5 The contextualised measuring of general pedagogical knowledge and skills: Exploring the use of knowledge in practice

Christian Brühwiler and Lena Hollenstein

Institute of Profession Research and Competence Development, St.Gallen University of Teacher Education, Switzerland

What do teachers need for knowledge-based practice? This is a key research question of great relevance to policy and practice. The chapter highlights the role of teachers' practice-based knowledge and situation-specific skills for transforming general pedagogical knowledge into effective practice. It also provides an overview of the state-of-the-art on measuring such knowledge and skills through contextualised measurements, including the suitability of different approaches for an international large-scale teacher assessment.

Introduction

Teachers' general pedagogical knowledge is seen as an important aspect of teachers' professional competence [see Guerriero (2017[1]) for an overview]. This specialised knowledge, combined with subject-specific knowledge beliefs, motivation and self-regulation [Baumert and Kunter (2011[2]) based on Shulman (1986[3]; 1987[4])], enables teachers to effectively foster learning processes.

Many teachers, however, experience difficulties in the transition phase from teacher education to teaching. They struggle to apply acquired knowledge in the classroom (Doyle, 2006[5]; Wanzare, 2007[6]). Additionally, research results show that it cannot necessarily be assumed that the pedagogical knowledge captured at the end of teacher education correlates with effective practice in the classroom (Brühwiler et al., 2017[7]; Cauet et al., 2015[8]). The common struggle of novice teachers showcases the need for a more comprehensive perspective on teachers' professional knowledge, and a deeper understanding of teacher transition from theory to practice.

Newer models of professional competences include situation-specific skills and highlight their importance for the transformation of knowledge into effective practice (Blömeke, Gustafsson and Shavelson, 2015[9]; Depaepe, Verschaffel and Star, 2020[10]; Krauss et al., 2020[11]). The development of such skills and more practical knowledge seems to require guided teaching experiences. This insight has led to teacher education reforms worldwide [e.g. Ball and Forzani, (2009[12])]: Diverse practical elements have been introduced into the curricula of most teacher education programmes, complemented by induction activities for novice teachers in many countries. Further empirical information on the transition process of teachers and the impact of such reforms would be of great value in terms of policy and practice. This requires knowledge assessments that cover more of those practical knowledge and skills.

The development of such assessments that are reliable and valid is as important as it is challenging [see Depaepe, Verschaffel and Star (2020[10]) for an overview]. Situated- and performance-oriented assessments have been developed that go beyond the limited scope of traditional paper-and-pencil assessments. Many of the more recent assessments require teachers to apply knowledge in hypothetical classroom situations (Blömeke, Gustafsson and Shavelson, 2015[9]; Depaepe, Verschaffel and Star, 2020[10]; Shavelson, 2010[13]). This approach has often been referred to as a contextualised assessment of teacher knowledge.

The aim of this chapter is to give an overview of the state-of-the-art on contextualised measurement of general pedagogical knowledge and skills. It seeks also to highlight the potentials and challenges for such an approach in the context of large-scale international studies. This chapter will first discuss the knowledge and skills teachers need to transform general pedagogical knowledge into effective practice. Then, different approaches to measure these knowledge and skills with contextualised assessments will be compared, including their suitability for an international large-scale assessment of teachers. The chapter ends with some main conclusions for a large-scale contextualised assessment of teacher knowledge across countries.

The role of practice-based knowledge and situation-specific skills for effective teaching

Initial models describe teachers' knowledge as an important aspect of teachers' professional competence, which includes subject-independent knowledge (general pedagogical knowledge) and subject-specific knowledge (such as content knowledge and pedagogical content knowledge). Each of these knowledge component is further broken down into its composing elements, for example the content areas and topics of teacher knowledge.

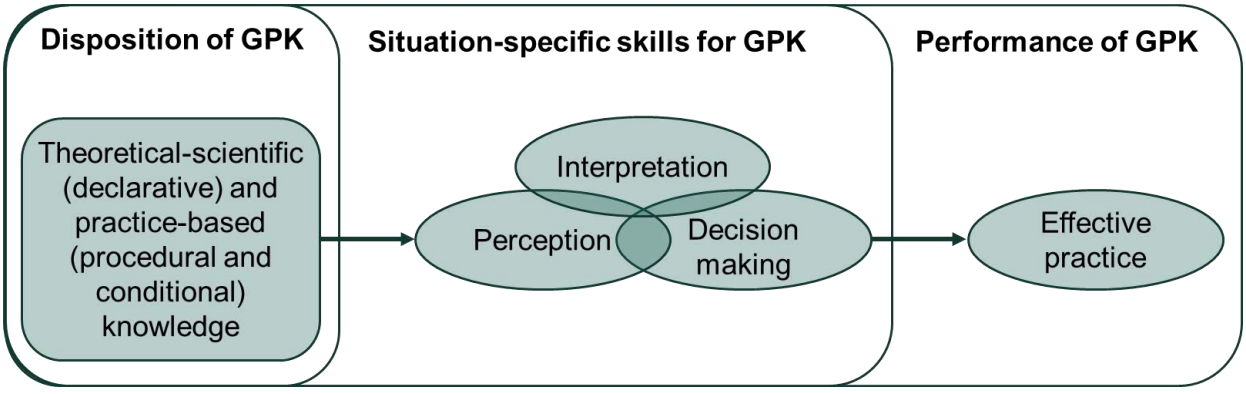
In his international review, König (2015[14]) identified three *content areas* that are common across existing knowledge frameworks: assessment, instruction and learning. These content areas, thus, need to be covered in an international assessment of teachers' general pedagogical knowledge, which is the case for the Teacher Knowledge Survey (TKS) (see Chaper 1 for a detailed description of these content areas). Some authors highlight the particular importance of two sub-dimensions within the broad content areas: (1) classroom management and (2) knowledge of the learning processes of students (Borko and Putnam, 1996[15]; Bromme, 2001[16]; Fennema and Loef Frank, 1992[17]).

Similar to other knowledge, general pedagogical knowledge can also be broken down into two different *knowledge types*: (1) theoretical-scientific (declarative) and (2) practice-based (procedural and conditional) knowledge [see Guerriero (2017[1])]. *Theoretical-scientific knowledge (declarative knowledge;* "knowing what") comprises the knowledge of pedagogical concepts, as well as contents and facts about learning and teaching from educational research (Anderson, 1996[18]).

