

Does it Matter What We Measure?

Domain-specific Professional Knowledge of Physics Teachers

Eva Cauet, Sven Liepertz, Andreas Borowski and Hans E. Fischer

Can we be confident that extensively validated tests for teachers on their professional knowledge actually measure what matters for effective teaching? This study investigated the relations between physics teachers' domain-specific professional knowledge, students' cognitive activation – as a measure for the quality of instruction in each of the teachers' classrooms – and the achievement of their students using multi-level analysis. Neither teachers' content knowledge (CK) nor their pedagogical content knowledge (PCK) correlated significantly with their support of students' cognitive activation in the classroom; nor did their professional knowledge explain any variance of student learning gains. While these results have to be interpreted carefully for various reasons, they question in particular the validity of the PCK test, which is dealing with content accepted in the community, but normatively set. Moreover, the findings of this study emphasize the importance of connecting professional knowledge to classroom and student variables in order to prove that what tests measure matters for effective teaching.

For more than four decades, professional knowledge of teachers and its different areas have been discussed as a precondition for successful teaching (Peterson, Carpenter & Fennema, 1989; Abell, 2007; Fischer, Borowski & Tepner, 2012), whereas science education only began to be involved in this discussion since the 1990s (Van Driel, Verloop & De Vos, 1998; Gess-Newsome & Lederman, 1999). Professional knowledge is considered as teachers' concepts and competencies required to solve more general pedagogical problems in the classroom, to address adequate and context-dependent teaching and learning issues but also to meet standards for teacher education agreed upon in democratic societies in particular. Regarding standards of teacher education, teacher educators should know which competencies are not only validly tested with samples of student teachers at university or in-service teachers but also relevant for successful teaching and learning, and therefore should be taught in teacher education. From a research point of view, the demand for practical relevance of standards connects teacher education with the classroom and the quality of school instruction.

Recently, standards of teacher education are normatively developed and only partially connected with teaching and learning practices based on the evidence from research findings. For example, the report of the AERA Panel on Research and Teacher Education (Cochran-Smith & Zeichner, 2005) and the programme of the National Academy of Education (Darling-Hammond & Bransford, 2005) or the German standards of the Secretary of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany for teacher education (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2008) defined the content and methods for teacher education. As Herzog (2005) puts it, there are only poor connections between standards and theories; the choice of competencies listed in the standards is characterised as more or less accidental. This is not astonishing because standards in education are designed as governance tools and not directly applicable in classroom settings. However, all these standards are more or less plausible in everyday teaching practice but are not evidence-based from a research perspective (for evidence, see Fischer, Boone & Neumann, 2014). The development of a theory on teachers' professional knowledge for good teaching under classroom conditions can be tracked from Shulman (1987) over Park and Chen (2012) to current approaches. Starting with Shulman's (1987) suggestion that professional knowledge has seven aspects, the common ground in recent research on teacher education (e.g., Baumert et al., 2010) appears in the following dimensions of professional knowledge: content knowledge (CK), pedagogical content knowledge (PCK) and pedagogical knowledge (PK). For example, CK for teaching physics requires «an understanding of physics subject matter as well as research experiences within the discipline», «a knowledge of science in general and of mathematics» and «an understanding of the nature of science, including its history, philosophy, and epistemology at levels that exceed those specified in science education reform documents» (Wenning et al., 2011, p.4). For PCK, the demands are even more vague like «an understanding of the main goal of science education, and an understanding of what it means to be scientifically literate» or «an understanding of the authentic best practices of physics teaching» (p.5). Most of these requirements are examples of the central dilemma of teacher education. Research on CK must clarify which parts of the subject content must be understood for high-quality teaching and what understanding actually means. Therefore, as science educators, we have to answer the following questions: What are the goals of science education and on which notion of scientific literacy and authentic best practices do we all agree? What are effective teaching practices for motivating students learning of biology, chemistry and physics? The question of what content should be taught at universities – to increase the probability of high-quality teaching for future teachers – cannot be answered for most of the content of national standards all over the world because, as the recommendations from Wenning et al. (2011) show, standards are far from being based on empirical research evidence.

As a consequence of the recent research situation, a coherent theory on classroom teaching and learning of the sciences is needed with evidence-based tracks from its elements to teachers' competencies, classroom activities and students' competencies, skills and interest; the most challenging aspect of developing such a theory might be that it has to take into account all possible influences and its elements that survive the demanding evidence of empirical research on these tracks.

From a Normative Set of Professional Knowledge to an Important Variable of Effective Teaching

Standards for professional knowledge are initially formulated as normative settings of institutions (Klauer & Leutner, 2007), and therefore can be taken only as the starting point to measure teachers' competencies (Kauertz et al., 2010). Professional knowledge mostly refers to declarative and theoretical knowledge of teachers – as a result of their teaching experience and their studies at university and teacher training seminars (Clandinin & Connelly, 1995) – but also attitudes, beliefs and emotions are considered as elements of professional knowledge (Barnett & Hodson, 2001; Moallem & Moallem, 1998). To be theoretical and declarative in our first attempt to consider professional knowledge, we have to discuss if there are procedural elements and if the theoretical and declarative elements have some impact on classroom teaching activities (Dann, 1994). Therefore, the test results of teachers on their standard-related professional knowledge should in addition be correlated with lesson activities in the classroom and students' learning outcomes to developed models for instructional quality, and if necessary, to revise standards. Nevertheless, it is still subject to discussion if a relationship between explicable knowledge and teaching actions exists at all (Riese, 2009). To close the gap between standards of teacher education and teaching and learning at school, a measuring model for instructional quality is needed – a model that takes into account not only commonly measured professional knowledge but also its relevance for the activities on the classroom level.

