Asking the right question - Using different types of student feedback in developing ePCK for in-service biology teachers in dynamic teaching environments at school and at university

Daniel Hartmuth, Daniel Schlichting, Jonathan Haudenschild, Helmut Küchenhoff & Birgit J. Neuhaus

Published online: 25 Nov 2025.

Submit your article to this journal ☑

Article views: 178

View related articles ☑

View Crossmark data ☑

co:gent

TEACHER EDUCATION & DEVELOPMENT|RESEARCH ARTICLE   OPEN ACCESS   Check for updates

Asking the right question - Using different types of student feedback in developing ePCK for in-service biology teachers in dynamic teaching environments at school and at university

Daniel Hartmuth[a] (iD), Daniel Schlichting[b], Jonathan Haudenschild[a], Helmut Küchenhoff[b] (iD) and Birgit J. Neuhaus[a] (iD)

[a]Biology Education, Faculty for Biology, LMU Munich, Munich, Germany; [b]Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Munich, Germany

ABSTRACT

In science education research, the RCM (refined consensus model of Pedagogical Content Knowledge) conceptualizes ePCK (enacted Pedagogical Content Knowledge) as an emergent situation-specific property, influenced by domain-specific factors (content knowledge, skills) and generic factors (cognitive activation, constructive support, classroom management). In-service teachers develop ePCK by iterating a plan-teach-reflect cycle. Student feedback, especially digital summarising feedback, supports this process. However, it remains unclear whether a simple question feedback, differentiated feedback, or a grand rubric score best improves teaching across educational levels. While student feedback on teaching quality is valuable, identifying the most effective feedback type for enhancing ePCK—especially when comparing high-school and university settings—remains unresolved. We analyzed three types of digital summative feedback (a simple question, differentiated responses, and a grand rubric score) from 56 high-school lessons and 175 university lectures in a within-subject design, examining correlations among these feedback types. Students' overall impression of lesson quality was highly correlated with differentiated feedback. A grand rubric score derived from this feedback could substitute for the single unspecific question. Results are consistent across both educational levels. A single-question feedback method may suffice for practical integration into the plan-teach-reflect cycle. However, the lack of detailed insights requires iterative feedback. Given the limited reliability of high-school data, further studies with in-service teachers are necessary. We emphasize the dynamic nature of the plan-teach-reflect cycle and propose student feedback as a key tool for sustaining instructional improvement throughout a teacher's career.

Introduction

Developing teaching strategies and teaching skills during the teaching process is dynamic and influenced by numerous interrelated variables with non-linear effects on outcomes (Gess-Newsome & Lederman, 1999; Hattie, 2009; Neuhaus, 2021). The Refined Consensus Model (RCM) of Pedagogical Content Knowledge (PCK) encompasses many of these variables and their relationships, providing a framework for detailed study (Carlson et al., 2019). The RCM emphasizes the dynamic and situational nature of teaching through the plan-teach-reflect cycle (Alonzo et al., 2019), whereas PCK is divided into three realms: collective (cPCK), personal (pPCK), and enacted (ePCK). cPCK refers to the collective consensus knowledge of the scientific community; pPCK represents a teacher's personal, cognitive, and articulable knowledge; ePCK is the behavioural manifestation of this knowledge (Carlson et al., 2019). The

relationship between pPCK and ePCK is complex (Mientus et al., 2022) often making it difficult to measure them as distinct entities (Park, 2019).

For the assessment of ePCK various methods are being proposed and tested, such as questionnaires, video observations and voice observations (Suresh et al., 2018). A practical method for accurately assessing teaching quality, and thus an approach to measuring ePCK is through summarising student feedback, which should prompt necessary adjustments in the teaching approach on the basis of the feedback received (Wisniewski et al., 2020). To develop ePCK based on cPCK and pPCK, a cyclical feedback process is essential for providing valuable information for a teacher's plan-teach-reflect cycle (Chan & Hume, 2019).

Rationale of the Study: While existing research has demonstrated the utility of student feedback in measuring teaching quality and informing the plan–teach–reflect cycle (Bijlsma et al., 2022; Wisniewski et al., 2020), there remains a critical gap regarding how different types of feedback drive improvements in enacted pedagogical content knowledge (ePCK). Specifically, it is not yet clear which feedback format—simple undifferentiated ratings, differentiated domain-specific items, or an aggregated rubric score—provides the most actionable insight for teachers operating in diverse educational contexts.

In practice, teachers often work under significant time and resource constraints, needing to make quick, practical decisions rather than exhaustive analyses. They therefore require fast-and-frugal feedback tools that balance precision with efficiency, enabling a workable compromise between depth of information and immediacy of insights (Bijlsma et al., 2021). Understanding which feedback formats yield the most actionable, rapid insights will help teachers optimise their reflective processes within real-world classroom pressures.

Clarifying these trade-offs will inform both theoretical models of PCK development and the practical design of digital feedback tools for in-service teachers. Thus, this study seeks to provide a systematic rationale for selecting and implementing each feedback type based on their potential impact on teachers' ePCK.

Variable under study: types of student feedback. In this study, we focus on three digital student feedback variables designed to capture different levels of instructional quality within the plan–teach–reflect cycle. First, simple-question feedback is operationalised as a single, undifferentiated item asking students to "Rate the lesson" on a five-star scale, providing a broad overall assessment of teaching effectiveness (Rollett et al., 2021). Second, differentiated feedback comprises three separate items, each corresponding to one generic domain of teaching quality—cognitive activation (e.g. "The pupils are actively engaged in the classroom activities"), constructive support (e.g. "The teacher supports the pupils individually in their learning process"), and classroom management (e.g. "The available lesson time is used for engaging with the learning content"). These items draw on validated measures from Göllner et al. (2021) to capture specific facets of enacted PCK. Third, we derive a grand rubric score by computing the arithmetic mean of the three differentiated feedback ratings, aggregating the domain-specific measures into a single composite indicator of ePCK (Chan et al., 2019). By comparing these three feedback variables—simple-question, differentiated domain-specific, and grand rubric—we aim to determine whether the simple-question feedback can serve as a valid proxy for the more detailed rubrics in both high-school and university teaching contexts.

Focusing on student feedback for educators through digital rating systems allows quick and practical feedback (Bijlsma et al., 2022), avoiding the loss of valuable insights that can occur with lengthy feedback questionnaires (Wisniewski et al., 2020).

To measure ePCK, we explored student feedback based on a single question (Rollett et al., 2021) and compared it with differentiated feedback from three different facets of instructional quality, namely cognitive activation, constructive support, and classroom management (Chan et al., 2019).

The goal of this study was to test simple student feedback for teachers against a broader set of ePCK measures, implement student feedback into the plan-teach-reflect-cycle of the RCM and thus enable successful measurement and development of ePCK. We developed a digital tool (Hartmuth et al., 2025) that enables the collection of student lesson ratings in a dynamic classroom. This approach enables teachers to monitor the difference in ratings of current versus prior lessons and assists them in adapting their teaching methods according to the student ratings, thus developing their ePCK.

Theory

ePCK as an emergent property within the RCM

The so-called consensus model of PCK has emerged in recent years by emphasising two factors (Berry et al., 2015): the dynamic nature of PCK as manifested in teaching situations, and the inclusion of explicit science teaching skills at subject, topic and concept levels (Drits-Esser et al., 2017). In its most recent consensus version, the model highlights the practitioner's viewpoint on PCK and underscores the numerous feedback loops involved in the development of PCK in practice. The illustration of these feedback loops shows the pedagogical reasoning cycles of teaching, also referred to as the plan-teach-reflect cycle (Carlson et al., 2019). The various layers of the model are linked through filters and enhancers that facilitate knowledge exchange, highlighting the model's inherently dynamic and non-linear nature. These hypothetical filters and enhancers should be considered as catalysts for knowledge exchange, rather than the causal or correlated factors expected in a linear model (Carlson et al., 2019). Although the RCM focuses on different domains within PCK, it acknowledges Shulman's first definition of PCK that the "(…) broader professional knowledge bases are foundational to teachers' PCK" (Carlson et al., 2019).

This collection of individual skills and practical tools, known as ePCK, is incorporated within the second element of PCK, referred to as personal PCK (pPCK). As soon as a teacher can articulate part of their ePCK, the conscious part of ePCK is transformed into pPCK. Conversely, as soon as a teacher uses part of pPCK in action, this part is transformed into ePCK (Irmer et al., 2023). Thus, pPCK emerges from situational ePCK and vice versa.

The collective pedagogical content knowledge (cPCK) of the scientific community serves as an additional source of information, alongside ePCK, for the development of pPCK. The relationship between pPCK and cPCK mirrors the symmetrical connection observed between ePCK and pPCK. A source for the development of cPCK of the whole scientific community is the expressible, articulable body of knowledge of all teachers (Seidel & Shavelson, 2007).