Practice-based knowledge (action-related knowledge) includes procedural and conditional knowledge (Anderson, 1996[18]). *Procedural knowledge* ("knowing how") corresponds to knowledge relating to concrete teaching-related actions, such as the typical procedures involved in planning and implementing lessons (Artelt and Wirth, 2014[19]). It contains primarily unconscious cognitive operations (Seel, 2003[20]) and helps by using typical procedures to recognise, plan and realise teaching, both stepwise and systematically (König and Blömeke, 2009[21]). *Conditional knowledge* or pragmatic knowledge ("knowing when and why") comprises the knowledge of conditions under which pedagogical theories and concepts (*declarative knowledge*) apply in a given classroom situation, and which teaching methods (*procedural knowledge*) are appropriate and effective in achieving a specific teaching goal (Lenske et al., 2016[22]). This knowledge is therefore necessary when deciding which declarative and procedural knowledge are appropriate in which situation (Woolfolk Hoy and Schönplug, 2008[23]), i.e. which teaching method is appropriate in achieving teaching targets.

Research into teacher expertise established that both theoretical-scientific and practice-based knowledge contribute to expert teachers' performance in the classroom (Bromme, 2001[16]). Yet, teachers also need situation-specific skills to transform their knowledge into practice (Blömeke, Gustafsson and Shavelson, 2015[9]; Depaepe, Verschaffel and Star, 2020[10]; Krauss et al., 2020[11]). It is assumed, therefore, that knowledge is a prerequisite for high-quality teaching and learning outcomes, but that it is an indirect relationship mediated by situation-specific skills (Blömeke, Gustafsson and Shavelson, 2015[9]; Ulferts, 2019[24]). Blömeke et al. (2015[9]) have extended the initial models to cover the mediating role of situation-specific skills. Figure 5.1 displays an adapted version of this competence model.

Figure 5.1. Model on the transformation of theoretical-scientific and practice-based knowledge to effective practice



Source: Adapted from the competence model of (Blömeke, Gustafsson and Shavelson, 2015[9])

The model assumes three *types of skills* as crucial for transforming general pedagogical knowledge into effective practice: Teachers' perception of classroom situations, interpretation and decision making (Guerriero, 2017[1]; Blömeke, Gustafsson and Shavelson, 2015[9]). Teachers constantly need to monitor the classroom and identify situations and cues that, from a professional perspective, are decisive for effective instructional practice and student learning (*perception*). In this regard, teacher knowledge acts as a filter helping teachers to direct their attention to those relevant acts. Such a knowledge-directed perception is indispensable in classroom teaching where numerous teaching and learning acts occur; some of particularly importance, others not. Teachers then have to process the information based on their knowledge of principles of teaching and learning (*interpretation*), in order to make instructional decisions.

It is assumed that *decision making* in classrooms benefits from a great breadth and depth of pedagogical knowledge (e.g. knowledge about various instructional approaches and how they impact student learning and when to apply them). Research is generally in accordance with the models' assumption. Studies have proven that general pedagogical knowledge helps teachers to perceive and interpret classroom situations and decide on an appropriate course of action in the classroom (Charalambous, 2020[25]; König and Kramer, 2016[26]; Krauss et al., 2020[11]).

König and colleagues (2014[27]) showed, for example, that general pedagogical knowledge, measured with the *Teacher Education and Development Study in Mathematics* (TEDS-M) instrument [cf. König et al. (2011[28])] and comprising declarative and partly procedural general pedagogical knowledge, correlates substantially with teachers' situation-specific skills to interpret. The current models see teachers' theoretical-scientific and practice-based knowledge as a necessary, but not sufficient prerequisite for situation-specific skills (perception, interpretation, decision making) and for teachers' effective practice (Blömeke, Gustafsson and Shavelson, 2015[9]; Kaiser and König, 2019[29]; Krauss et al., 2020[11]). Professional perception, interpretation and decision making requires specialised knowledge but also specific training. Gold, Förster and Holodynski (2013[30]), for instance, showed that guided video analysis can be used to train teachers' perception abilities. Consequently, assessing teachers' situation-specific skills alongside their pedagogical knowledge is key for understanding how well they are equipped for high-quality teaching.

Contextualised assessments of practice-based knowledge and situation-specific skills

Only in recent decades has there been a shift from self-assessment to more objective measures of teachers' professional competence (Hill, Beisiegel and Jacob, 2013[31]). By now, several measurement instruments have been developed that measure theoretical-scientific and practice-based knowledge in a more contextualised way and include situation-specific skills (König et al., 2011[28]). They frequently use (text or video) vignettes that confront teachers with typical classroom situations (Gasteiger et al., 2020[32]), or challenging situations in which teachers fail to provide an adequate response (Brühwiler et al., 2017[7]; Brühwiler and Vogt, 2020[33]; Lenske et al., 2015[34]). Table 5.1 lists various recently developed contextualised assessments. The table does not represent an exhaustive list but rather a selection of instruments differing in assessment approach.

Content areas

It becomes obvious when comparing the existing instruments that they vary in content covered. Some instruments cover a broad range of topics from different content areas (Brühwiler et al., 2017[7]; Lenske et al., 2015[34]; Charalambous, 2020[25]), whereas other instruments focus on certain content dimensions and sub-dimensions, e.g. classroom management or assessment (Brühwiler and Vogt, 2020[33]; König and Kramer, 2016[26]; Wildgans-Lang et al., 2020[35]).