Impact of Professional Knowledge

In Germany, teacher education at university has a strong focus on imparting theoretical knowledge. Therefore, there were different attempts to measure science or mathematics teachers' theoretical professional knowledge with paper-and-pencil-based test instruments (e.g., Baumert et al., 2010; Riese, 2009; Borowski et al., 2010). Those test instruments are often based upon normatively set content widely accepted in the community (e.g., Magnusson, Krajcik

& Borko, 1999). The models usually are validated by using expert ratings, by analysing the correlations between CK, PCK and PK, by referring the results to distal components or by verifying expected differences between different groups (Borowski & Riese, 2010; Dollny, 2011). In a few cases, validation includes comparing results of other models and instruments which target at measuring the same construct (Borowski, Olszewski & Fischer, 2010). But without any connection to students' outcomes, researchers cannot be sure if the tested knowledge is relevant for effective teaching. The analysis of the relationship between professional knowledge, quality of instruction and students' outcome is therefore essential.

There are only a few studies addressing the connection between teachers' knowledge and student achievement for learning mathematics and sciences. In mathematics teaching, the Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy (COACTIV) study (Baumert et al., 2010) showed that PCK had a positive influence on student achievement, which was mediated by the level of cognitive activation in class. Hill, Rowan and Ball (2005) also showed that the construct *content knowledge for teaching mathematics*, which corresponds to CK and PCK, was a significant predictor of student learning gains during the first and third grades. In studying physics teaching at primary school, Lange (2010) and Ohle (2010) were able to find an influence of teachers' PCK and CK, respectively, on student achievement, in both studies, by controlling for additional variables. The Quality of Instruction in Physics (QuIP) project compared physics education in Finland, Germany and Switzerland (Fischer, Labudde, Neumann & Viiri, 2014). For the joint German and Swiss subsample in QuIP, Ergöneç, Neumann and Fischer (2014) found a correlation between physics teachers' PCK and students' cognitive activation. However, the Finnish students in QuIP showed the greatest learning gains (Spoden & Geller, 2014), despite that the Finnish teachers were found to have the lowest level of PCK. This could indicate that an important part of PCK of Finnish teachers – which might be necessary for effective teaching also in other countries – was not addressed.

In general, there is a lack of large-scale studies regarding the empirical investigation of the impact of teachers' professional knowledge on student achievement (Gess-Newsome, 2013). In particular, there are very few studies dealing with physics teaching at secondary schools; and therefore, science educators still do not know what professional knowledge of teachers evidently has an impact on student achievement at secondary schools.

Domain-specific Professional Knowledge in ProwiN Project

The study reported in the following sections was intended to close the gap between theory and practice and to test professional knowledge of teachers and

their cognition-related activities in the classroom. It was part of the second phase of a large project called Professional Knowledge in Science (ProwiN) funded by the German Federal Ministry of Education. The ProwiN project aimed at closing the above-mentioned research gap by analysing the relationship between teachers' CK, PCK and PK, quality of instruction and student achievement and motivation in biology, chemistry and physics lessons. The project referred to the different subjects because *science* is not a common subject in most of the German states. In the first phase of ProwiN, a model for science teachers' professional knowledge was developed in order to quantify and analyse teachers' CK, PCK and PK (Tepner et al., 2012; for physics teaching, see Kirschner, 2013). The ProwiN model focuses on certain facets of each dimension, which are assumed to be important for successful teaching of the three science subjects. Even so, these facets do not reflect the full scope of what can be considered as professional knowledge (Park & Oliver, 2008). The PCK model covers three facets: experiments, concepts and students' preconceptions, whereas the PK model covers four facets: classroom management, teaching methods, individual learning processes and assessment of performance. In physics education, the CK model focuses on school knowledge and advanced school knowledge relating to mechanics. According to the model, paper-and-pencil-based test instruments for physics teachers' CK and PCK were developed and validated by Kirschner (2013); and their construct validities were analysed by using the ProwiN PK test and comparing different dimensional models. It could be shown that CK, PCK and PK are separable dimensions of professional knowledge. As expected, the analyses showed that domain-specific CK and PCK had a higher correlation to each other than to PK. Content validity was ensured by aligning content with curricula and the literature, by consulting with experts and developing a model-based test. To confirm criteria validity, expected differences between well-known groups were analysed. As expected, physics teachers scored significantly higher in the CK and PCK tests than did teachers of other subjects; and student physics teachers at university performed significantly worse than did the more experienced pre-service and in-service physics teachers at school.

In the second phase of the ProwiN project, a video study in all three subjects was conducted in order to analyse if higher professional knowledge leads to better teaching and higher student achievement and motivation.

Research Focus: Does it Matter What We Measure?