The individual and dynamic framework of the RCM emphasises the individual importance of each teacher's PCK. Both the set of knowledge and skills used in carrying out the teaching cycle (ePCK) and the conscious knowledge of one's teaching (pPCK) are intertwined. As soon as a teacher acts, ePCK unfolds; as soon as a teacher articulates, describes or explains the act of teaching, pPCK is expressed (Morine-Dershimer & Kent, 1999). From a scientific point of view, it is difficult to disentangle ePCK and pPCK; moreover, ePCK and pPCK are considered emergent properties of other teacher knowledge bases (Shulman, 1986), which are non-linear, dependent and dynamic. Currently, the debate about PCK as an integrative concept rather than a transformative property leans heavily towards the latter (Carlson et al., 2019; Gess-Newsome & Lederman, 1999), and this raises the question of how to foster individual, situation-dependent ePCK and pPCK as concepts that are more than, but different from, the sum of their parts (Morine-Dershimer & Kent, 1999).

Positioning within TPACK

The Technological Pedagogical Content Knowledge (TPACK) framework extends Shulman's original PCK by explicitly integrating technological knowledge (TK) alongside pedagogical (PK) and content knowledge (CK) (Mishra & Koehler, 2006). This highlights how technology both frames and enables teaching and learning. While the RCM's enacted PCK (ePCK) emphasises the practical interplay of pedagogical and content knowledge in a plan–teach–reflect cycle, TPACK prompts us to see technology not just as context, but as an active constructive element of teacher knowledge (Koehler et al., 2012). In our study, the Teacher Tunes platform embodies TK in action. The QR-mediated digital star-rating shapes how digital feedback is collected and interpreted, thereby becoming part of the enacted knowledge itself. Contrasting ePCK of the RCM with TPACK shows that our digital feedback tool both mediates and enriches pedagogical enactment (Harris et al., 2009).

Development of ePCK in practice through the teaching cycle

The RCM highlights the significance of developing ePCK through practical experience at an individual level, placing the teaching cycle at its core (Blömeke & Kaiser, 2014). This cycle involves a continuous

feedback loop of reflection, planning, and teaching, which relies on ongoing observational data about a teacher's progress to meet the informational demands of the teaching cycle (Chan & Hume, 2019).

For the development of ePCK, teaching experience is needed (Davidowitz & Potgieter, 2016); in the absence of practical training of contextualized teaching situations, individual ePCK does not improve (Drits-Esser et al., 2017; Förtsch et al., 2016). To a large extent the question of how to improve ePCK systematically in practice remains unclear (Mientus et al., 2022; Sorge et al., 2019).

The connection between theory and practice can be explored using the RCM as a framework, which outlines two pathways for the development of a teacher's ePCK. The first involves progressing through the teaching cycle in a continuous feedback loop, while the second focuses on engaging with the scientific community to access the broader cPCK, supporting the personal development of ePCK (Mientus et al., 2022). These insights, which emphasise the growth of individual PCK through practical experience and action, align with contemporary constructivist learning theories, particularly the concept of experiential learning (Kolb & Kolb, 2005; Morris, 2020). The development of ePCK is an individual process, shaped by action and influenced by the specific learning context.

In response to diverse and evolving learning environments, and based on their current level of ePCK available in a given context, teachers develop a repertoire of skills and knowledge that enhances their performance across various settings, thereby advancing their ePCK (Tobin & McRobbie, 1999). However, this development is limited by the constraints of their specific learning environment and their existing pPCK and ePCK (Irmer et al., 2024). When there is a mismatch between these constraints and a chosen teaching strategy, teachers repeatedly struggle to implement the intended interventions successfully (Behling et al., 2022; Shavelson, 2020). Effective organization of contributing factors to enhance teaching quality is only possible when they are considered collectively (Praetorius et al., 2012; Praetorius & Gräsel, 2021). The RCM illustrates this approach through its central cyclical structure, highlighting the development of teaching skills as an iterative process similar to a feedback loop, as noted in research within the natural sciences (Behling et al., 2022). By sequentially following this plan-teach-reflect-cycle, a teacher develops additional ePCK, putting reflected pedagogical content knowledge into action (Hartmuth et al., 2025; Irmer et al., 2023).

ePCK and facets of instructional quality

The concept of teaching quality can be approached through the eyes of science education as well as from the perspective of general research on teaching and learning, encompassing different foci of two disciplines. General research on teaching and learning often compares different variables, a line of research that has led to a such basic dimensions of teaching quality as cognitive activation, classroom management and constructive support (Hattie, 2009; Kunter & Voss, 2013). On the other hand, science education often focuses on domain-specific variables of quality of instruction, such as use of models and conceptual orientation (Förtsch et al., 2018). In order to build a bridge between these different approaches to teaching quality, efforts have been made to integrate the facets of subject-specific quality traits into the broader dimensions of general teaching quality (Chan & Hume, 2019; Praetorius & Gräsel, 2021). These endeavors are in line with current findings that generic and subject-specific variables are intertwined (Dorfner et al., 2018; Steffensky et al., 2018) and that no categorial distinction between general and subject-specific traits is justified (Neuhaus, 2021). Several recent studies analyzing the interrelationship of different variables of teaching quality hint at the usefulness of bundling subject-specific variables under the broader generic categories (Chan et al., 2019; Dorfner et al., 2018; Heinitz et al., 2022; Neuhaus, 2021; Praetorius & Gräsel, 2021).

This tendency to incorporate subject-specific quality traits as aspects of more general dimensions is mirrored in the RCM through the emergence of ePCK from subject-specific and cross-cutting knowledge bases during the teaching cycle (Gess-Newsome et al., 2019). The concept of ePCK acknowledges a generic background whilst stressing the importance of subject-specific traits of teaching quality (Carlson et al., 2019). Within the RCM, this is accomplished by placing the individual teaching performance at the center of the model and emphasizing the dynamic nature of the plan-teach-reflect cycle when all knowledge bases manifest themselves in the situation-specific performance of ePCK (Mientus et al., 2022).

Process orientated measurement of ePCK

Assessing ePCK has recently become of interest, especially as several studies suggest that teaching performance affects the quality of student learning success (Baumert et al., 2010; Hattie, 2009). In practice, these serve as the primary benchmark and proxy for evaluating teachers' ePCK success (Alonzo et al., 2012). Research has demonstrated that ePCK directly influences student outcomes (Keller et al., 2017) and competence (Mahler et al., 2017).

Theoretically, the problem of measuring ePCK lies in the notion that it is fed from a myriad of interrelated variables with non-linear effects (Carlson et al., 2019; Dorfner et al., 2019). Several attempts have been made to construct a rubric that captures a wide range of generic and subject-specific knowledge bases of ePCK (Chan & Hume, 2019; Mientus et al., 2022; Park & Kyung Suh, 2015). Agreement is needed on a set of empirically derived variables constituting ePCK in order to direct research and communicate findings within a model or framework (Chan et al., 2019).

Practically the RCM imposes the difficulty that although ePCK is manifested in an individual and situation-specific setting, and has a clear dynamic component (Gess-Newsome et al., 2019), the knowledge bases that ePCK require, such as pPCK of an individual teacher, cPCK and other broader and generic knowledge bases, are considered more static (Berry et al., 2015). A more dynamic focus requires measurement of ePCK during the act of teaching, while the more static interpretation of PCK encompasses a focus on a teacher's knowledge and reasoning about teaching practices (Gess-Newsome et al., 2019).

Methodologically the problem of measuring ePCK lies in finding the appropriate match between the conceptualization of ePCK and the corresponding fit with a data source (Chan et al., 2019). Studies with a more static framework often focus on measuring ePCK through post-hoc knowledge questionnaires and interviews (Davidowitz & Potgieter, 2016), whereas empirical studies that focus on the dynamic nature of ePCK use more practice-orientated data sources such as specialist observations and videos, or deep learning structures of actual teaching performances in the classroom (Chan & Hume, 2019; Suresh et al., 2022). During the pre-service teacher (PST) training expert colleagues often assess ePCK on the basis of empirically founded criteria (Kramer et al., 2020; Wang et al., 1993). During the more theoretical training phases, individual PCK is often addressed through pen-and-paper articulation of pPCK in relation to video-based teaching scenarios. (Irmer et al., 2023; Seidel & Stürmer, 2014; Shavelson, 2012). Information about the assessment or acquisition of teaching quality, or the development of teaching skills, is often assumed to be transferable from PST formation to the in-practice school environment (Shavelson, 2020).

Thus, theoretical, practical and methodological considerations must be taken into account, placing ePCK on a continuum in all three dimensions and influencing how ePCK is captured through measurement (Carlson et al., 2019) (Figure 1).