Table 5.1. Overview of contextualised assessments of general pedagogical knowledge and skills

Study/Instrument	Pedagogical-psychological teaching knowledge (PPTK)	Professional knowledge in natural science (ProWiN)	Classroom management expertise (CME)	Video test for adaptive teaching competency (ATC video test)	Teaching simulation	Simulated online environment
Reference	Brühwiler et al. (2017[7])	Lenske et al. (2015[34])	König and Kramer (2016[26]); König (2015[36])	Brühwiler and Vogt (2020[33])	Charalambous (2020[25])	Wildgans-Lang et al. (2020[35])
Construct measured	Pedagogical-psychological teaching knowledge	Pedagogical-psychological knowledge	Classroom management expertise	Teachers' adaptive performance competency	Performance in the simulation (action-related competence)	Teachers' diagnostic competence
Content area	Instruction (and learning)	Instruction and learning	Instruction	Instruction and learning	Instruction and learning	Assessment
Types of knowledge and skills	Situation-specific skills (interpretation and decision making)	Practice-based knowledge (procedural and conditional)	Situation-specific skills (perception and interpretation)	Situation-specific skills (perception, interpretation and decision making)	Situation-specific skills (perception, interpretation and decision making)	Theoretical-scientific and practice-based knowledge
Task format	Text vignettes	Text vignettes	Video vignettes	Video vignettes	Digital simulations	Digital simulations
Response format	Closed	Closed	Closed and open	Open (oral answers)	Open	Open
Evaluation and scoring method	Pairwise agreement with experts	Pairwise agreement with experts	Absolute agreement with experts	Scoring based on a coding system	Evaluation as correct/incorrect based on a coding system	Evaluation as correct/incorrect based on a coding system

Note: The table does not present an exhaustive list of instruments for a contextualised assessment of general pedagogical knowledge and skills, but it gives an overview of instruments differing in assessment approach.

Type of knowledge and skills

The various instruments are designed to measure different types of knowledge and skill, as explained earlier. In particular, the instruments used to measure classroom management expertise [CME] and adaptive teaching competency [ATC] video test focus on teachers' interpretation of a specific classroom situation. Other instruments (e.g. pedagogical-psychological teaching knowledge [PPTK], ATC video test) include teachers' decision making. The Professional knowledge in natural science (ProWiN) study covered practice-based knowledge (procedural and conditional knowledge) (Lenske et al., 2015[34]), whereas the simulated online environment by Wildgans-Lang et al. (2020[35]) includes teachers' theoretical-scientific and practice-based knowledge in its assessment.

Task format and administration mode

As shown in Table 5.1, the task format corresponds closely to the types of knowledge and skills captured. Whereas traditional uncontextualised tasks seem appropriate for measuring teachers' theoretical-scientific knowledge, contextualised approaches using vignettes (hereafter, *vignette approaches*) are essential for assessing teachers' practice-based knowledge and situation-specific skills. The contextualisation can either be realised with text- or video-vignettes (Brühwiler et al., 2017[7]; Gasteiger et al., 2020[32]; König and Kramer, 2016[26]; Krauss et al., 2020[11]; Lenske et al., 2015[34]). With regard to video-vignettes, two approaches can be distinguished: (1) participants are shown a short, completed video sequence, which they then have to evaluate [e.g. the video-vignette test to assess classroom management expertise (König, 2015[36])] and (2) participants are shown a longer video sequence, in which they are asked to stop the video themselves as soon as they notice a situation that was not handled adequately. Once they stop

the video, they should express their thoughts and suggest a more adequate alternative to the teacher's action [e.g. the video test to assess teachers' adaptive teaching competence; (Brühwiler and Vogt, 2020[33])]. Whereas the former approach mainly focuses on teachers' interpretation of the situation shown in the video, the latter allows insights into their decision making as the teachers have to suggest a more adequate alternative action.

Empirical studies regarding the validity of these instruments (e.g. can the quality of the teaching be predicted) confirm that both assessment approaches, text- and video-based vignettes, capture knowledge and skills relevant for instructional quality (Brühwiler et al., 2017[7]; Hollenstein, Affolter and Brühwiler, forthcoming[37]; König and Kramer, 2016[26]; Lenske et al., 2016[22]; Lenske, Wirth and Leutner, 2017[38]).

The most recent developments in measurement instruments are digital simulations, which also use typical classroom situations as their stimulus (Charalambous, 2020[25]; Wildgans-Lang et al., 2020[35]). For example, in a virtual mathematics lesson, participants indicate how the teacher should interact with the students concerning topics such as providing explanations, using representations, analysing student work or contributions, and responding to students' requests for help. This is used, for example, in the simulation of Charalambous (2020[25]) to assesses (pre-service) teachers' action-related competence. Wildgans-Lang et al. (2020[35]) developed an instrument with which (pre-service) primary school teachers should diagnose virtual students' competence levels, while the students solved mathematical problems. Digital simulations are promising tools in the assessment of characteristics, closely linked to classroom performance (Charalambous, 2020[25]; Wildgans-Lang et al., 2020[35]). Nevertheless, to date, few studies exist that evidence their relationship to instructional quality (Charalambous, 2020[25]).

Response format, evaluation and scoring method

A further distinction between the measuring instruments can be made in terms of response formats. In principle, the instruments can be categorised into closed or open answer formats. There are several options for the closed response formats and the choice among them determines, to a certain extent, the methods for evaluating and scoring the answers obtained from teachers. In the context of large-scale assessments, typically multiple choice items, Likert-scale items or short answer formats are used.

Multiple choice items force teachers to choose one out of several options provided. The underlying assumption is that answers are either correct or incorrect, i.e. in agreement or disagreement with scientific theories and evidence. Consequently, answers are scored binary (e.g. no point for incorrect answers and one point for every correct answer). Partial credit models allow for a more differentiated evaluation of teachers' responses: Responses may indicate some knowledge and are, thus, partially credited (e.g. teachers receive one point for a partially correct answer and two points for an entirely correct answer).

Multiple choice items might be appropriate for assessing theoretical-scientific knowledge, but for situation-specific skills *Likert scales* seem more appropriate. The perception and interpretation of classroom situations as well as decision making in teaching is not a question of either-or but of choosing among options that are more or less appropriate. Likert scales can take account of this by asking teachers to judge the degree of appropriateness, effectiveness or utility of different options for teaching: Brühwiler and colleagues, for instance, asked teachers to judge on 4-point Likert scales whether the various options proposed are more or less useful (Brühwiler et al., 2017[7]).

The scoring happens via comparison with an expert solution, either through absolute agreements (as done for the instrument by König (2015[36]) measuring Classroom management expertise) or relative agreements with experts [as done for the instrument measuring pedagogical-psychological knowledge by Lenske et al. (2015[34])]. Absolute agreements require teachers to rate the options exactly like experts, whereas relative agreements consider if teachers rank the options similarly as experts.