This study was an attempt to find out if the domain-specific professional knowledge of physics teachers – measured by intensively validated test instruments such as the ProwiN CK and PCK tests – really matters for effective teaching. The consideration of professional knowledge as being important for

effective teaching implies the assumption that teachers with higher professional knowledge teach better by applying their knowledge in classroom situations, and are therefore, more successful in initiating student learning. Therefore, the relationships between physics teachers' CK and PCK, quality of instruction, and student achievement are taken into account. Following the results of aforementioned studies such as COACTIV (Baumert et al., 2010), we analysed – with respect to quality of instruction – teachers' actions in supporting students' cognitive activation in classroom lessons.

Cognitive activation is regarded as an important dimension of quality of instruction and can be supported by encouraging students to engage in higher-level thinking, by exploring students' prior knowledge and ways of thinking, by dealing with students' preconceptions in an evolutionary way and by avoiding the use of transmissive teaching methods (Lipowsky et al., 2009). In this attempt, cognitive activation is not necessarily expected as staying stable across different lessons. Based on theoretical considerations, researchers argue that the investigation of cognitive activation should be restricted to introductory lessons (Praetorius, Pauli, Reusser, Rakoczy & Klieme, 2014). As a consequence, if the ProwiN CK and PCK tests measure knowledge which is relevant for effective teaching, we would expect significant correlations between CK and PCK and the teachers' ability to cognitively activate their students when introducing a new concept. Teachers with higher CK should be able to create more challenging learning opportunities for their students in order to engage them in higher-level thinking. Correlations of PCK to cognitive activation in the classroom should be even higher because teachers with higher PCK should additionally be more interested in exploring students' prior knowledge and ways of thinking. Moreover, knowledge about students' preconceptions should enable teachers to better deal with the preconceptions. Therefore, teachers' CK and PCK should also be significant predictors of student knowledge acquirement at the end of a physics unit.

Design and Methodology

This study took place in the federal state of North Rhine-Westphalia, Germany. The professional knowledge of 23 physics teachers (35% female, $M_{Age}=44$ years, $SD_{Age}=12$ years) – teaching students in grades eight and nine at grammar school (Gymnasium) (Bonsen, Bos & Frey, 2008) – was gathered. For each of the participating teachers, the introductory lesson on the force concept within a course in mechanics was videotaped in order to analyse cognitive activation. Dependent on the school, lesson length ranged between 45 to 90 minutes. Because student content knowledge is used as a criterion for effective teaching in this study, teachers were asked to plan the lesson with concept development as primary learning goal. Student achievement was measured before and after

the teaching of the whole mechanics unit; 610 of the 660 participating students (56% female, $M_{Age}=14$ years, $SD_{Age}=1$ year) took the student content knowledge pretest and posttest. The teachers were asked to report on how many lessons they had taught within the mechanics unit. The number of lessons ranged from 12 to 59 (standardised to 45 minute per lesson). *Lesson time* that we use in this report is equal to the number of lessons multiplied by the length of a lesson in minutes. Data were also gathered on students' cognitive abilities and migration background which were considered as additional control variables. While 21% of the students had a migration background, this percentage varied among the classes from 0% to 46%.

Teacher and Student Tests

Physics teachers' professional knowledge was measured using a paper-and-pencil test developed and validated in ProwiN I (Kirschner, 2013). The CK test consisted of 12 items and the PCK test consisted of 11 items; and in both tests, open and multiple-choice questions were used. The content of the CK test was mainly related to classical mechanics. The PCK test additionally referred to teaching physics in general and covered the facets already described above: experiments, concepts and students' preconceptions. All the teachers' responses to the CK and PCK items were coded by two raters with good to very good inter-rater agreement (CK: $ICC_{2-fact, unjust} \geq .96$; PCK: $ICC_{2-fact, unjust} \geq .85$, except for one item with $ICC_{2-fact, unjust} = .77$) (Wirtz & Caspar, 2002). Teachers' CK and PCK were estimated in separate analyses within a Rasch model using Joint Maximum Likelihood Estimation. As the teacher sample was too small for Rasch analysis, data collected in a previous study (ProwiN I: $N = 79$, 37% female, $M_{Age}=44$ years, $SD_{Age}=10$ years) was included to enlarge the sample. Rasch person measures for teachers' professional knowledge were then calculated using the whole sample of 102 teachers. One CK item and one PCK item had to be rejected because of their significant misfit to the Rasch model ($MNSQ > 1.2$, $ZSTD > 2$; Bond & Fox, 2007).

Student content knowledge (SCK) was measured with a multiple-choice test in a multi-matrix pre-post design with two anchored test booklets, each of which contained 24 items including nine anchor items common to the two booklets (i.e., there were 39 different items in total). The test that was self-developed covered different subtopics of mechanics and included items from Trends in International Mathematics and Science Study (TIMSS) Assessment (Olson, Martin & Mullis, 2008), the Force Concept Inventory (Hestenes, Wells & Swackhamer, 1992) and the Mechanics Baseline Test (Hestenes & Wells, 1992). The SCK test was administered to the students before and after the teaching of the unit on mechanics. Both test booklets were used at the pre- and post-measurement points. At the posttest, students used the test booklet they had not