Measuring ePCK with the help of student feedback

Another option for teachers' ePCK assessment is to use feedback from observers (Behnke, 2016). In order to gather valuable information about teaching performance, the feedback process has to be structured in specific and well-researched ways (Hattie & Timperley, 2007). When feedback is collected regularly and the information obtained used to adapt teaching practices and strategies, feedback is one of the stronger factors contributing to teaching quality (Hattie & Clarke, 2019). Research has shown that student feedback especially has a strong influence on the development of teaching quality (Wisniewski et al., 2019). The idea is to ask students in class (Kämpfe, 2009) to observe teachers in action (Hattie & Clarke, 2019), since recourse to expert observers or advanced technological methods is often unfeasible (Bijlsma et al., 2022; Chan & Yung, 2018; Hense, 2013). Students possess a key to evaluating overall teaching quality by means of feedback (Wisniewski et al., 2019); that the information is gathered via their ratings, which are reliable indicators of teaching quality (Kämpfe, 2009), and make efficient use of time and resources (Bijlsma et al., 2019). Building on this efficiency, mean student rating values can serve as a viable and valid measure of ePCK (van der Scheer et al., 2019).
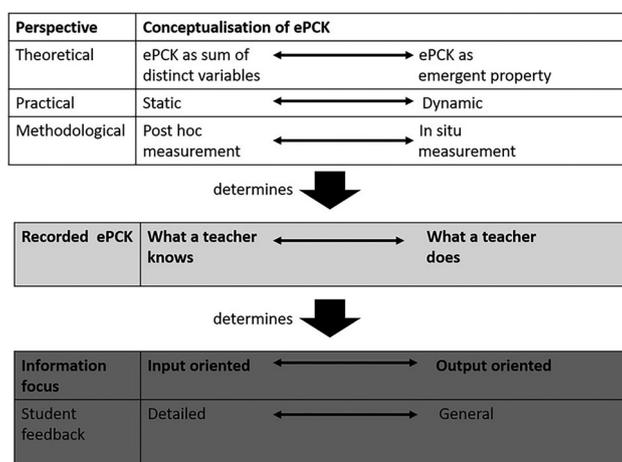
Figure 1. Recorded ePCK: Only a specific part of ePCK is recorded on a continuum determined by the conceptualization of ePCK from theoretical, practical and methodological perspectives. Information focus: The concepualisation of ePCK determines the appropriate kind of student feedback.

Rubric scores: Determining the effectiveness of student feedback for ePCK measures

Student feedback can be collected in different ways, time frames, frequency and detail (Narciss & Huth, 2006), but information to be acted upon is what teachers seek by gathering student feedback (Wisniewski & Zierer, 2021). Study results show that gathering student feedback along generic dimensions of teaching quality with the help of rubric scores gives teachers actionable information to improve their teaching skills and overall teaching quality (Wisniewski et al., 2018). Rubric scores provide a structured framework for assessing teaching quality with the help of student feedback by defining clear performance criteria (Chan et al., 2019). They serve as an effective tool for teachers by offering a transparent and systematic approach to evaluating instructional quality and thus make student feedback more useful (Panadero & Jonsson, 2020). Within the classroom, rubric-based assessments allow teachers to reflect on their practice, identify areas for improvement, and make data-driven adjustments to enhance student learning experience (Allen & Tanner, 2006).

Rubrics, when aligned with key generic teaching dimensions such as cognitive activation, constructive support, and classroom management, offer a reliable measure of ePCK (Chan et al., 2019). Rubric-based feedback can capture essential aspects of teaching performance in real-time, allowing educators to adapt their strategies dynamically (Suresh et al., 2022). A key advantage of rubric scores is their ability to provide consistent and objective insights into teaching quality across different educational settings. Unlike broad, unstructured feedback, rubrics help standardise student responses and reduce subjectivity in evaluations. This ensures that teachers receive actionable feedback, which can be integrated into the plan-teach-reflect cycle to refine their instructional approaches (Mientus et al., 2022).

The three generic domains with the largest descriptive value of overall teaching quality and the strongest impact on teaching quality are cognitive activation, constructive support and classroom management (Dorfner et al., 2018; Hattie & Clarke, 2019; Praetorius & Gräsel, 2021). Whether this student feedback is collected with the help of questions referring to generic or subject-specific variables depends on the goal of the feedback receiver. In order to improve overall teaching quality, items belonging to the generic domains of teaching are appropriate (Rollett et al., 2021). It has been shown that these generic domains of teaching quality can be assessed by means of student feedback (Ruth-Herbein et al., 2022). Moreover this generic domain information on teaching quality helps teachers improve their teaching (Manzi et al., 2022). If the goal of a teacher is to improve subject-specific skills, questions gathering feedback on these specific practices has to be collected (Praetorius et al., 2012). In order to construct a fitting rubric for the measure of ePCK via student feedback, the conceptualization of ePCK used has to be justified and placed on the continuum from a theoretical, practical and methodological perspective (Chan et al., 2019).

When ePCK is conceptualized as consisting of distinct variables (theoretical perspective) in a static environment (practical perspective) a student feedback questionnaire is often employed once

(methodological perspective), because the assessment of distinct facets of ePCK requires clear criteria (Wisniewski & Zierer, 2021). In this case the focus of the student feedback lies in gathering detailed input information on strengths and weaknesses, or effects of specific variables of ePCK (Neuhaus, 2021).

When conceptualising ePCK as an emergent property with many interacting variables (theoretical perspective) in a dynamic teaching situation (practical perspective) with interest in the specific ePCK of a teacher (methodological perspective), student feedback is better collected regularly and in less detail owing to feasibility and time constraints (Bijlsma et al., 2019). This leads to the gathering of information about behavioural components of ePCK. This orientation of teaching performance by means of more general student feedback accepts the cost of reduced detail about specific facets of ePCK and focuses on the feasibility of collecting information about a teacher's ePCK in situ. The use of undifferentiated, single-question feedback has gained prominence across various fields owing to its simplicity, efficiency, and effectiveness in capturing essential user perceptions (Allen et al., 2022). This concept pertains to the practice of collecting user feedback through a single, straightforward question, similar to the undifferentiated star ratings used by platforms such as WhatsApp and Adobe. This straightforward approach has become a global standard, demonstrating that complex feedback systems are not always necessary to produce meaningful insights (Reichheld, 2003). In the field of education, similarly, research indicates that even in complex learning environments, an undifferentiated single-item measure can reliably assess core aspects of the learning experience (Atroszko et al., 2019); and shorter surveys yield higher response and completion rates than longer ones (Kost & da Rosa, 2018). The use of Audience Response Systems, which often employ simple, undifferentiated questions, boosts student engagement (Funnell, 2017). Beyond fostering engagement, implementing a simple single-question feedback also provides real-time feedback, allowing lecturers to adjust their teaching strategies promptly (Schmidt et al., 2020). Together, these findings underline that simple, one-question feedback processes are not only practical but also highly informative, offering robust insights without overburdening the stakeholders in the feedback process.

The assessment of ePCK via general student feedback comes at the cost of information loss, and loss of resolution concerning the facets of ePCK. Thus, capturing ePCK in situ via student feedback is only feasible if the information captured is useful (Figure 1).

Both the pedagogy literature on effective student feedback and the RCM, with its subject-specific focus, emphasize a cyclical structure where reflection is closely linked to lesson planning and the teaching process. However, research indicates that teachers often struggle to incorporate feedback about mistakes or potential improvements into their teaching practices (Shavelson, 2020). Feedback has been shown to act as a catalyst for reflection within the classroom, significantly influencing lesson planning, teaching approaches, and consequently, the development of ePCK (Hattie & Clarke, 2019). Incorporating student feedback into the plan-teach-reflect cycle emphasizes the importance of information gathering to support the personal development of ePCK (Elstad et al., 2017). This is particularly relevant as a teacher's behavior in the classroom has a direct impact on teaching quality (Seidel & Shavelson, 2007; Seidel & Stürmer, 2014).

Research objectives and hypotheses

Incorporating regular student feedback into classroom routines could implement effectively a feedback process within a teacher's plan-teach-reflect cycle of the RCM. Such an approach would systematically improve their ePCK and teaching methods by supplying the necessary data for the teacher.

The main goal of the present study was to find out if simple and unspecific student feedback on lesson quality is able to yield as valuable information as that from a rubric composed of student feedback on the three main domains of teaching quality, cognitive activation, constructive support and classroom management. Furthermore, we investigated the transferability of student feedback measures between university and high school environments. This approach enables teachers to monitor the ratings of current versus prior lessons and to adapt their teaching methods to the student ratings, thus developing their ePCK.

Prior studies suggest high correlations between overall satisfaction scores and detailed quality measures in classroom settings (Bijlsma et al. 2022). This indicates that simple-question feedback can capture key variance in teaching quality to serve as a practical proxy under time constraints (Rollett et al., 2021).

