structure and learning opportunities. *Journal of Science Teacher Education*, *26*(3), 291–318. doi:10.1007/s10972-015-9423-6

Großschedl, J., Mahler, D., Kleickmann, T., & Harms, U. (2014). Content-related knowledge of biology teachers from secondary schools: Structure and learning opportunities. *International Journal of Science Education*, *36*(14), 2335–2366. doi:10.1080/09500693.2014.923949

Gröschner, A., Seidel, T., & Shavelson, R. J. (2013). Methods for studying teacher and teaching effectiveness. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 240–242). New York, NY: Routledge.

Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses relating to achievement*. London: Routledge.

Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität* [Instructional quality and teachers' professional competence]. Seelze-Velber: Klett/Kallmeyer.

Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M., & Rowan, B. (2007). Validating the ecological assumption: The relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research & Perspective*, *5*(2–3), 107–118. doi:10.1080/15366360701487138

Houtveen, A. A. M., van de Grift, W. J. C. M., & Creemers, B. P. M. (2004). Effective school improvement in mathematics. *School Effectiveness and School Improvement*, *15*(3–4), 337–376. doi:10.1080/09243450512331383242

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Jatzwauk, P., Rumann, S., & Sandmann, A. (2008). Der Einfluss des Aufgabeneinsatzes im Biologieunterricht auf die Lernleistung der Schüler - Ergebnisse einer Videostudie [The effect of tasks in biology education on learning performance – a video study]. *Zeitschrift für Didaktik der Naturwissenschaften*, *14*, 263–283.

Jordan, A., Krauss, S., Löwen, K., Blum, W., Neubrand, M., Brunner, M., … Baumert, J. (2008). Aufgaben im COACTIV-Projekt: Zeugnisse des kognitiven Aktivierungspotentials im deutschen Mathematikunterricht [Tasks in COACTIV: potential to cognitive activate in German mathematics instruction]. *Journal für Mathematik-Didaktik*, *29*(2), 83–107.

Jüttner, M., Boone, W., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, *25*(1), 45–67. doi:10.1007/s11092-013-9157-y

Jüttner, M., & Neuhaus, B. J. (2012). Development of Items for a Pedagogical Content Knowledge Test Based on Empirical Analysis of Pupils' Errors. *International Journal of Science Education*, *34*(7), 1125–1143. doi:10.1080/09500693.2011.606511

Jüttner, M., & Neuhaus, B. J. (2013a). Das Professionswissen von Biologielehrkräften: Ein Vergleich zwischen Biologielehrkräften, Biologen und Pädagogen [Professional knowledge of biology teachers: Comparing biology teachers, biologists, and educationalists]. *Zeitschrift für Didaktik der Naturwissenschaften*, *19*, 31–49.

Jüttner, M., & Neuhaus, B. J. (2013b). Validation of a Paper-and-Pencil Test Instrument Measuring Biology Teachers' Pedagogical Content Knowledge by Using Think-Aloud Interviews. *Journal of Education and Training Studies*, *1*(2), 113–125. doi:10.11114/jets.v1i2.126

Kahan, J., Cooper, D., & Bethea, K. (2003). The role of mathematics teachers' content knowledge in their teaching: A framework for research applied to a study of student teachers. *Journal of Mathematics Teacher Education*, *6*(3), 223–252.

Kleickmann, T. (2012). *Kognitiv aktivieren und inhaltlich strukturieren im naturwissenschaftlichen Sachunterricht* [Cognitively activating and structuring content in science lessons]. Kiel: IPN Universität Kiel.

Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimension und Wirksamkeit von Mathematikunterricht: Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts 'Pythagoras' [Dimension of quality and effectiveness of mathematics instruction: theoretical bases and selected results of the project 'Pythagoras']. In M. Prenzel & L. Allolio-Näcke

(Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* [Research on quality of education in school] (pp. 127–146). Münster: Waxmann.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janík & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster: Waxmann.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: 'Aufgabenkultur' und Unterrichtsgestaltung [Mathematics instruction in secondary education: task culture and instructional processes]. In Bundesministerium für Bildung und Forschung (Ed.), *Timss – Impulse für Schule und Unterricht* [TIMSS – impetus for school and teaching] (pp. 43–57). Bonn: BMBF.