worked on at the pretest. The development and validation of the SCK test was described in Cauet (2015). The SCK test showed good construct and criteria validities. Although curricular validity could only be found using the statewide curriculum but not with respect to what was actually taught in each individual class, validation results showed that a fair comparison of students' content knowledge between different classes is possible if lesson time is controlled for. Students' content knowledge was estimated within a Rasch model in separate analyses for the pretest and the posttest. One SCK item had to be rejected in all analyses because of a misleading task formulation. In the posttest analysis, four other items had to be rejected because of significant misfit to the Rasch model ($MNSQ > 1.2$, $ZSTD > 2$) (Bond & Fox, 2007). Therefore, only 34 of the original 39 SCK items were analysed as shown in Table 1. In order to calculate students' pretest and posttest measures on the same scale, the posttest item measures were fixed in the pretest analysis. Items which did not fit in the posttest were also rejected in the pretest. Although there were then some misfitting items in the pretest, no items were rejected because this misfit was most likely a result of fixing the item measures (Linacre, 2011).

Lesson time was estimated based on the teachers' report on the number of lessons taught within the unit. Students' migration background was operationalised by students' home language (Quesel, Möser & Husfeldt, 2014). Students' cognitive abilities were surveyed by the subscale N2 (A) of the Cognitive Abilities Test (CAT) (Heller & Perleth, 2000). There were different test booklets for grade eight and grade nine students, each of which contained 25 items including 20 anchor items, that is, there were 30 different CAT items in total (see Table 1). Cognitive abilities are estimated within a Rasch model, too. With reference to the CAT-Manual, no item of the cognitive abilities test was rejected (Heller & Perleth, 2000), although there was some significant misfit to the Rasch model.

In interpreting the reported Rasch person reliabilities as shown in Table 1, researchers have to consider that those values are often lower than Cronbach's Alpha values (Linacre, 2011). The low SCK_{pre} person reliability values can be explained by the fact that student knowledge in the pretest is much less structured and that students are more likely to guess than in the posttest. All in all, the content of the student test covers the quite heterogeneous construct «mechanics» which consists of several subtopics. Moreover, the multi-matrix design of the tests could also cause a reduction of person reliability (Linacre, 2011).

Table 1: Reliabilities of Teacher and Student Instruments

	Teacher Instruments		Student Instruments		
	CK	PCK	SCK _{Pre}	SCK _{Post}	CAT
N _{Persons}	102	102	640	630	640
N _{Items}	11	10	34	34	30
Person Reliability	0.73	0.59	0.51	0.61	0.84
Item Reliability	0.97	0.96	0.99	0.99	0.99

Analysing Cognitive Activation in Classroom Lessons

Teacher actions supporting students' cognitive activation (CA) in the lessons were rated on a 3-point Likert scale (1= «disagree», 2= «partly agree» and 3= «agree») using an adapted rating instrument from Vogelsang (2014) which was based on rating scales from Racozy and Pauli (2006), Kunter (2005), Widodo and Duit (2004), Clausen (2002) and Clausen, Reusser and Klieme (2003) (as cited in Vogelsang, 2014). The CA rating consisted of 39 indicators which can be attributed to seven subscales: *Making students aware of the learning status* (5 indicators), *exploration of prior knowledge* (5 indicators), *exploration of students' ways of thinking* (6 indicators), *evolutionary dealing with students' preconceptions* (6 indicators), *mediation* (8 indicators), *receptive teaching* (4 indicators) and *challenging learning opportunities* (5 indicators). Four indicators, which could not be rated in more than 20% of the lessons, were rejected. For two other indicators with missing values in just one lesson, those values were imputed by the mean value of the indicator. Four indicators had to be rejected because of low discriminatory power ($<.2$). The subscale *Making students aware of the learning status* with 5 indicators was rejected because it was not reliable ($\alpha_C=.13$). The reliability of the 25 remaining indicators and the reliabilities of the remaining subscales are reported in Table 2. Although raters were trained for 2 months, the inter-rater agreement was not satisfactory. Therefore, three raters independently rated each lesson ($.19 < ICC_{(2,1)unjust} < .58$ on subscale level for single measures) and shortly afterwards, the ratings for each indicator were discussed among the raters until consensus was reached. For each lesson, the CA score was calculated as the average rating over all indicators (Vogelsang, 2014). In order to ensure that the CA scores are valid indicators for teachers' quality of instruction, these scores were analysed to find out if they can explain a significant amount of the between-class variance in students' content knowledge in the posttest measures. The corresponding results are reported in the next section.

Table 2: Reliability of the Cognitive Activation Rating Instrument

Subscales of Cognitive Activation	N _{Lessons}	N _{Indicators}	α_C
Exploration of Prior Knowledge	23	4	.71
Exploration of Students' Ways of Thinking	23	4	.79
Evolutionary Dealing with Students' Preconceptions	23	4	.73
Mediation	23	6	.75
Receptive Teaching	23	3	.52
Challenging Learning Opportunities	23	4	.76
Cognitive Activation (Overall scale)	23	25	.93

Learning Gains

Pretest-posttest comparison of students' content knowledge revealed a significant but small learning gain during the mechanics unit on the student level, $t(610) = 10.50$, $p < .001$, $d = .43$. These results are in line with the results from other studies such as PISA 2003 (Prenzel et al., 2006) or the QuIP project (Fischer et al., 2014) in which similar results were found regarding learning gains in learning physics in German classes. Pant et al. (2013) pointed out that within Germany, the students in North Rhine-Westphalia – where this study was conducted – even learned less than did the students in other Federal German States. The results of one-tailed paired t -tests carried out separately for each class showed that there were nine classes (39% of the sample) without significant learning gains. By calculating the ICC(1,1) (Shrout & Fleiss, 1979), the proportion of variance in the posttest caused by differences between classes can be quantified. For SCK_{POST} , 10.4% of the variance was due to the between-class differences.