Teaching contexts vary across high school and university settings. Thus, examining comparing results enhances validity and offers the option of tailored recommendations (Chan et al., 2019).

We considered the following hypotheses:

H1: A simple one-question student feedback is correlated highly with feedback question answers on the three domains of teaching quality, namely cognitive activation, constructive support and classroom management.

H2: A simple one-question student feedback is correlated highly with a grand rubric score constructed from the three domains of teaching quality, namely cognitive activation, constructive support and classroom management (Chan et al., 2019).

H3: The results from H1 and H2 show no diverging tendencies between university and school background (Shavelson, 2020).

Methods

Sample

255 lectures and lessons were rated in total, 184 at universities and 71 at a school. To prevent single student impressions obscuring the data, lectures and lessons with fewer than 3 ratings (9 lectures) and those with missing values (15 lessons with no student rating at all) were omitted. Ultimately, the remaining material comprised 175 university lectures from 20 different courses delivered by seven lecturers and 56 school lessons from three classes taught by the same teacher. The data is described in the subsequent section.

Study design

For the current study, lessons and lectures at a university and a high-school were rated by students. The study took place from October 2023 – February 2024. Before the commencement of the study, all lecturers and teachers at the university and school were familiarized with the feedback tool. All lectures and lessons during this period were rated. Data were analyzed on the level of the classes. Each teacher/ lecturer was rated after each lesson with four feedback questions in a within-study design. Each student had to rate the teacher with the broad student feedback question and with one feedback question from all three domains of generic teaching quality. Those questions were taken from an evaluated and standardized questionnaire on generic teaching quality from Göllner and colleagues currently in use in several states in Germany (Göllner et al., 2021). All four questions were randomly shuffled for each student at each rating event (Figure 2).

October	Timeline		February

High School	**Lesson 1**	**Lesson 2**	**Lesson N**
Teacher 1/Class 1	Ø Student feedback	Ø Student feedback	Ø Student feedback
Teacher 1/Class N	Ø Student feedback	Ø Student feedback	Ø Student feedback
Teacher 2...			

University	**Lesson 1**	**Lesson 2**	**Lesson N**
Lecturer 1/Course 1	Ø Student feedback	Ø Student feedback	Ø Student feedback
Lecturer 1/Course N	Ø Student feedback	Ø Student feedback	Ø Student feedback
Lecturer 2...			

Figure 2. Timeline and study design.

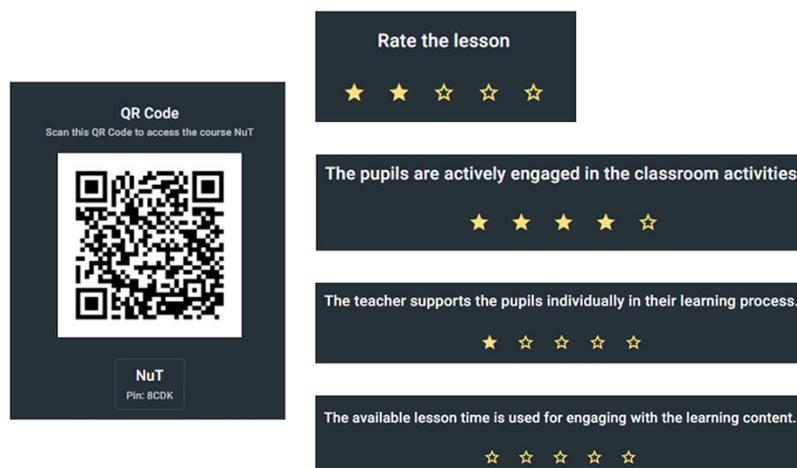**Figure 3.** Front-end of the teachertunes.de website.

Figure 4. Star rating scale for student feedback: students can rate the lesson from one to five stars and submit their rating by pressing the "confirm" button. Each device can only be used to submit one rating.

Instruments

The student rating environment teacher Tunes

The study utilized the web-based feedback platform Teacher Tunes (teacher-tunes.de), created at LMU Munich (Hartmuth et al., 2025). This tool gathers student feedback on lessons through a rating system. Teachers sign up with individual accounts and create distinct courses for each participating class (Figure 3).

After each lesson, students used their smartphones and a QR code (Figure 4) to rate the lesson with several questions on a five-star scale (Figure 4). In order to collect reliable data for the three dimensions of teaching quality cognitive activation, constructive support and classroom management we used the items proposed and validated in the research project "Promoting Teaching Quality through Classroom Observation and Feedback" (Manzi et al., 2022; Ruth-Herbein et al., 2022).

For full replicability, the exact item stems we used from Göllner et al. (2021) are listed below, all on a five-star scale:

Cognitive activation: The pupils are actively engaged in the classroom activities.
Constructive support: The teacher supports the pupils individually in their learning process.
Classroom management: The available lesson time is used for engaging with the learning content.
Category: broad student rating: Question: Rate the lesson.
Category: student activation: Question: The pupils are actively engaged in the classroom activities.
Category: constructive support: Question: The teacher supports the pupils individually in their learning process
Category: classroom management: Question: The available lesson time is used for engaging with the learning content. 5 Star rating scale for student rating: the students can rate the lesson from one to five stars and submit their rating with the "confirm" button. Rating is possible only once from each device.

Data description

The objective was to measure and enhance the quality of teaching at schools by incorporating student feedback in the plan-teach-reflect cycle of individual teachers (Chan et al., 2019). However, the collection
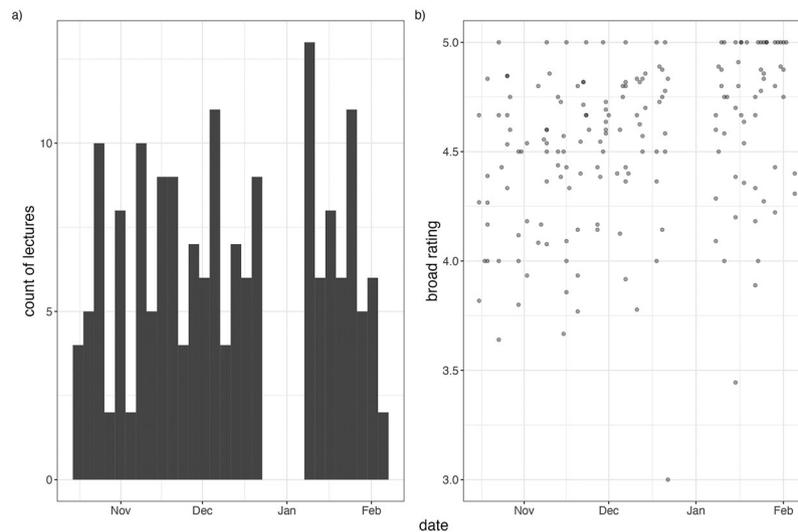
Figure 5. a) Histogram of lectures held per day during the study period. b) Date with average broad rating. Each dot represents a lecture, with darker dots indicating a greater number of lectures with the specified combination of values.

of data involving minors is legally not straightforward, so for this study only 56 school lessons could be rated. The opportunity to gather data from a high school setting represents a good opportunity for advancing our understanding of educational sciences. Nevertheless, the data clearly lack representativeness as they were collected from three classes taught by the same teacher at the same school. Accordingly, the hypotheses were analyzed from the university data, which will be contrasted with the school data in the section dedicated to hypothesis three below.

Figure 5a illustrates the number of lectures that were rated over the study period. There is no discernible trend over time, and the data do not indicate a decline towards the end of the study, as observed in previous research on student feedback (Hartmuth et al., 2025). On the right in Figure 5b, the lecture data are visualized together with the corresponding average broad rating. Each dot is a rated lecture and the darker the dot the more lectures obtained that rating on the same day. Remarkably, there were no average broad ratings below 3, and 75% of the ratings lay above 4.33 with a median of 4.60 and a standard deviation of 0.37. There was a slight increase in average lecture rating over time, however no structural change. As this study focused on the correlation between the question answers rather than the development of ratings over time, this aspect was not subjected to further investigation.

The associations of the average broad rating with the three questions about specific dimensions of teaching quality are depicted in Figure 6a (cognitive activation), b (constructive support) and c (classroom management), and with the rubric score in Figure 6d. The range of the dimensions cognitive activation and constructive support is similarly small and even smaller for the dimension classroom management and, hence, also the rubric score. All three dimensions and the rubric score were strongly positively correlated with the average broad rating.

Statistical analysis

A histogram of the number of lectures held per day during the study period in a scatter plot with each dot representing a lecture were used to detect changes in rating tendencies and rating behavior over time.

As the associations appear to be linear, to answer hypotheses 1 and 2, we used the Pearson correlation coefficient for further investigation. Confidence Intervals (CIs) were computed to denote the actual size of the correlation.

Moreover, a rating score, referred to as a rubric score, was calculated as the arithmetic mean of the three teaching dimensions, i.e. rubric score = (average cognitive activation + average constructive support + average classroom management)/3.