Expert ratings have already been successfully used in international, large-scale assessments. For example, in the 2009 round of PISA (Programme for International Student Assessment), pairwise

comparisons with expert ratings were used to measure students' reading strategies (OECD, 2010[39]) (for more detail see Box 5.1). Similarly, TALIS Starting Strong tested expert rankings and other forms of rankings to measure pre-school teachers' responses on several situational judgement items (Nielsen et al., 217[40]). Though both did not measure teacher knowledge (the latter measured staff practices and the former students' reading strategies), these examples show the feasibility of the approach for international surveys.

Box 5.1. Pairwise comparison with expert rating

Using a pairwise comparison with experts' answers is based on the assumption that practice-based knowledge or situation-specific skills become apparent, particularly when teachers are able to distinguish between adequate and inadequate teacher actions, and are able to rank teacher actions according to their adequacy (Lenske et al., 2016[22]). Besides PISA (OECD, 2010[39]), other text-vignette instruments use the pairwise comparison as an evaluation method (Schlagmüller and Schneider, 2007[41]; Lenske et al., 2015[34]; Brühwiler et al., 2017[7]; Rutsch et al., 2018[42]).

A pairwise comparison with expert rating means that the participants' and the experts' answers are related to one another. There are different ways of doing this. If the item relation of the test person (a > b) corresponds to the item relation of the expert solution (a > b), e.g. 2 points are awarded. Only 1 point is given if the test respondent considered the items to be equivalent (a = b). If the item relation of the test respondent (a < b) is opposite to the item relation of the expert answer (a > b), the test respondent receives 0 points (Brühwiler et al., 2017[7]; Lenske et al., 2016[22]; Rutsch et al., 2018[42]). In PISA 2009 (OECD, 2010[39]) a binary coding was applied. Participants received 1 point for a correct relation or 0 points for an incorrect relation, compared to the experts' solution.

Open response items ask teachers to produce short answers without providing any options. Answers can be given in writing (König, 2015[36]) or orally (Brühwiler and Vogt, 2020[33])]. Participants' answers are usually scored on several criteria by trained experts according to a structured coding scheme. The schemes describe the different criteria for rating the teacher answers and provides examples of correct answers. It also details the points awarded for each criterion that is met. Coding schemes can be developed deductively (e.g. from a theoretical framework) or inductively (from the empirical teacher responses, for example via content analysis). Coding schemes can vary in their prescriptiveness and, thus, the amount of inferences required by the raters. Low-inferent codes allow to code every response with the least possible amount of inferences by the raters.

Advantages and disadvantages of different approaches in the context of international large-scale surveys

As the previous section has highlighted, various options for designing teacher assessments exist. Each choice comes with certain advantages and disadvantages and they are not equally suitable for an international large-scale assessment (see Table 5.2 for an overview).

Choice 1: Narrow vs. broad content coverage

The breadth of content covered with the assessment impacts the generalisability of obtained results, as well as the efficiency of data collection and the psychometric quality of obtained data. Instruments covering a broad range of knowledge and skills assess the constructs more comprehensively than instruments with a more narrow focus (as these only provide information on specific content areas and information on other areas is missing).

In contrast, a narrow approach can render a more detailed coverage of topics in a specific area (e.g. detailed information on the knowledge and skills needed for effective classroom management). Further advantages of a narrower assessment are that sufficient reliability in terms of internal consistency can generally achieved with fewer items and less testing time.

It would be of great importance to select content that is relevant for teaching across countries such as classroom management and further knowledge about instruction, learning and assessment (see Chapter 2).

Table 5.2. Advantages and disadvantages of different design choice

Design Choice	Advantages	Disadvantages
	Choice 1	
Broad coverage of content	• a more comprehensive measurement of knowledge and skills	• requires more items items to obtain sufficient internal consistency for all areas covered • requires more time to complete
Narrow coverage of content	• allows for a more detailed coverage of a specific aspect • sufficient reliability in terms of internal consistency may be achieved with less items, limiting the response burden for teachers	• insights into teachers' knowledge and skills is limited to a certain aspect, leaving out others
	Choice 2	
Theoretical-scientific knowledge	• prerequisite for situation-specific skills and effective teaching • closely linked to the content of initial teacher education and suited for measuring its output	• insights into teachers' knowledge and skills is limited to a certain aspect, leaving out others
Practice-based knowledge	• prerequisite for situation-specific skills and effective teaching • important for evaluating the practical elements of teacher education	• indirectly linked to effective teaching practice
Situation-specific skills	• more direct relationship with effective teaching • important for understanding the transformation of knowledge into practice; and evaluating the practical elements of teacher education	• standardised measurement of situation-specific skills comes with additional challenges
	Choice 3	
Uncontextualised tasks	• widely used in international assessments • less developmental effort • suitable for assessing theoretical-scientific knowledge	• no reference to classroom situations • not suitable for assessing practice-based knowledge and situation-specific skills
Video vignettes	• cover the complexity and authenticity of classroom situations • suitable for assessing practice-based knowledge and situation-specific skills	• achieving standardisation and comparability across diverse cultures and educational contexts is challenging • requires technical equipmentt
Text vignettes	• provide a neutral description and, thus, easier applicable to various contexts than video vignettes • less cost and resource intensive than video vignettes • suitable for assessing practice-based knowledge and situation-specific skills	• cover less well the complexity and authenticity of classroom situations
Digital simulations	• close to actual experience of teachers in classrooms • various factors can be systematically varied • suitable for assessing practice-based knowledge and situation-specific skills	• very demanding to create an internationally comparable, simulated classroom environment • time consuming in the development stage • scarcity of empirical evidence for its relationship to teaching quality and student outcomes • limited experience of the scientific community, and, thus, is a risky choice