Findings from Multilevel Analyses

In order to find out if the between-class variance in the posttest can be explained by teachers' CK or PCK, multilevel analyses were conducted. The baseline model (Model 1) included only the control variables. On the student level, students' pretest content knowledge, cognitive abilities and migration background were included as predictors. While these variables explained 31% ($SE = .03$, $p < .001$) of the within-class variance, controlling for them reduced the between-class variance to 5.3%. On the class level, lesson time was included as a predictor in the model and explained 63% ($SE = .15$, $p < .001$) of the between-class variance in the posttest. In Models 2 and 3, teachers' CK and PCK were included as additional predictors on the class level. Neither CK nor PCK could explain a significant additional amount of the posttest variance between the classes (Model

2: $\beta_{\text{Stand.}}=-.14$, $SE=.19$, $p=.474$; $R^2=65\%$, $SE=.13$, $p<.001$; Model 3: $\beta_{\text{Stand.}}=-.19$, $SE=.14$, $p=.183$; $R^2=67\%$, $SE=.16$, $p<.001$). Moreover, physics teachers' PCK showed a tendency to have a negative impact on students' content knowledge in the posttest. In Model 4, the score for cognitive activation was included as a predictor on the class level. Compared to the baseline model, the CA score explained an additional amount of 15% of the between-class variance in the posttest ($\beta_{\text{Stand.}}=.40$, $SE=.19$, $p<.05$; $R^2=78\%$, $SE=.13$, $p<.001$). Therefore, the CA score can be interpreted as a valid indicator for quality of instruction.

Relationship between teachers' CK and PCK and Cognitive Activation in Classroom Lessons

There were no significant correlations between teachers' CK and PCK and cognitive activation in the analysed lessons (see Table 3). It has to be noted, however, that for the analysed sample size only large effects could become significant on a 5%-level. The correlation between CK and cognitive activation becomes at least significant when one-tailed testing is used. This correlation is predominantly caused by a significant correlation between CK and the CA subscale *Challenging Learning Opportunities* ($r=.57$, $SE=.12$, $p<.01$).

Table 3: Pearson's Correlations between Teachers' Domain-specific Professional Knowledge and Cognitive Activation in Classroom Lessons

N=23	CK	PCK
	r	.15
Cognitive Activation in Classroom Lessons	SE	.20
	p	0.504

Discussion

This study was an attempt to relate teachers' domain-specific professional knowledge – measured with the ProwiN I test instruments – to quality of instruction and student achievement. When controlling for students' cognitive abilities, migration background and length of the unit, we could not find a significant relationship between physics teachers' CK and PCK and their support of students' cognitive activation in the classroom or student learning gains as result of a teaching sequence on mechanics in grades 8 and 9 at secondary schools. However, the reported results need to be carefully interpreted because they were generated from a rather small sample of 23 physics teachers. The aim of the ProwiN project to analyse a sample of 40 physics teachers could not be

realised because only this small sample of teachers agreed to participate in a video study. After two years of data collection for this study, data of 12 more physics teachers and their classes was collected in another subproject of ProwiN, which focused on content structure as a different dimension of quality of instruction. The data of the whole sample of 35 physics teachers has not yet been analysed, but results for the relationship between teachers' CK and PCK, and content structure for the subsample of the 23 physics teachers are similar to the findings presented here. Although content structure can as well explain an additional amount of 19% of the posttest variance between the classes – compared to the baseline model ($\beta_{\text{Stand.}} = .44$, $SE = .15$, $p < .01$; $R^2 = 82\%$, $SE = .12$, $p < .001$), and therefore, content structure can be regarded as another valid indicator for quality of instruction – there were no correlations of content structure to teachers' CK ($r = .05$, $SE = .22$, $p = .831$) or PCK ($r = -.09$, $SE = .23$, $p = .682$).

With the assumption that the correlations might become significant in a larger sample, it seems that teachers with higher CK are able to create more challenging learning opportunities for their students, for example, by putting emphasis on tasks or questions which stimulate students to think or which require cognitively demanding activities such as comparing and analysing. However, teachers' CK cannot explain any differences between the classes in students' content knowledge at the end of a unit. This might indicate that the CK test cannot differentiate between the amount of teachers' CK which is necessary for effective teaching in physics classes and the level above which more CK does not necessarily lead to better teaching. This interpretation is supported by findings from Hill, Rowan and Ball (2005) who found a non-linear effect of primary teachers' content knowledge for teaching mathematics (CKT-M) on student learning gains. There was little systematic relationship between teachers CKT-M and student gains above a certain level of knowledge. The most surprising result of this study is the correlation between teachers' PCK and student learning gains which showed a tendency to be negative and the finding that, compared to teachers' CK, their PCK showed a much smaller correlation with students' cognitive activation. The PCK test had a focus on students' preconceptions; and therefore, we expected it to show higher correlation with students' cognitive activation compared to the CK test. Those findings support the impression that although the ProwiN PCK test for physics teachers was validated and dealing with content widely accepted in the community (but normatively set), it does not measure knowledge related to effective teaching. Our findings might imply in particular that physics teachers' PCK has to be measured differently; they are in line with those findings from another recently conducted study in which no relationship between prospective physics teachers' PCK – measured with a different but similar test – and several dimensions of quality of instruction could be found (Vogelsang, 2014). However, Vogelsang's study also had a rather small sample size ($N = 22$). Therefore, it remains to be seen if the results for the whole ProwiN sample of 35 physics teachers are in line with our results presented here.