In order to answer hypothesis 3 we ran t-tests for independent random sampling.
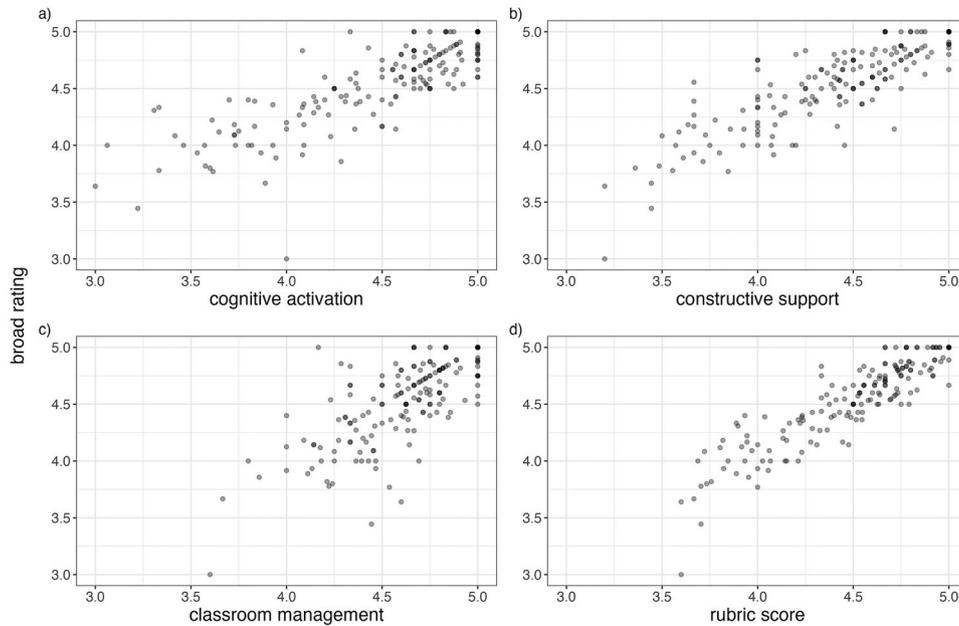
Figure 6. Average single question rating (broad rating -ordinate) plotted against Average rating of (abscissa) for a) cognitive activation, b) constructive support c) classroom management and d) the rubric score. Each dot represents a lecture with darker dots indicating a greater number of lectures with the specified combination of values.
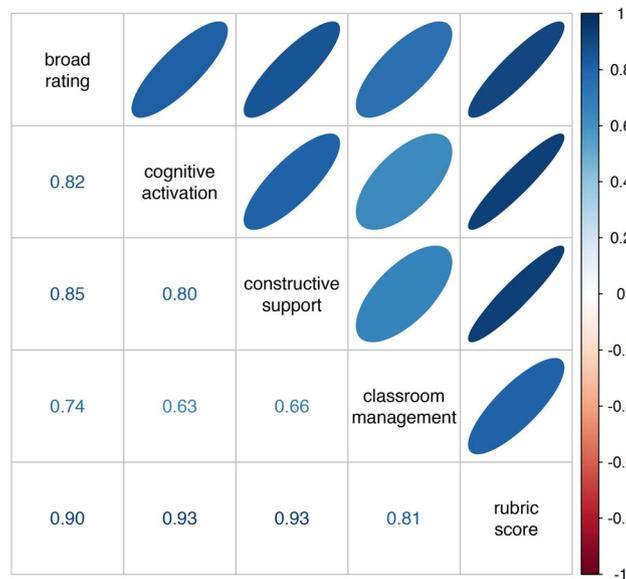
Figure 7. Pairwise Pearson correlation coefficients, both numerically and graphically for the university data (N = 184). The filled ellipses represent the correlation coefficients; a thin ascending ellipse represents a high correlation coefficient, a more circular ellipse a low correlation coefficient.

To give a graphic representation of our findings for hypothesis 3, boxplots for the four questions and the rubric score grouped by cohort were constructed.

Results

Correlation analysis

Figure 7 presents all-pairwise correlations of the Pearson correlation coefficient, both numerically and graphically. Four associations are of primary interest for hypotheses 1 and 2: the correlation of the

Table 1. Correlation coefficients and corresponding 95%-CI for the pairwise correlation between average broad rating and the averages of the three dimensions of teaching quality and the rubric score.

Correlation between average broad rating and …	Coefficient value	Lower 95%-CI bound	Upper 95%-CI bound
Cognitive activation	0.82	0.76	0.86
Constructive support	0.85	0.81	0.89
Classroom management	0.74	0.67	0.80
Rubric score	0.90	0.87	0.93

average broad rating with each of the dimensions of teaching quality and with the rubric score. They are shown in the first row and column of Figure 7. The remaining correlations are included for completeness.

Unsurprisingly, the values are all positive and rather large. Testing the hypothesis that the correlation coefficient differs from 0 will certainly yield a significant correlation. More insightful are confidence intervals (CIs) that specify an uncertainty region and suggest the actual size of the correlation. Table 1 denotes the coefficient values and lower and upper bounds of the 95%-CI for the four associations defined in hypotheses 1 and 2. In 95 of 100 cases such an interval includes the true correlation value. Calculations were performed in the open source software R and its function cor.test. Statistical details can be found in Fahrmeir et al. (2023).

In all cases, the value of zero is not covered by the interval corresponding to the coefficients significantly different from zero. The average ratings of the three dimensions of teaching quality are highly correlated with the average broad rating, especially cognitive activation and constructive support. The combination of each dimension, i.e. the rubric score, has an even larger positive association with the average broad rating.

Evaluation of hypothesis 1

Hypothesis 1 was about the degree of correlation between our single question student feedback (overall student feedback) and three questions of student feedback corresponding to the three main domains of teaching quality (cognitive activation, constructive support and classroom management). Our data show that the overall student feedback question result is highly and significantly correlated with all three questions corresponding to the three main domains of overall teaching quality.

For the comparison of the overall student feedback with the question corresponding to the domain cognitive activation we found a correlation coefficient of 0.82 with the 95% confidence interval ranging from 0.76 to 0.86. Thus, the null hypothesis can be rejected The observed correlation is very high, showing that measurement of both categories in situ yields comparable results.
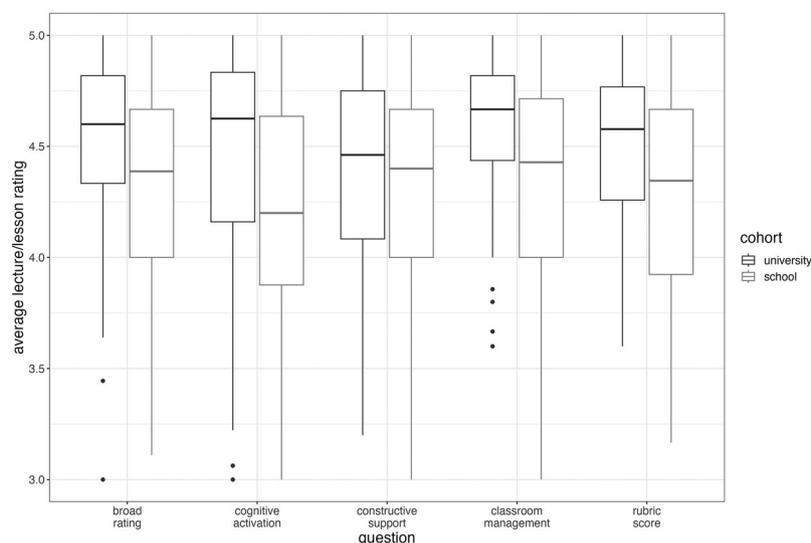
For the comparison of the overall student feedback with the question corresponding to the domain constructive support we found a correlation coefficient of 0.85 with the 95% confidence interval ranging from 0.81 to 0.89. Thus, the null hypothesis can be rejected. The observed correlation is very high, showing that measurement of both categories in situ yields comparable results.

For the comparison of the overall student feedback with the question corresponding to the domain classroom management we found a coefficient of correlation of 0.74 with the 95% confidence interval ranging from 0.67 to 0.80. Thus, the null hypothesis can be rejected. The observed correlation is very high, showing that measurement of both categories in situ yields comparable results.

Evaluation of hypothesis 2

For hypothesis 2 (whether the rubric score constructed from the three student feedback questions on the three domains of general teaching quality is correlated with the simple overall student feedback question),we opted for the simplest option to weight all three questions equally in constructing the rubric, from the myriad of ways of constructing such a score, by weighting the domains according to their contribution to the overall score.

Our data show that the overall student feedback question is highly and significantly correlated with the rubric score, higher than coefficients form the three questions corresponding to the three main overall teaching quality domains.

Figure 8. Boxplots for the four questions and the rubric score grouped by cohort.