Design Choice	Advantages	Disadvantages
	Choice 4	
Multiple choice items scored as correct/incorrect or using partial credit models	• widely used in international assessments • take little time to answer and score • provide reliable and valid information • partial credit allows a more differentiated measurement	• do not allow respondents to express their own ideas • force teachers to choose one of the proposed alternatives, though none may reflect their preferred option • cannot capture more complex tasks of teachers in classrooms (e.g. decision making and professional judgement) • teacher decisions and actions may not be correct or incorrect, but may vary in effectiveness, depending on the situational context
Likert-scale items using agreement with experts for scoring	• have been shown to provide reliable and valid information in international large-scale assessments • reflects the typical choice of teachers among various more or less suitable options and require a moderate time to complete • can use the degree of alignment with expert opinion, rather than judging teachers' decision and action as correct or incorrect	• do not allow respondents to express their own ideas • requires additional efforts for developing a reliable and valid expert rating system such as the involvement of (international) expert
Open response format scored with a coding system	• respondents are free in their response • allows a highly individualised and differentiated assessment of knowledge and skills • takes account of the specificities of different cultures and educational systems • allows for a differentiated evaluation of teachers' answers and perspectives on complex classroom situations	• requires expressive and reflective skills and a high motivation of teachers • very time consuming to complete, and, thus, lead to high response burdens for teachers • coding is time consuming requires high expertise and training

Choice 2: Theoretical-scientific knowledge, practice-based knowledge or situation-specific skills

Teacher education is largely concerned with teaching theories, concepts and principals of teaching and learning. Thus, instruments focusing on theoretical-scientific knowledge of teachers are suitable tools for measuring the output of teacher education to a certain extent. Insights from these instrument into the transition process from theory to practice of teachers are limited. Understanding why some teachers (especially novice teachers) struggle to apply acquired knowledge in the classroom, requires instruments that cover practice-based knowledge and situation-specific skills of teachers. Yet, a standardised measurement of situation-specific skills in an international context is challenging. A particular attention to the administration mode and task formats are required as well as to the classroom situations selected for the vignettes.

Choice 3: Uncontextualised tasks, vignettes or simulations

Uncontextualised tasks (mostly multiple choice tasks) are widely and successfully used in many international assessments. Many reliable and valid instruments already exist and the development effort is lower than for more recent approaches. Such tasks are suitable for measuring theoretical-scientific knowledge. They need to be complemented with text- or video vignettes that confront teachers with authentic and typical classroom situations, in order to provide information on situation-specific skills and practice-based knowledge of teachers.

Existing vignette instruments have their origin in a particular cultural and educational context. Adapting the vignettes and items for other contexts may be quite challenging and will not simply be a matter of translation (Hambleton, Merenda and Spielberger, 2005[43]). Text vignettes seem to be a more feasible choice for an

international assessment than video vignettes, as they are probably easier to adapt and develop. Video vignettes, which show classroom situations with teachers and students in a specific national and cultural context, may not work in other contexts and cultures. A text vignette can provide a more neutral description of a classroom situation. Respondents can interpret the described situation within their specific cultural and educational context and translate the description into a mental image that fits their specific experience. Furthermore, text vignettes have lower technical requirements and are, thus, a more economic means of creating contextualised assessments than video vignettes.

However, the question arises as to whether text vignettes can adequately represent the complexity of a classroom situation. Generally, text vignettes provide only brief summaries of classroom situations and acts. Conversely, video vignettes can capture the complexity of classroom situations where multiple exchanges happen simultaneously and also non-verbally. They are, therefore, closer to the authentic situation and pose particular, real life challenges to teacher perception, interpretation and decision making (Kramer et al., 2020[44]; Stürmer, Konings and Seidel, 2013[45]).

Though promising, *digital simulations* are relatively new to the field. Given the scarcity of empirical evidence, developing an internationally validated, simulated classroom environment would be very demanding and time consuming. In addition, the lack of experience of using them in diverse contexts would be a risky choice for an international assessment. Furthermore, their additional benefit has not yet been empirically tested; in particular, there is a lack of empirical evidence of their relationship with teaching quality and student outcomes. Consequently, at this time, the implementation of digital simulations in large-scale assessments is not recommended.

Choice 4: Open or closed response format and choice of a scoring and evaluation method

Developing cross-cultural valid items and agreeing on responses across countries is a major challenge, regardless of the answer format used. Closed response items (such as multiple-choice or Likert scales) may, however, be the most efficient option: They take little time to answer and - once the scoring grid is developed - to score. The use of partial credit models to score answers allows both a speedy and more differentiated scoring. A major drawback is that they do not allow teachers to express their own ideas and force them to choose one of the proposed alternatives, even though none may reflect their preferred option. They are also less suited for measuring practice-based knowledge and situation-specific skills. In most classroom situations, there is not a right choice but a most adequate one (e.g. choosing between different teaching approaches varying in effectiveness and suitability for a specific situational context).

In view of the complex situations in classrooms, using Likert scales and comparing the responses to expert answers seems like an economic alternative. As already mentioned, the feasibility of such an approach for international assessments has been successfully demonstrated (OECD, 2010[39]; Nielsen et al., 217[40]). However, developing an expert rating system requires additional efforts and the involvement of (international) experts.

Open formats enable teachers to provide differentiated judgements, allowing for a more individualised assessment of teacher knowledge. As a result, the specificities of cultural and educational systems are accounted for. Yet, answering open response items takes time and requires motivation and more complex skills (expressive and reflective skills, as well as writing skills for written responses). Furthermore, it is very challenging and resource intensive to create an internationally valid and reliable coding system for the evaluation of answers. Open response items also require expertise and coding training for those categorising the items.

Conclusion

Based on the aforementioned advantages and disadvantages of different assessment approaches, considerable added value is expected from text-vignettes. They should describe typical and challenging classroom situations that most teachers are confronted with in their classrooms. Vignettes allow for a measurement of practice-based knowledge and situation-specific skills. The response burden for teachers, as well as the costs and resources, can be kept modest if Likert scales and comparisons with expert ratings are used. In the following, two examples of text vignettes for a contextualised assessment are described in more detail.

Example text vignettes and items for a large-scale assessment

Two example vignette items are shown in Table 5.3. They stem from an adapted version of the Swiss instrument measuring "pedagogical-psychological teaching knowledge" [PPTK; (Brühwiler et al., 2017[7]; Brühwiler et al., forthcoming[46])] and aim at capturing situation-specific skills. Both vignettes describe situations that most teachers around the globe face and represent both typical and challenging situations.

The first vignette describes a situation in which a teacher returns corrected papers to students after an exam. Solving the items requires knowledge about the attribution theory of achievement motivation (Weiner, 1985[47]). The second vignette is related to the context of classroom management and represents a typical situation in which a pupil is inattentive because he is bored. Each vignette is followed by two sets of Likert-scale items immediately after the vignette, which measure different situation-specific skills (Blömeke, Gustafsson and Shavelson, 2015[9]): The first set asks how the respondent would act in this situation and, therefore, require *professional decision making* of the responding teachers. The second set asks teachers for a *professional interpretation* of the described classroom situation.