Moreover, the results of the chemistry and biology subproject of ProwiN could clarify if our findings indicate problems in the physics-specific operationalization of PCK or if the normatively set facets of the PCK model do not capture the relevant knowledge for effective teaching in chemistry and biology either. Some researchers argue that measuring professional knowledge with paper-and-pencil tests might not be sufficient in order to measure the knowledge relevant for teachers' actions in classroom teaching (Aufschnaiter & Blömeke, 2010). However, using so-called *vignettes* (written descriptions of authentic classroom situations) is seen as a possible solution for the problem of not capturing this knowledge with written tests (Aufschnaiter & Blömeke, 2010). Whereas the ProwiN PCK test as well as the PCK test used by Vogelsang included some less complex vignettes, there are currently efforts in studies on physics teacher education to measure student teachers' PCK with a vignette test using authentically complex teaching contexts (Brovelli, Bölsterli, Rehm & Wilhelm, 2014). Validity arguments for this test instrument are based on a theory-driven test development, expert ratings and group comparisons.

Although this study could not clarify which teacher professional knowledge is relevant for effective teaching, it shows that validating test instruments for measuring professional knowledge – by comparing well-known groups and by analysing the relationship between the different dimensions of professional knowledge – seems not to be sufficient for claiming that such test instruments measure teacher knowledge which is relevant for effective teaching. As a decision criterion in teacher education research, the relevance of the measured professional knowledge of teachers for successful teaching has to be proven for every test instrument.

References

- Abell, S. (2007). Research on science teachers' knowledge. In S.K. Abell & N.G. Lederman (Ed.), *Handbook of research on science education* (pp. 1105-1149). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Aufschnaiter, C. V. & Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 361-367
- Barnett, J. & Hodson, D. (2001). Pedagogical context knowledge: Toward a fuller understanding of what good science teachers know. *Science Education*, 85, 426-453.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T. & Jordan, A. et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47 (1), 133-180.
- Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences (2nd ed.)*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Bonsen, M., Bos, W. & Frey, K. A. (2008). Germany. In I. V. S. Mullis, M. O. Martin, J. F. Olson, D. R. Berger, D. Milne & G. M. Stanco (Ed.), *TIMSS 2007 Encyclopedia. A guide to mathematics and science education around the world. Volume I.* (pp. 203-216). Chestnut Hill, MA: Boston College.
- Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D., Sandmann, A. & Sumfleth, E. (2010). Professionswissen von Lehrkräften in den Naturwissenschaften

- (ProwiN): Kurzdarstellung des BMBF-Projekts. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 341-349.
- Borowski, A., Olszewski, J. & Fischer, H. E. (2010). Fachdidaktisches Wissen von Physikreferendaren. *Der mathematische und naturwissenschaftliche Unterricht*, 63 (5), 260-263.
- Borowski, A. & Riese, J. (2010). Physikalisch-fachdidaktisches Wissen: Was kommt in der Praxis an? *Praxis der Naturwissenschaften - Physik in der Schule*, 59 (5), 5-8.
- Brovelli, D., Bölsterli, K., Rehm, M. & Wilhelm, M. (2014). Using Vignette Testing to Measure Student Science Teachers' Professional Competencies. *American Journal of Educational Research*, 2 (7), 555-558.
- Cauet, E. (2015). *Testen wir relevantes Wissen? – Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten*. Unveröffentlichte Dissertation, Universität Duisburg-Essen.
- Clandinin, D. J. & Connelly, F. M. (1995). *Teachers' professional knowledge landscapes*. New York: Teachers College Press.
- Cochran-Smith, M. & Zeichner, K. M. (Ed.). (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum.
- Dann, H.-D. (1994). Pädagogisches Verstehen: Subjektive Theorien und erfolgreiches Handeln von Lehrkräften. In Reusser, K. & Reusser-Weyeneth, M. (Hrsg.): *Verstehen* (S. 163-182). Bern: Huber.
- Darling-Hammond, L. & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. Hoboken, NJ: Jossey-Bass.
- Dollny, S. (2011). *Entwicklung und Evaluation eines Testinstruments zur Erfassung des fachspezifischen Professionswissens von Chemielehrkräften*. Studien zum Physik- und Chemielernen: Vol. 127. Berlin: Logos.
- Ergönce, J., Neumann, K. & Fischer, H. (2014). The impact of pedagogical content knowledge on cognitive activation and student learning. In: H. E. Fischer, P. Labudde, K. Neumann & J. Viiri (Ed.) (2014). *Quality of instruction in physics. Comparing Finland, Germany and Switzerland* (pp. 145-159), Münster: Waxmann Verlag.
- Fischer, H. E., Boone, W. J. & Neumann, K. (2014). Quantitative research designs and approaches. In N. Lederman & S.K. Abell (Ed.), *Handbook of research in science teaching. Volume II* (pp. 18-37). London: Routledge.
- Fischer, H. E., Borowski, A. & Tepner, O. (2012). Professional knowledge of science teachers. In B. Fraser, K. Tobin & C. McRobbie (Ed.), *Second international handbook of science education*. New York: Springer
- Fischer, H. E., Labudde, P., Neumann, K. & Viiri, J. (Ed.) (2014). *Quality of instruction in physics. Comparing Finland, Germany and Switzerland*, Münster: Waxmann Verlag.
- Gess-Newsome, J. & Lederman, N. G. (Ed.) (1999). *Examining pedagogical content knowledge*. Dordrecht: Kluwer.
- Gess-Newsome, J. (2013). Pedagogical content knowledge. In J. Hattie & E. M. Anderman (Ed.), *Educational psychology handbook series. International guide to student achievement* (pp. 257-259). New York: Routledge.
- Heller, K. A. & Perleth, Ch. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+R)*. Göttingen: Hogrefe.
- Herzog, W. (2005). Müssen wir Standards wollen? Skepsis gegenüber einem theoretisch (zu) schwachen Konzept. *Zeitschrift für Pädagogik*, 51 (2), 252-258.
- Hestenes, D. & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, 30, 159-165.
- Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30 (3), 141.
- Hill, H. C., Rowan, B. & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42 (2), 371-406.