For the comparison of the overall student feedback with the rubric score corresponding equally to all three domains of classroom management we found a correlation coefficient of 0.90 with the 95% confidence interval ranging from 0.87 to 0.93. Thus, the null hypothesis can be rejected. The observed correlation is extremely high, showing that measurement of either of both ways in situ yields comparable results.

Comparison of university and high school data

Figure 8 displays boxplots comparing the four questions and the score for the two cohorts. The line of greater thickness situated inside the box represents the median, the lower border marks the first quartile, and the upper border the third quartile. The box represents the central 50% of the data set, while the lines extending beyond the box encompass the remaining two quartiles. Points are used to indicate values that are outliers.

In all cases the median average ratings and the rubric score at school were lower than at university and the variance in the school data were slightly larger. Besides these differences, the values in both cohorts all lay above three and the size of the boxes was large, except for classroom management. Values for constructive support was similar in both cohorts. The associations among the variables were essentially similar to those observed in the university data. Figure 9 presents the analogous correlation matrix for the school data. The sole distinction between the two cohorts was that in classroom management the correlation was stronger than in the university data.

To ascertain whether there were discrepancies among the cohorts, t-tests for the means were conducted, and the results are in Table 2. The calculations were again performed using the open-source software R and its t-test function. The statistical details can be found in Fahrmeir et al. (2023). Note that the means are compared here, whereas the boxplots in Figure 8 depict the medians (Fahrmeir et al., 2023).

Evaluation of hypothesis 3

Hypothesis 3 was whether the two data sources in our study yielded comparable results.

Firstly, we looked at the correlation of the simple broad student feedback question with the other three feedback questions, analogous to the procedure described for RQ1 and RQ2.

For the comparison of the broad student feedback with the question corresponding to the domain cognitive activation we found a coefficient of correlation of 0.87; a high value indicating that measurement of both categories in situ yields comparable results.
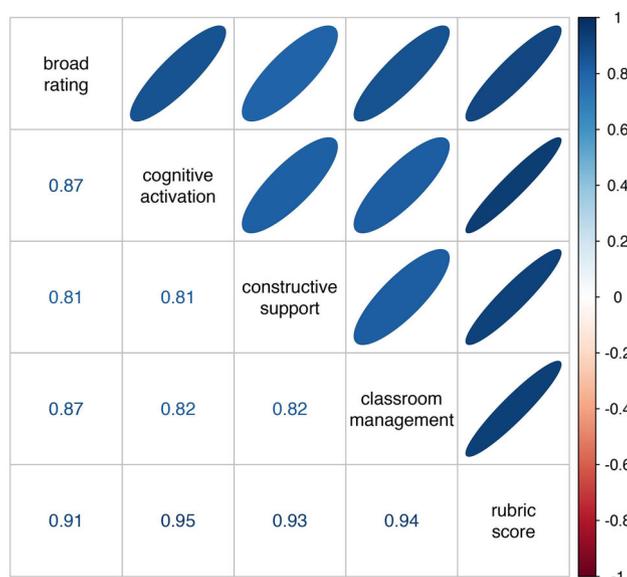
Figure 9. Pairwise Pearson correlation coefficients, both numerically and graphically for the school data (N = 71).

Table 2. Comparison of means between university and school data for the four questions and the rubric score and p-value if the means significantly (95%) differed from each other.

Question	Mean university	Mean high school	p-value
broad rating	4.54	4.30	0.002
Cognitive activation	4.47	4.17	0.001
Constructive support	4.40	4.32	0.215
Classroom management	4.62	4.34	<0.001
Rubric score	4.50	4.28	0.002

For the comparison of the broad student feedback with the question corresponding to the domain constructive support we found a correlation R = 0.81.

For the comparison of the broad student feedback with the question corresponding to the domain classroom management we found R = 0.87.

These findings parallel our findings from the university data set.

Secondly, we compared rating behavior on all four student feedback questions, as well as with our calculated rubric score.

For the comparison of the mean ratings of the broad student feedback question, results showed that students at university provided significantly higher ratings (mean 4.54) than high school pupils (mean 4.30).

For the comparison of the mean ratings of the student feedback question corresponding to the dimension cognitive activation results showed that students at university significantly rated higher (mean 4.47) than those at high school (mean 4.17).

For the comparison of the mean ratings of the student feedback question corresponding to the dimension constructive support results showed that students at university rate slightly, but not significantly higher (mean 4.40) than school pupils (mean of 4.32).

For the comparison of the mean ratings of the student feedback question corresponding to the dimension classroom management results showed that students at university rated significantly higher (mean 4.62) than pupils at high school (mean 4.34).

For the comparison of the grand rubric or score corresponding equally to all three domains of teaching quality results showed that students at university significantly rated higher (mean 4.50) than those in school (mean 4.28).

The null hypothesis could be rejected in all cases except for the student rating question corresponding to the domain constructive support, and we conclude that in general students at university show comparable rating behavior for all evaluated student rating questions, but generally rated teaching quality more positively than pupils at school.

Discussion

The aim of this study was to investigate whether a simple overall student feedback yields valuable information on teaching quality and the measurement and development of a rubric score of ePCK in the classroom environment.

In the following, results referring to the three hypotheses are interpreted and discussed.

Correlation of overall student feedback with domains of teaching quality

Our study on student feedback for science teachers delivered valuable insights. Our data show that if student feedback is collected regularly, ratings between the posed questions will correlate highly. We conclude that students are able to evaluate teaching quality in a given lesson with a simple overall student feedback question just as well as with a more complex questionnaire on different domains of teaching quality. Taking in situ restraints such as sparsity of time and complexity of interpretation of complex data into account we argue that a simple overall student feedback question is preferable to a series of more complex student feedback questions or a grand rubric (Chan et al., 2019). We also conclude that using relative data (change of performance) instead of absolute values in analyzing student feedback is a suitable option, drawing focus towards development of ePCK and away from specific facets of teaching quality that will all yield comparable relative results when measured in the classroom (Davidowitz & Potgieter, 2016).

Correlation of overall student feedback with the rubric score

The highest correlation found was between the overall student feedback and our rubric score. The purpose of a rubric score is to guide instruction and improvement, to facilitate self-assessment, and to enhance communication (Chan et al., 2019). This is achieved by providing consistent evaluation and clear criteria and standards. Collecting data on the various facets of teaching quality and creating a rubric score provides a reliable indicator of a teacher's ePCK (Shavelson, 2020). Our data show that this score parallels our overall student feedback results. Considering that the rubric score itself expresses a measure of ePCK, and not the items it is constructed from (Park & Kyung Suh, 2015), we suggest that it may be just as useful and more feasible to collect overall student feedback instead. Especially in a dynamic environment, where ePCK is seen as an emergent property, the total rubric score is the information with the highest predictive value on future performance (Kahneman et al., 2021), hence what is needed for a teacher

Comparison of student feedback between university and high school

Our data suggest that using overall student feedback in both high school and university environments provide similar pre-service teacher advice. This finding is in line with previous studies comparing feedback in both environments (Seidel & Shavelson, 2007). The observation that mean values in university are higher than those from high school may be due to factors we did not control for in our study such as class size or grading (Wisniewski & Zierer, 2021)When relative performance is used to reflect development of ePCK, the intercept or absolute values of ePCK are not considered directly.

Influence of student feedback on ePCK development

Overall student feedback is an easy-to-collect proxy for the ePCK of a teacher in a specific lesson (Rollett et al., 2021). It mimics to a high degree a rubric score, which is more time-consuming to collect. Moreover overall student feedback can be used to track teacher ePCK over time and over the course of several lessons (Bijlsma et al., 2022). Measuring ePCK in each lesson shows a teacher in real time his or her teaching performance, and this information is what a teacher needs to adapt his or her plan-teach-reflect cycle and improve by iteration (Carlson et al., 2019). Making information on ePCK accessible for

individual teachers gives teachers the opportunity to try out different teaching strategies and receive immediate information on the success of their efforts (Manzi et al., 2022).

Limitations

In this study the number of school lessons in which student feedback was collected was considerably lower than, and not as representative as, the broader database for the university. In order to draw conclusions for the high school setting with the same confidence as for the university setting, a larger study with more participants would be beneficial.

However, our high-school dataset comprises only 56 lessons taught by a single teacher, whereas the university sample covers 175 lectures taught by multiple instructors, which limits the generalisability of our findings across teaching contexts.

Despite the small sample size for our high school data set, which can affect the validity of results (Döring & Bortz, 2016), we could report reliable results for hypothesis 3 and derive generalizations.

Not all research on ePCK focuses on the dynamics of the teaching environment and not all research groups portray ePCK as an emergent property, where the construct is more than the sum of its parts. When ePCK is framed as a more static construct of facets that can be summarised, the student feedback on those facets (in our study cognitive activation, constructive support and classroom management) becomes more valuable, and cannot easily be ignored or substituted.