Teacher responses were scored using the pairwise comparison. The experts' rating (also shown in Table 5.3.) was generated by consulting a total of 16 experts. Experts had either a strong teaching expertise *or* research expertise. The answer most frequently chosen by the experts was used as the expert rating. Discrepancies were discussed among a smaller group of experts, in order to reach consent (Brühwiler et al., 2017[7]; Brühwiler et al., forthcoming[46]). If consent could not be achieved, the item was excluded.

Brühwiler and colleagues (2017[7]) tested the assumption that situation-specific skills are closely related to effective practice, as explained earlier. The results show that PPTK predicts both instructional quality and the pupils' academic achievements. These findings underline the potential of contextualised assessments.

Table 5.3. Examples of text vignettes focusing on typical classroom situations

Vignette 1: Students wrote an exam in class. The teacher returns the corrected papers.

a) How would you act in this situation? Which action is likely, which is unlikely? *Please tick one box per line.*		very unlikely	unlikely	Likely	very likely
A	Karin receives a bad grade in her exam and the teacher says to her: "I know that you practiced a lot, but this is not one of your strengths."	☒	☐	☐	☐
B	Peter receives a very good grade in his exam. The teacher says to him, "It doesn't seem to have been difficult."	☒	☐	☐	☐
C	Michael receives a good grade in his exam with the comment: "You see, your diligence has really paid off."	☐	☐	☐	☒
D	To Anna she says, "Your talent is obvious once again." Anna receives a very good grade.	☐	☐	☒	☐

b) Analyse the reaction of the student Karin (first statement above). What effect could the teacher's feedback have on Karin? Which of the effects outlined below is likely, which is unlikely? *Please tick one box per line.*		very unlikely	unlikely	Likely	very likely
A	Karin goes home reassured because she now knows that the teacher is not disappointed in her.	☐	☒	☐	☐
B	Karin will try harder next time. She really wants to show the teacher that she is good at maths.	☒	☐	☐	☐
C	Karin will not feel like practicing a great deal for the next exam, as it has not been of any use so far.	☐	☐	☐	☒
D	The next time Karin achieves a good result, she will believe that she has finally been able to show her abilities.	☒	☐	☐	☐

Vignette 2: A teacher explains a difficult maths task in the classroom, which almost no student was able to solve. Patrick, a very good student in mathematics, paints the paper in front of him during the lesson without disturbing any of his classmates.

a) How would you act in this situation? Which action is likely, which is unlikely? *Please tick one box per line.*		very unlikely	unlikely	Likely	very likely
A	I ask a question to assess comprehension and then call Patrick.	☐	☒	☐	☐
B	I try to make eye contact with Patrick.	☐	☐	☐	☒
C	I ignore Patrick's behaviour.	☐	☐	☒	☐
D	I say to Patrick: "Please stop painting now."	☐	☒	☐	☐
E	While I continue the lesson, I stand close to Patrick's table.	☐	☐	☐	☒
F	I ask Patrick what is going on.	☐	☒	☐	☐
G	I send Patrick to the next room.	☒	☐	☐	☐
H	I give Patrick extra homework.	☒	☐	☐	☐
I	I give Patrick the opportunity to explain the task to his fellow students in his own words.	☐	☐	☐	☒

b) Analyse the situation. What could be the difficulty? Which statements are likely, which are unlikely? *Please tick one box per line.*		very unlikely	unlikely	Likely	very likely
A	Patrick is one of the few students who solved the maths problem correctly.	☐	☐	☐	☒
B	Patrick is trying to annoy the teacher with his behaviour.	☒	☐	☐	☐
C	Patrick paints the paper in front of him out of boredom.	☐	☐	☒	☐
D	The task is too difficult for Patrick.	☒	☐	☐	☐

Note: Vignettes were scored using a pairwise comparison with expert ratings. Crosses indicate the experts' answers; adapted version of the German-language instrument (Brühwiler et al., 2017[7]; Brühwiler et al., forthcoming[46]); Vignette 1 relates to attribution theory, vignette 2 relates to teaching disturbances (classroom management).

Further recommendations for a contextualised assessment of teacher knowledge and skills

The question of which type of knowledge and skills teachers need to transform general pedagogical knowledge into effective teaching practice is highly relevant. To date, however, it remains largely

unanswered. Further empirical information, which would allow for a deeper understanding of these transformation processes, would be of great value in terms of both policy and practice. As it seems that practice-based knowledge and situation-specific skills play a crucial role in this transformation process, it would be of great importance to include a sufficient number of contextualised items in an international survey on teacher knowledge.

Based on the considerations in this chapter, considerable added value could be expected by developing text vignettes which describe difficult classroom situations in which teachers' practice-based pedagogical knowledge and situation-specific skills are required to solve the challenging situation. These situations should be typical and relevant for most teachers across countries. The development of text vignettes instead of video vignettes is recommended not only for economic reasons, but also because text vignettes can provide a more neutral description of a classroom situation and are, therefore, easier to develop for various cultural and educational contexts.

The development of vignettes and items that are valid and reliable across all participating countries is crucial. They should be based on theory and practice. The OECD Global Teaching InSights (GT) study could be a fruitful starting point for identifying typical and challenging classroom situations relevant across countries (OECD, 2020[48]). Applying a rater-scoring system represents an efficient and appropriate scoring method that has successfully been used in other international studies (OECD, 2010[39]; Nielsen et al., 217[40]).

An extensive validation of the items and expert scoring system in a pilot study and in the field trail is recommended. An important aspect is measurement invariance testing (Milfont and Fischer, 2010[49]): It must be ensured that vignettes are comparable across countries. Additionally, it would be important to test the predictive or at least concurrent validity of the developed instrument (König, 2015[14]), i.e. whether the measured knowledge and skills are actually related to effective teaching practice. The TALIS includes various scales on teaching practices (e.g. classroom management and cognitive activation), which could be used for the validity testing.