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., … Elsner, J. (2013). Mathematics teachers' domain-specific professional knowledge: Conceptualization and test construction in COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 147–174). New York, NY: Springer.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716–725. doi:10.1037/0022-0663.100.3.716

Kremer, K., Fischer, H. E., Kauertz, A., Mayer, J., Sumfleth, E., & Walpuski, M. (2012). Assessment of standards-based learning outcomes in science education: Perspectives from the German project ESNaS. In S. Bernholt, K. Neumann, & P. Nentwig (Eds.), *Making it tangible. Learning outcomes in science education* (pp. 201–218). Münster: Waxmann.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.). (2013a). *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*. New York, NY: Springer.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013b). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805–820. doi:10.1037/a0032583

Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., … Tsau, Y.-M. (2007). Linking aspects of teacher competence to their instruction: Results from the COACTIV project. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG priority programme* (pp. 39–59). Münster: Waxmann.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education*, *30*(10), 1343–1363. doi:10.1080/09500690802187058

Lenske, G., Thillmann, H., Wirth, J., Dicke, T., & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests [Pedagogical knowledge of teachers: Evaluation of the ProwiN test]. *Zeitschrift für Erziehungswissenschaft*, *18*(1), 225–245. doi:10.1007/s11618-015-0627-5

Linacre, J. M. (2015). *A user's guide to Winsteps/Ministep: Rasch-model computer programs*. Retrieved from http://www.winsteps.com/manuals.htm [01.12.2015]

Lipowsky, F. (2009). Unterricht [Instruction]. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* [Educational psychology] (pp. 73–101). Berlin: Springer.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, *19*(6), 527–537. doi:10.1016/j.learninstruc.2008.11.001

Loyens, S. M. M., & Gijbels, D. (2008). Understanding the effects of constructivist learning environments: Introducing a multi-directional approach. *Instructional Science*, *36*(5–6), 351–357. doi:10.1007/s11251-008-9059-4

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Lawrence Erlbaum Associates.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*(1), 14–19. doi:10.1037/0003-066X.59.1.14

Mayer, R. E. (2009). Constructivism as a theory of learning versus constructivism as a prescription for instruction. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction. Success or failure?* (pp. 184–200). New York, NY: Routledge.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide: Seventh edition*. Los Angeles, CA: Muthén & Muthén.

Nachreiner, K., Spangler, M., & Neuhaus, B. J. (2015). Begründung eines an Basiskonzepten orientierten Unterrichts [Teaching based on biological core ideas. A theoretical foundation]. *Der mathematische und naturwissenschaftliche Unterricht*, *68*(3), 172–177.

National Research Council [NRC] (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press. Retrieved from www.nap.edu/catalog.php?record_id=13165# [16.12.2015]

Neuhaus, B. J. (2007). Unterrichtsqualität als Forschungsfeld für empirische biologiedidaktische Studien [Instructional quality as a research field for empirical studies in biology education]. In D. Krüger & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung* [Theories in biology educational research] (pp. 143–154). Berlin: Springer.

Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education*, *8*(3), 545–563.

Neumann, K., Kauertz, A., & Fischer, H. E. (2012). Quality of instruction in science education. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 247–258). Berlin: Springer.

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*(3), 261–284. doi:10.1007/s11165-007-9049-6

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

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. doi:10.3758/BRM.40.3.879

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. doi:10.1016/j.learninstruc.2013.12.002

Reusser, K. (2009). Unterricht [Instruction]. In S. Andresen, R. Casale, T. Gabriel, R. Horlacher, S. Larcher Klee, & J. Oelkers (Eds.), *Handwörterbuch Erziehungswissenschaft* [Dictionary educational science] (pp. 881–896). Weinheim: Beltz.

Riese, J. & Reinhold, P. (2010). Empirische Erkenntnisse zur Struktur professioneller Handlungskompetenz von angehenden Physiklehrkräften. [Empirical findings regarding the structure of future physics teachers' competence regarding professional action] *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 167–187.

Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the 'cardiovascular system'. *International Journal of Science and Mathematics Education*, *11*(6), 1369–1390. doi:10.1007/s10763-012-9384-6

Schmidt-Atzert, L. (2006). Leistungsrelevante Rahmenbedingungen/Leistungsmotivation [Achievement-related framework conditions/achievement motivation]. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* [*Achievement and achievement diagnostics*] (pp. 224–241). Berlin: Springer.

Sczudlek, M., Kirschner, S., Lenske, G., Tepner, O., Borowski, A., Fischer, H., … Neuhaus, B. J. (2016). *Secondary science teachers' PCK, CK and PK: Their interplay*, Manuscript in preparation.
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. doi:10.3102/0034654307310317
Shayer, M. (1999). Cognitive acceleration through science education II: Its effects and scope. *International Journal of Science Education*, *21*(8), 883–902. doi:10.1080/095006999290345
Shayer, M., & Adhami, M. (2007). Fostering cognitive development through the context of mathematics: Results of the CAME project. *Educational Studies in Mathematics*, *64*(3), 265–291. doi:10.1007/s10649-006-9037-1
Shuell, T. J. (2001). Learning theories and educational paradigms. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioural sciences* (pp. 8613–8620). Amsterdam: Elsevier.
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.
Shulman, L. S. (1987). Knowledge and teaching of the new reform. *Harvard Educational Review*, *57*, 1–22.
Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, *21*(2), 141–157. doi:10.1016/0883-0355(94)90029-9
Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, *2*(1), 50–80.
Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., … Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften [Model for the development of test items measuring science teachers' professional knowledge]. *Zeitschrift für Didaktik der Naturwissenschaften*, *18*, 7–28.
Vogelsang, C., & Reinhold, P. (2013). Gemessene Kompetenz und Unterrichtsqualität: Überprüfung der Validität eine Kompetenztests mit Hilfe der Unterrichtsvideografie [Measured competence and instructional quality: validating a competence test using videos of instruction]. In U. Riegel & K. Macha (Eds.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* [Educational research on competencies with videos] (pp. 319–334). Münster: Waxmann.
Voss, T., & Kunter, M. (2013). Teachers' general pedagogical/psychological knowledge. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 207–228). New York, NY: Springer.
Wadouh, J., Liu, N., Sandmann, A., & Neuhaus, B. J. (2014). The effect of knowledge linking levels in biology lessons upon students' knowledge structure. *International Journal of Science and Mathematics Education*, *12*(1), 25–47. doi:10.1007/s10763-012-9390-8
Waldis, M., Grob, U., Pauli, C., & Reusser, K. (2010). Der Einfluss der Unterrichtsgestaltung auf Fachinteresse und Mathematikleistung [The influence of instruction on interest and mathematics achievement]. In K. Reusser, C. Pauli, & M. Waldis (Eds.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* [Instruction and instructional quality: Results of an international and Swiss video study in mathematics] (pp. 209–251). Münster: Waxmann.
Werner, S. (2016). *Zusammenhänge zwischen dem fachspezifischen Professionswissen einer Lehrkraft, dessen Unterrichtsgestaltung und Schülervariablen am Beispiel eines elaborierten Modelleinsatzes* [Teachers' domain-specific professional knowledge, instructional quality using the example of elaborated model use, and students' variables] (Dissertation). Ludwig-Maximilians-Universität, München.
Werner, S., Förtsch, C., von Kotzebue, L., & Neuhaus, B. J. (2016). *Modeling-based biology teaching to foster students' content knowledge.* Manuscript submitted for publication.

Widodo, A., & Duit, R. (2004). Konstruktivistische Sichtweisen vom Lehren und Lernen und die Praxis des Physikunterrichts [Constructivist perspectives on teaching and learning and the practice of instruction]. *Zeitschrift für Didaktik der Naturwissenschaften*, *10*, 233–255.

Wild, E., Gerber, J., Exeler, J., & Remy, K. (2001). *Dokumentation der Skalen- und Item-Auswahl für den Kinderfragebogen zur Lernmotivation und zum emotionalen Erleben* [Documentation of the scales and items of the questionnaire on motivation and emotional experience]. Germany: Universität Bielefeld.

Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität* [Interrateragreement and interraterreliability]. Göttingen: Hogrefe.