Although our results show a strong correlation between a single feedback item and more elaborate rubrics, relying solely on a one-question format risks oversimplifying the feedback process. In reality, the precision of student evaluations may also hinge on factors such as learners' cognitive maturity, their familiarity with the rubric's dimensions, and cultural or motivational biases that shape how they rate instruction.

Overall, this study supports the idea of using simple overall student feedback to gauge and track individual teacher's ePCK in an adaptive feedback cycle and encourage further investigation in this field.

Ethics statement

Informed consent was obtained from all subjects involved in the study. No personalized data were collected during the study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Authors

Daniel Hartmuth is teacher on secondment and PhD candidate at the Institute of Biology Education of the Faculty for Biology, LMU Munich.

Daniel Schlichting is a Masters student and student assistant of StabLab (Statistical Consulting Unit) at LMU Munich.

Jonathan Haudenschild has a Masters degree in Human-Computer Interaction and works as a programmer at the Institue of Biology Education oft he Faculty for Biology, LMU Munich.

Helmut Küchenhoff is a Professor and Head of StabLab (Statistical Consulting Unit) at LMU Munich.

Birgit J. Neuhaus is a Professor and chair holder of the Chair for Biology Education at LMU Munich, Germany. Her research interests are mainly connected to biology teachers' professional competence, instructional quality of biology lessons and simulations in biology teacher education.

ORCID

Daniel Hartmuth http://orcid.org/0009-0002-8063-307X
Helmut Küchenhoff http://orcid.org/0000-0002-6372-2487
Birgit J. Neuhaus http://orcid.org/0000-0001-6031-9660

Data availability statement

References

Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 273–288). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-5898-2_12

Alonzo, A. C., Kobarg, M., & Seidel, T. (2012). Pedagogical content knowledge as reflected in teacher–student interactions: Analysis of two video cases. *Journal of Research in Science Teaching*, *49*(10), 1211–1239. https://doi.org/10.1002/tea.21055

Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sciences Education*, *5*(3), 197–203. https://doi.org/10.1187/cbe.06-06-0168

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment*, *38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

Atroszko, B., Czerwiński, S. K., & Uzarska, A. (2019). Single-item self-report measure of learning engagement: What does it measure?. In *Reviewed Proceedings of the Interdisciplinary Scientific International Conference for PhD Students and Assistants QUAERE 2019* (Vol. IX, pp. 806–816). Magnanimitas.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U. [Uta], Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Behling, F., Förtsch, C., & Neuhaus, B. J. [Birgit J.] (2022). Using the plan–teach–reflect cycle of the refined consensus model of PCK to improve pre-service biology teachers' personal PCK as well as their motivational orientations. *Education Sciences*, *12*(10), 654. https://doi.org/10.3390/educsci12100654

Behnke, K. (2016). *Umgang mit Feedback im Kontext Schule. [Dealing with feedback in the school context.]*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-10223-4

Berry, A., Friedrichsen, P., & Loughran, J. (Eds.) (2015). *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education: October 2012, … Pck Summit Working conference in Colorado Springs (First published.)*. Routledge.

Bijlsma, H., van der Lans, R., Mainhard, T., & Brok, P. D (2021). A reflection on student perceptions of teaching quality from three psychometric perspectives: CCT, IRT and GT. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools* (pp. 15–29). Springer International Publishing. https://doi.org/10.1007/978-3-030-75150-0_2

Bijlsma, H. J. E., Glas, C. A. W., & Visscher, A. J. (2022). Factors related to differences in digitally measured student perceptions of teaching quality. *School Effectiveness and School Improvement*, *33*(3), 360–380. https://doi.org/10.1080/09243453.2021.2023584

Bijlsma, H. J., Visscher, A. J., Dobbelaer, M. J., & Veldkamp, B. P. (2019). Does smartphone-assisted student feedback affect teachers' teaching quality? *Technology, Pedagogy and Education*, *28*(2), 217–236. https://doi.org/10.1080/1475939X.2019.1572534

Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *Advances in mathematics education. International perspectives on teacher knowledge, beliefs and opportunities to learn* (pp. 19–47). Springer Netherlands. https://doi.org/10.1007/978-94-007-6437-8_2

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., … Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–94). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-5898-2_2

Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 3–76). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-5898-2_1

Chan, K. K. H., Rollnick, M., & Gess-Newsome, J. (2019). A grand rubric for measuring science teachers' pedagogical content knowledge. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 253–271). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-5898-2_11

Chan, K. K. H., & Yung, B. H. W. (2018). Developing pedagogical content knowledge for teaching a new topic: More than teaching experience and subject matter knowledge. *Research in Science Education*, 48(2), 233–265. https://doi.org/10.1007/s11165-016-9567-1

Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. *International Journal of Science Education*, 38(9), 1483–1503. https://doi.org/10.1080/09500693.2016.1196843

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften [Research Methods and Evaluation in the Social and Human Sciences] (5., vollständig überarb. Aufl.)*. Springer.

Dorfner, T., Förtsch, C., Boone, W., & Neuhaus, B. J. [. J.]. (2019). Instructional quality features in videotaped biology lessons: Content-independent description of characteristics. *Research in Science Education*, 49(5), 1457–1491. https://doi.org/10.1007/s11165-017-9663-x

Dorfner, T., Förtsch, C., & Neuhaus, B. J. [. J.]. (2018). Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction. *Learning and Instruction*, 56, 42–53. https://doi.org/10.1016/j.learninstruc.2018.03.001

Drits-Esser, D., Gess-Newsome, J., & Stark, L. A. (2017). Examining the sustainability of teacher learning following a year-long science professional development programme for inservice primary school teachers. *Professional Development in Education*, 43(3), 375–396. https://doi.org/10.1080/19415257.2016.1179664

Elstad, E., Lejonberg, E., & Christophersen, K. -A. (2017). Student evaluation of high-school teaching: Which factors are associated with teachers' perception of the usefulness of being evaluated? *Journal for Educational Research Online*, 9(1), 99–117. https://doi.org/10.25656/01:12968

Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., & Tutz, G. (2023). *Statistik. [Statistics.]*. Springer. https://doi.org/10.1007/978-3-662-67526-7

Förtsch, C., Werner, S., Kotzebue, L. V., & Neuhaus, B. J. [J.]. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, 38(17), 2642–2666. https://doi.org/10.1080/09500693.2016.1257170

Förtsch, C., Werner, S., Kotzebue, L. V., & Neuhaus, B. J. [J.]. (2018). Effects of high-complexity and high-cognitive-level instructional tasks in biology lessons on students' factual and conceptual knowledge. *Research in Science & Technological Education*, 36(3), 353–374. https://doi.org/10.1080/02635143.2017.1394286

Funnell, P. (2017). Using audience response systems to enhance student engagement and learning in information literacy teaching. *Journal of Information Literacy*, 11(2), 28–50. https://doi.org/10.11645/11.2.2238

Gess-Newsome, J., & Lederman, N. G. (Eds.) (1999). *Science & technology education library. Vol. 6. Examining pedagogical content knowledge: The construct and its implications for science education*. Kluwer Academic Publ. http://www.loc.gov/catdir/enhancements/fy0813/99040998-d.html

Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D., & Stuhlsatz, M. A. M. (2019). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, 41(7), 944–963. https://doi.org/10.1080/09500693.2016.1265158

Göllner, R., Fauth, B., & Wagner, W. (2021). Student ratings of teaching quality dimensions: Empirical findings and future directions. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools* (pp. 111–122). Springer International Publishing. https://doi.org/10.1007/978-3-030-75150-0_7

Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types. *Journal of Research on Technology in Education*, 41(4), 393–416. https://doi.org/10.1080/15391523.2009.10782536

Hartmuth, D., Amin, R. M., Schlichting, D., Schött, S., Haudenschild, J., Küchenhoff, H., & Neuhaus, B. J. (2025). Developing science teachers' enacted pedagogical content knowledge through integration of student feedback into the Refined Consensus Model of pedagogical content knowledge. *Cogent Education*, 12(1), 2470563. https://doi.org/10.1080/2331186X.2025.2470563

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://www.taylorfrancis.com/books/9780203887332 https://doi.org/10.4324/9780203887332

Hattie, J., & Clarke, S. (2019). *Visible learning: Feedback*. Routledge. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5584342

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Heinitz, B., Szogs, M., Förtsch, C., Korneck, F., Neuhaus, B. J. [. J.]., & Nehring, A. (2022). Unterrichtsqualität in den Naturwissenschaften. Eine vergleichende Gegenüberstellung von Ansätzen zwischen Fachspezifik und Generik. [Teaching quality in the natural sciences. A comparative comparison of approaches between subject-specific and generic approaches]. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 28(1), 169–186. https://doi.org/10.1007/s40573-022-00146-5

Hense, J. (Ed.) (2013). *Forschung über Evaluation: Bedingungen, Prozesse und Wirkungen. [Research on evaluation: conditions, processes and effects.]*. Waxmann. https://elibrary.utb.de/doi/book/10.31244/9783830978299

Irmer, M., Traub, D., Böhm, M., Förtsch, C., & Neuhaus, B. J. [J.]. (2023). Using video-based simulations to foster pPCK/ePCK—new thoughts on the refined consensus model of PCK. *Education Sciences*, 13(3), 261. https://doi.org/10.3390/educsci13030261

Irmer, M., Traub, D., & Neuhaus, B. J. (2024). Fostering ePCK in pre-service biology teacher education using adaptive scaffolding in a video-based simulation. *Cogent Education*, *11*(1), 2422272. https://doi.org/10.1080/2331186X.2024.2422272

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. William Collins.