To sum up, it is less a question of ''whether'' but ''how'' to assess general pedagogical knowledge and situation-specific skills in a large-scale assessment (see Table 8.1 in Chapter 8 for the main takeaways from this chapter for TALIS and the TKS assessment module). Empirical information would be of great value for policy and research, as it would allow for a greater scientific understanding of how knowledge is transformed in practice, and why some teachers struggle with this. The inclusion of more contextualised items, therefore, would strengthen the value of the TKS assessment module.

References

Anderson, J. (1996), "ACT: A simple theory of complex cognition", *American Psychologist*, Vol. 51/4, pp. 355-365, http://dx.doi.org/10.1037/0003-066X.51.4.355. [18]

Artelt, C. and J. Wirth (2014), "Kognition und Metakognition [Cognition and Metacognition]", *in Pädagogische Psychologie [Pedagogical Psychology]*, Beltz, Weinheim. [19]

Ball, D. and F. Forzani (2009), "The work of teaching and the challenge for teacher education", *Journal of Teacher Education*, Vol. 60/5, pp. 497–511, http://dx.doi.org/10.1177/0022487109348479. [12]

Baumert, J. and M. Kunter (2011), "Das Kompetenzmodell von COACTIV [The competence model of COACTIV]", *in Professionelle Kompetenz von Lehrkräften [Teachers' Professional Competence]*, Waxmann, Münster, https://doi.org/10.1007/978-3-658-00908-3_13. [2]

Blömeke, S., J. Gustafsson and R. Shavelson (2015), "Beyond dichotomies competence viewed as a continuum", *Zeitschrift für Psychologie*, Vol. 223/1, pp. 3-13, http://dx.doi.org/10.1027/2151-2604/a000194. [9]

Borko, H. and R. Putnam (1996), "Learning to teach", *in Handbook of Educational Psychology*, MacMillan, New York. [15]

Bromme, R. (2001), "Teacher expertise", *in International Encyclopedia of the Social and Behavioral Sciences*, Pergamon, London. [16]

Brühwiler, C. et al. (2017), "Welches Wissen ist unterrichtsrelevant? [What knowledge is relevant to teaching?]", *Zeitschrift für Bildungsforschung*, Vol. 7/3, pp. 209-228, http://dx.doi.org/10.1007/s35834-017-0196-1. [7]

Brühwiler, C. et al. (forthcoming), "Situationsspezifische Fähigkeiten im Bereich Klassenführung: Ein Vergleich zwischen einem textbasierten und einem videobasierten Erhebungsinstrument [Situation-specific skills in classroom management]". [46]

Brühwiler, C. and F. Vogt (2020), "Adaptive teaching competency: Effects on quality of instruction and learning outcomes", *The Journal of Educational Research*, Vol. 12/1, pp. 119-142, http://dx.doi.org/10.1037/0003-066X.51.4.355. [33]

Cauet, E. et al. (2015), "Does it matter what we measure? Domain-specific professional knowledge of physics teachers", *Swiss Journal of Educational Research*, Vol. 37/3, pp. 462-479, http://dx.doi.org/10.24452/sjer.37.3.4963. [8]

Charalambous, C. (2020), "Reflecting on the troubling relationship between teacher knowledge and instructional quality and making a case for using an animated teaching simulation to disentangle this relationship", *ZDM Mathematics Education*, Vol. 52, pp. 219–240, http://dx.doi.org/10.1007/s11858-019-01089-x. [25]

Depaepe, F., L. Verschaffel and J. Star (2020), "Expertise in developing students' expertise in mathematics: Bridging teachers' professional knowledge and instructional quality", *ZDM Mathematics Education*, Vol. 52, pp. 179-192, http://dx.doi.org/10.1007/s11858-020-01148-8. [10]

Doyle, W. (2006), "Ecological management and classroom management", *in Handbook of Classroom Management*, Lawrence Erlbaum, New York. [5]

Fennema, E. and M. Loef Frank (1992), "Teachers' knowledge and its impact", *in Handbook of Research on Mathematics Teaching and Learning*, MacMillan, London. [17]

Gasteiger, H. et al. (2020), "Mathematical pedagogical content knowledge of early childhood teachers: a standardized situation-related measurement approach", *ZDM Mathematics Education*, Vol. 52, pp. 193-205, http://dx.doi.org/10.1007/s11858-019-01103-2. [32]

Gold, B., S. Förster and M. Holodynski (2013), "Evaluation eines videobasierten Trainingsseminars zur Förderung der professionallen Wahrnehmung von Klassenführung im Grundschulunterricht t [Evaluation of a video-based training program to enhance professional perception of classroom leadership]", *Zeitschrift für pädagogische Psychologie*, Vol. 27, pp. 141–155. [30]

Guerriero, S. (2017), *Pedagogical Knowledge and the Changing Nature of the Teaching Profession*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264270695-de. [1]

Hambleton, R., P. Merenda and C. Spielberger (eds.) (2005), "Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures", *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*, Lawrence Erlbaum Associates, New Jersey, https://doi.org/10.4324/9781410611758. [43]

Hill, H., M. Beisiegel and R. Jacob (2013), "Professional development research: Consensus, crossroads, and challenges", *Educational Researcher*, Vol. 42/9, pp. 476-487, http://dx.doi.org/10.3102/0013189X13512674. [31]

Hollenstein, L., B. Affolter and C. Brühwiler (forthcoming), "The importance of primary school teachers' pedagogical-psychological knowledge in mathematics teaching and learning". [37]

Kaiser, G. and J. König (2019), "Competence measurement in (mathematics) teacher education and beyond: Implications for policy", *Higher Education Policy*, Vol. 32/4, pp. 597-615, http://dx.doi.org/10.1057/s41307-019-00139-z. [29]

König, J. (2015), *Background Document: Designing an International Assessment to Assess Teachers' General Pedagogical Knowledge (GPK)*, OECD Website,, http://www.oecd.org/education/ceri/Assessing%20Teachers%E2%80%99%20General%20Pedagogical%20Knowledge.pdf. [14]

König, J. (2015), "Measuring classroom management expertise (CME) of teachers: A video-based assessment approach and statistical results", *Cogent Education*, Vol. 2/1, pp. 1-15, http://dx.doi.org/10.1080/2331186X.2014.991178. [36]