- Kauertz, A., Fischer, H. E., Labudde, P., Härtig, H., Ropohl, M., Walpuski, M. & Sumfleth, E. (2010). Measurement of Content Related Aspects of Science Competence. In: Tasar, M.F. & Cakmakci, G. (Ed.), *Contemporary science education research: Assessment of student learning and development*. (pp. 377-394). Ankara, Turkey: Pegem Akademi.
- Kirschner, S. (2013). *Modellierung und Analyse des Professionswissens von Physiklehrkräften*. Studien zum Physik- und Chemielernen: Vol. 161. Berlin: Logos.
- Klauer, K.J. & Leutner, D. (2007). *Lehren und Lernen*. Weinheim: Beltz.
- Lange, K. (2010). *Zusammenhänge zwischen naturwissenschaftsbezogenem fachspezifisch-pädagogischem Wissen von Grundschullehrkräften und Fortschritten im Verständnis naturwissenschaftlicher Konzepte bei Grundschülerinnen und -schülern*. Dissertation, Westfälische Wilhelms-Universität Münster. Zugriff am 14.11.2014 unter <http://nbn-resolving.de/urn:nbn:de:hbz:6-75459654103>
- Linacre, J. M. (2011). *A user's guide to Winsteps® Ministep: Rasch-model computer programs (3.72.3th ed.): Winsteps.com*.
- 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.
- Magnusson, S., Krajcik, J. & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teacher. In J. Gess-Newsome & N. G. Lederman (Ed.), *Science & technology education library: Vol. 6. Examining pedagogical content knowledge. The construct and its implications for science education*. Dordrecht: Kluwer Academic Publishers.
- Moallem, A. & Moallem, M. (1998). Systemic change in vocational training institutions in France. *International Journal of Disability, Development and Education*, 45, 17-33.
- Ohle, A. (2010). *Primary school teachers' content knowledge in physics and its impact on teaching and students' achievement*. Studien zum Physik- und Chemielernen: Vol. 110. Berlin: Logos.
- Olson, J. F., Martin, M. O. & Mullis, I. V. (Ed.) (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013) *IQB-Ländervergleich 2012: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe*. Münster: Waxmann.
- Park, S. & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49 (7), 922-941.
- Park, S. & Oliver, S.J. (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.
- Peterson, P.L., Carpenter, T.P. & Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlating and case analysis. *Journal of Educational Psychology*, 81 (4), 558-569.
- 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.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M.,...Schiefele, U. (Hrsg.). (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Quesel, C., Möser, G. & Husfeldt, V. (2014). Auswirkungen sozialer Belastungen auf das Schul-, Unterrichts- und Arbeitsklima obligatorischer Schulen in der Schweiz. *Schweizerische Zeitschrift für Bildungswissenschaften*, 36 (2) 2014, 283-306.
- Riese, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften*. Studien zum Physik- und Chemielernen: Vol. 97. Berlin: Logos.

- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2008). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. München: Luchterhand.
- Shrout, P. E. & Fleiss, J. L. (1979). Interclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–3428.
- Shulman, L. S. (1987). Knowledge and teaching: foundation of the new reform. *Harvard Educational Review*, 57 (1), 1–22.
- Spoden, C. & Geller, C. (2014). Uncovering country differences in physics content knowledge and their interrelations with motivational outcomes in a latent change analysis. In: H. E. Fischer, P. Labudde, K. Neumann, & J. Viiri (Ed.) (2014). *Quality of instruction in physics. Comparing Finland, Germany and Switzerland* (pp. 49–63). Münster: Waxmann Verlag.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B., Sandmann, A., Sumfleth, E., Thillmann, H. & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7–28.
- 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.
- Vogelsang, C. (2014). *Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften: Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*. Studien zum Physik- und Chemielernen: Vol. 174. Berlin: Logos.
- Wenning C., Wester, K., Donaldson, N., Henning, S., Holbrook, T., Jabot, M., MacIsaac, D., Merrell, D., Riendeau, D. & Truedson, J. (2011) Professional knowledge standards for physics teacher educators: Recommendations by the CeMaST Commission on NIPTE. *Journal of Physics Teacher Education Online*, 6 (1), Spring 2011. Accessed on 14.11.2014 under [http://www2.phy.ilstu.edu/~wenning/jpteo/issues/jpteo6\(1\)spr11.pdf](http://www2.phy.ilstu.edu/~wenning/jpteo/issues/jpteo6(1)spr11.pdf).
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*, Göttingen: Hogrefe.

Keywords: Professional knowledge, quality of instruction, video-based classroom analysis, cognitive activation, validity

Messen wir relevantes Wissen ? Fachspezifisches Professionswissen von Physiklehrkräften

Zusammenfassung

Kann man davon ausgehen, dass aufwendig validierte Professionswissenstests Wissen messen, das auch wirklich relevant für effektives Unterrichten ist? In dieser Studie werden Zusammenhänge zwischen Fachwissen (CK) und fachdidaktischem Wissen (PCK) von Physiklehrkräften, der kognitiven Aktivierung der Lernenden (als Maß für die Qualität des Unterrichts der Lehrkräfte) und der Leistung der von ihnen unterrichteten Schülerinnen und Schüler in einer Mehrebenenanalyse untersucht. CK und PCK der Lehrkräfte korrelierte weder signifikant mit der kognitiven Aktivierung, noch erklärte es Varianz in den Schülerleistungszuwächsen. Obwohl diese Ergebnisse vorsichtig interpretiert werden müssen, stellen sie insbesondere die Validität des PCK-Tests in Frage, der auf in der Forschung breit akzeptierten, aber normativ gesetzten Wissens-

facetten basiert. Darüber hinaus unterstreichen die Befunde, wie wichtig es ist, Bezüge zwischen Professionswissen, Unterricht und Schülervariablen herzustellen, um die Relevanz des gemessenen Wissens nachzuweisen.

Schlagworte: Professionswissen, Unterrichtsqualität, Video-basierte Unterrichtsforschung, kognitive Aktivierung, Validität

Quelle est la pertinence de ce que nous mesurons? La connaissance disciplinaire professionnelle des enseignants de physique

Résumé

Pouvons-nous être certains que les tests largement validés des connaissances professionnelles des enseignants mesurent effectivement ce qui est pertinent pour un enseignement efficace? Cette étude examine, à l'aide d'analyses multi-niveaux, les relations entre les connaissances disciplinaires professionnelles des enseignants de physique, l'activité cognitive des étudiants – en tant que mesure de la qualité de l'enseignement dans chacune des classes des enseignants – et les performances des étudiants. Ni les connaissances disciplinaires (CD) des enseignants, ni leurs connaissances pédagogiques disciplinaires (CPD) ne corréleront significativement avec le degré d'activation cognitive de leurs étudiants. En outre, leurs connaissances professionnelles ne contribuent pas à expliquer la variance des gains des étudiants. Bien que l'interprétation de ces résultats appelle la prudence, ils interrogent la pertinence des tests CPD qui concernent des savoirs acceptés par la communauté mais qui sont établis normativement. De plus, les résultats soulignent l'importance de corréler les connaissances professionnelles à des variables relatives à la classe et aux élèves afin de prouver que ce que les tests mesurent soit effectivement pertinent pour l'efficacité de l'enseignement.

Mots-clés: Connaissances professionnelles, qualité de l'enseignement, analyse de l'activité en classe sur la base de vidéos, activation cognitive, validité

Le conoscenze valutate sono pertinenti? Le conoscenze specialistiche degli insegnanti di fisica

Riassunto

Possiamo davvero partire dal presupposto che i costosi test professionali attitudinali convalidati siano in grado di valutare le conoscenze rilevanti dal punto di vista di una didattica efficace? In questo studio, utilizzando l'analisi multi-livello, vengono analizzati i rapporti tra le conoscenze specialistiche (CK) e le conoscenze pedagogiche (PCK) degli insegnanti di fisica, l'attivazione cognitiva degli

studenti (come misura della qualità della didattica dei docenti) e le prestazioni degli studenti. La CK e la PCK degli insegnanti non sono correlate significativamente con l'attivazione cognitiva, né spiegano la variabilità del rendimento degli studenti. Anche se questi risultati vanno interpretati con cautela, essi mettono in discussione, in particolar modo, la validità del test PCK, che si basa su aspetti ampiamente accettati nel mondo della ricerca, ma stabiliti a livello normativo. Inoltre, i risultati sottolineano quanto sia importante stabilire collegamenti tra le conoscenze specialistiche, la didattica e le variabili rappresentate dagli studenti, con l'obiettivo di dimostrare la pertinenza delle conoscenze valutate.

Parole chiave: Conoscenze specialistiche, qualità dell'insegnamento, analisi didattica basata sui video, attivazione cognitiva, convalida