Kämpfe, N. (2009). Schülerinnen und Schüler als Experten für Unterricht [Students as experts in instruction]. *DDS – Die Deutsche Schule*, *101*(2), 149–163. https://doi.org/10.25656/01:25587

Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, *54*(5), 586–614. https://doi.org/10.1002/tea.21378

Koehler, M. J., Shin, T. S., & Mishra, P. (2012). How do we measure TPACK? Let me count the ways. In R. N. Ronau, C. R. Rakes, & M. L. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact* (pp. 16–31). IGI Global. https://doi.org/10.4018/978-1-60960-750-0.ch002

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education*, *4*(2), 193–212. https://doi.org/10.5465/amle.2005.17268566

Kost, R. G., & da Rosa, J. C. (2018). Impact of survey length and compensation on validity, reliability, and sample characteristics for ultrashort-, short-, and long-research participant perception surveys. *Journal of Clinical and Translational Science*, *2*(1), 31–37. https://doi.org/10.1017/cts.2018.18

Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. [. J.]. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, *7*(1), 1823155. https://doi.org/10.1080/2331186X.2020.1823155

Kunter, M., & Voss, T. (2013). The model of instructional quality in COACTIV: A multicriteria analysis. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 97–124). Springer US. https://doi.org/10.1007/978-1-4614-5149-5_6

Mahler, D., Großschedl, J., & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, *39*(2), 213–237. https://doi.org/10.1080/09500693.2016.1276641

Manzi, J., Sun, Y., & García, M. R. (2022). *Teacher evaluation around the world*. Springer International Publishing. https://doi.org/10.1007/978-3-031-13639-9

Mientus, L., Hume, A., Wulff, P., Meiners, A., & Borowski, A. (2022). Modelling STEM teachers' pedagogical content knowledge in the framework of the refined consensus model: A systematic literature review. *Education Sciences*, *12*(6), 385. https://doi.org/10.3390/educsci12060385

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Morine-Dershimer, G., & Kent, T. (1999). The complex nature and sources of teachers' pedagogical knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Science & technology education library: Vol. 6. Examining pedagogical content knowledge: The construct and its implications for science education* (Vol. 6, pp. 21–50). Kluwer Academic Publ. https://doi.org/10.1007/0-306-47217-1_2

Morris, T. H. (2020). Experiential learning – a systematic review and revision of Kolb's model. *Interactive Learning Environments*, *28*(8), 1064–1077. https://doi.org/10.1080/10494820.2019.1570279

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, *16*(4), 310–322. https://doi.org/10.1016/j.learninstruc.2006.07.003

Neuhaus, B. J. [J.]. (2021). Unterrichtsqualität aus der Perspektive der Biologiedidaktik. [Teaching quality from the perspective of biology didactics.]. *Unterrichtswissenschaft*, *49*(2), 273–283. https://doi.org/10.1007/s42010-021-00114-x

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, *30*, 100329. https://doi.org/10.1016/j.edurev.2020.100329

Park, S. (2019). Reconciliation between the refined consensus model of PCK and extant PCK models for advancing PCK research in science. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 119–130). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-5898-2_4

Park, S., & Kyung Suh, J. (2015). From portraying toward assessing PCK: Drivers, dilemmas, and directions for future research. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Teaching and learning in science series. Re-examining pedagogical content knowledge in science education* (pp. 114–129). Routledge. https://doi.org/10.4324/9781315735665-14

Praetorius, A. -K., & Gräsel, C. (2021). Noch immer auf der Suche nach dem heiligen Gral: Wie generisch oder fachspezifisch sind Dimensionen der Unterrichtsqualität? [Still searching for the holy grail: How generic or subject-specific are dimensions of teaching quality?]. *Unterrichtswissenschaft*, *49*(2), 167–188. https://doi.org/10.1007/s42010-021-00119-6

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

Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, *81*(12), 46–54, 124.

Rollett, W., Bijlsma, H., & Röhl, S. (Eds.) (2021). *Student feedback on teaching in schools*. Springer International Publishing. https://doi.org/10.1007/978-3-030-75150-0

Ruth-Herbein, E., Maier, J. L., & Fauth, B. (2022). Promoting teaching quality through classroom observation and feedback: Design of a program in the German state of Baden-Württemberg. In J. Manzi, Y. Sun, & M. R. García (Eds.), *Teacher education, learning innovation and accountability. Teacher evaluation around the world* (pp. 271–289). Springer International Publishing. https://doi.org/10.1007/978-3-031-13639-9_12

Schmidt, T., Gazou, A., Rieß, A., Rieß, O., Grundmann-Hauser, K., Falb, R., Schadeck, M., Heinrich, T., Abeditashi, M., Schmidt, J., Mau-Holzmann, U. A., & Schnabel, K. P. (2020). The impact of an audience response system on a summative assessment, a controlled field study. *BMC Medical Education*, *20*(1), 218. https://doi.org/10.1186/s12909-020-02130-4

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, *51*(4), 739–771. https://doi.org/10.3102/0002831214531321

Shavelson, R. J. (2012). Assessing business-planning competence using the Collegiate Learning Assessment as a prototype. *Empirical Research in Vocational Education and Training*, *4*(1), 77–90. https://doi.org/10.1007/BF03546509

Shavelson, R. J. (2020). Research on teaching and the education of teachers: Brokering the gap. *BzL - Beiträge Zur Lehrerinnen- Und Lehrerbildung*, *38*(1), 37–53. https://doi.org/10.36950/bzl.38.1.2020.9310

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Sorge, S., Kröger, J., Petersen, S., & Neumann, K. (2019). Structure and development of pre-service physics teachers' professional knowledge. *International Journal of Science Education*, *41*(7), 862–889. https://doi.org/10.1080/09500693.2017.1346326

Steffensky, M., Neuhaus, B. J., & Birgit, J. (2018). Unterrichtsqualität im naturwissenschaftlichen Unterricht. [Teaching quality in science lessons.]. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (pp. 299–313). Springer. https://doi.org/10.1007/978-3-662-56320-5_18

Suresh, A., Jacobs, J., Harty, C., Perkoff, M., Martin, J. H., & Sumner, T. (2022). The talkmoves dataset: K-12 mathematics lesson transcripts annotated for teacher and student discursive moves. https://doi.org/10.48550/arXiv.2204.09652

Suresh, A., Sumner, T., Huang, I., Jacobs, J., Foland, B., & Ward, W. (2018). Using deep learning to automatically detect talk moves in teachers' mathematics lessons. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 5445–5447). IEEE. https://doi.org/10.1109/BigData.2018.8621901

Tobin, K., & McRobbie, C. (1999). Pedagogical content knowledge and co-participation in science classrooms. In J. Gess-Newsome & N. G. Lederman (Eds.), *Science & technology education library: Vol. 6. Examining pedagogical content knowledge: The construct and its implications for science education* (Vol. 6, pp. 215–234). Kluwer Academic Publ. https://doi.org/10.1007/0-306-47217-1_9

van der Scheer, E. A., Bijlsma, H. J. E., & Glas, C. A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement*, *30*(1), 30–50. https://doi.org/10.1080/09243453.2018.1539015

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*(3), 249–294. https://doi.org/10.3102/00346543063003249

Wisniewski, B., & Zierer, K. (2021). Functions and success conditions of student feedback in the development of teaching and teachers. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools* (pp. 125–138). Springer International Publishing. https://doi.org/10.1007/978-3-030-75150-0_8

Wisniewski, B., Zierer, K., Dresel, M., & Daumiller, M. (2020). Obtaining secondary students' perceptions of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction*, *66*, 101303. https://doi.org/10.1016/j.learninstruc.2020.101303

Wisniewski, B., Zierer, K., & Hattie, J. (2018). *Visible Feedback: Ein Leitfaden für erfolgreiches Unterrichtsfeedback (2. erweiterte Auflage)*. Schneider Verlag Hohengehren GmbH.

Wisniewski, B., Zierer, K., & Hattie, J. (2019). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087