König, J. and S. Blömeke (2009), "Pädagogisches Wissen von angehenden Lehrkräften [Pedagogical knowledge of feature teachers]", *Zeitschrift für Erziehungswissenschaft*, Vol. 12, pp. 499-527, http://dx.doi.org/10.1007/s11618-009-0085-z. [21]

König, J. et al. (2014), "Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach", *Teaching and Teacher Education*, Vol. 38, pp. 76-88, http://dx.doi.org/10.1016/j.tate.2013.11.004. [27]

König, J. et al. (2011), "General pedagogical knowledge of future middle school teachers: On the complex ecology of teacher education in the United States, Germany, and Taiwan", *Journal of Teacher Education*, pp. 188-201, http://dx.doi.org/10.1177/0022487110388664. [28]

König, J. and C. Kramer (2016), "Teacher professional knowledge and classroom management: On the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME)", *ZDM Mathematics Education*, Vol. 48/1-2, pp. 139-151, http://dx.doi.org/10.1007/s11858-015-0705-4. [26]

Kramer, C. et al. (2020), "Classroom videos or transcripts? A quasi-experimental study to assess the effects of media-based learning on pre-service teachers' situation-specific skills of classroom management", *International Journal of Educational Research*, Vol. 1-13, p. 101624, http://dx.doi.org/10.1016/j.ijer.2020.101624. [44]

Krauss, S. et al. (2020), "Competence as a continuum in the COACTIV-Study—"The cascade model"", *ZDM Mathematics Education*, Vol. 52/3, pp. 311-327, http://dx.doi.org/10.1007/s11858-020-01151-z. [11]

Lenske, G. et al. (2015), "Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests [Pedagogical professional knowledge of teachers: Evaluation of the ProwiN test]", *Zeitschrift für Erziehungswissenschaft*, Vol. 18/2, pp. 225-245, http://dx.doi.org/10.1007/s11618-015-0627-5. [34]

Lenske, G. et al. (2016), "Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht [The importance of pedagogical knowledge for classroom management and for students' achievement]", *Zeitschrift für Erziehungswissenschaft*, Vol. 19/1, pp. 211-233, http://dx.doi.org/10.1007/s11618-015-0659-x. [22]

Lenske, G., J. Wirth and D. Leutner (2017), "Zum Einfluss des pädagogisch-psychologischen Professionswissens auf die Unterrichtsqualität und das situationale Interesse der Schülerinnen und Schüler [The impact of pedagogical-psychological knowledge on instructional quality]", *Zeitschrift für Bildungsforschung*, Vol. 7, pp. 229-253, http://dx.doi.org/10.1007/s35834-017-0200-9. [38]

Milfont, T. and R. Fischer (2010), "Testing measurement invariance across groups: Applications in cross-cultural research", *International Journal of Psychological Research*, Vol. 3/1, pp. 111-130, http://dx.doi.org/10.21500/20112084.857. [49]

Nielsen, T. et al. (217), "Measuring process quality in early childhood education and care through Situational Judgement Questions: Findings from TALIS Starting Strong 2018 Field Trial", *OECD Education Working Papers*, pp. 1-57, https://doi.org/10.1787/19939019. [40]

OECD (2020), *Global Teaching InSights: A Video Study of Teaching*, OECD Publishing, Paris, https://dx.doi.org/10.1787/20d6f36b-en. [48]

OECD (2010), *PISA 2009 Results: Learning to Learn: Student Engagement, Strategies and Practices (Volume III)*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264083943-en. [39]

Rutsch, J. et al. (2018), "Modellierung der Testletstruktur bei vignettenbasierten Testverfahren mit geschlossenem Antwortformat [Modelling of the testlet structure of vignette-based tests with closed response format]", *in Effektive Kompetenzdiagnose in der Lehrerbildung*, Springer, Wiesbaden. [42]

Schlagmüller, M. and W. Schneider (2007), *Würzburger Lesestrategie-Wissenstest für die Klassen 7–12 [Würzburg Reading Strategy Knowledge Test for Grades 7–12]*, Hogrefe, Göttingen. [41]

Seel, N. (2003), *Psychologie des Lernens: Lehrbuch für Pädagogen und Psychologen [Psychology of Learning: Textbook for Educators and Psychologists]*, UTB, Stuttgard. [20]

Shavelson, R. (2010), "On the measurement of competency", *Empirical Research in Vocational Education and Training*, Vol. 2/1, pp. 41-63, http://dx.doi.org/10.1007/BF03546488. [13]

Shulman, L. (1987), "Knowledge and teaching: Foundations of the new reform", *Harvard Educational Review*, Vol. 57/1, pp. 1-23, http://dx.doi.org/10.17763/haer.57.1.j463w79r56455411. [4]

Shulman, L. (1986), "Those who understand: Knowledge growth in teaching", *Educational Researcher*, Vol. 15/2, pp. 4-14, http://dx.doi.org/10.2307/1175860. [3]

Stürmer, K., K. Konings and T. Seidel (2013), "Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning", *The British Journal of Educational Psychology*, Vol. 83/3, pp. 467–483, http://dx.doi.org/10.1111/j.2044-8279.2012.02075.x. [45]

Ulferts, H. (2019), "The relevance of general pedagogical knowledge for successful teaching: Systematic review and meta-analysis of the international evidence from primary to tertiary education", *OECD Education Working Papers*, No. 212, OECD Publishing, Paris, https://dx.doi.org/10.1787/ede8feb6-en. [24]

Wanzare, Z. (2007), "The transition process: The early years of being a teacher", *in Handbook of Teacher Education*, Springer, Wiesbaden, https://doi.org/10.1007/1-4020-4773-8_23. [6]

Weiner, B. (1985), "An attributional theory of achievement motivation and emotion", *Psychological Review*, Vol. 92/4, pp. 548-573, http://dx.doi.org/10.1037/0033-295X.92.4.548. [47]

Wildgans-Lang, A. et al. (2020), "Analyzing prospective mathematics teachers' diagnostic processes in a simulated environment", *ZDM Mathematics Education*, Vol. 52, pp. 241–254, http://dx.doi.org/10.1007/s11858-020-01139-9. [35]

Woolfolk Hoy, A. and U. Schönplug (2008), *Pädagogische Psychologie [Pedagogical Psychology]*, Pearson Deutschland GmbH, Hallbergmoos. [23]