ZDM - Mathematics Education (2025) 57:1235–1250
https://doi.org/10.1007/s11858-025-01748-2

ORIGINAL PAPER

Does more teaching experience mean more expertise? A nonlinear perspective from the COACTIV study

Stefan Krauss[1] · Stefanie Hohenleitner[1] · Nicole Steib[1] · Alfred Lindl[2] · Sven Hilbert[2]

Received: 27 December 2024 / Accepted: 25 September 2025 / Published online: 16 October 2025

Abstract

In expert–novice research paradigms on the teaching profession, the length of experience is—despite many warnings—used quite often as a key criterion for assigning teachers to the expert group. Using data from the COACTIV project (a study on the professional competence of mathematics teachers in PISA classes) we analyzed the relationship between 13 expertise indicators and length of experience (i.e., years of teaching). By ordering all COACTIV teachers regarding each of the 13 indicators and considering the median rank of a teacher according to these indicators, we were able to form an overall expertise (OE) index. When modeling linear correlations, the 13 indicators as well as the OE index showed negative or zero correlations with length of experience. When allowing for nonlinear models, we identified slight differential relationships between years of teaching and OE within three distinct intervals: a positive relation during the first 8 years of teaching, a negative relation in midcareer, and again a positive relation in the last 10–15 years. Similar curves also emerged in subsamples (academic- vs. vocational-track teachers). We discuss these findings in light of the cross-sectional nature of the COACTIV data. Our results call into question if long experience is a sufficient condition for becoming an expert teacher but strikingly also contradict the widely held view that professional experience is a necessary condition.

Keywords COACTIV · Teacher · Professional competence · Expertise · Experience

1 Introduction

The question of what makes a good teacher is widely discussed in international literature under the concept of *teacher expertise* or, particularly in German-speaking countries, *teachers' professional competence* (for an overview of research paradigms see Table A1 in the online supplementary information). Various theoretical models describe teacher expertise, instructional quality, or the postulated causal chain from teachers' professional competence to students' learning gains (for an "integrative" model, see, e.g., the cascade model, Fig. 1a). In their recent meta-review on expertise of teachers, Anderson and Taner (2023) highlighted two key desiderata for future research: (1) "to investigate potentially causative relationships between features of the expert teacher prototype," and (2) to analyze "the extent to which adaptive expertise is dependent on experience" (p. 14). The call to analyze the dependency of expertise on professional experience automatically raises the question of which facets of expertise should be considered. Anderson and Taner (2023), for instance, identified 73 different features of expert teachers. Consequently, research must address how these facets relate to one another and whether they are linked to professional experience in a comparable way.

In individual empirical studies, typically only a few expertise indicators of teachers are assessed and examined, owing to high costs, time constraints, or insufficient field access, especially over longer periods of time. For the same reason, longitudinal studies that follow teachers over multiple years—and thus allow for reliable conclusions about intraindividual long-term developments—are largely lacking. Consequently, study samples in this field often include preservice teachers, for whom it is not (yet) possible to simultaneously assess the quality of their teaching and the learning progress of their students. Accordingly, it is difficult to empirically link relevant indicators of teacher expertise within

✉ S. Krauss
stefan.krauss@ur.de

1 Faculty of Mathematics, University of Regensburg, Regensburg, Germany

2 Faculty of Human Sciences, University of Regensburg, Regensburg, Germany

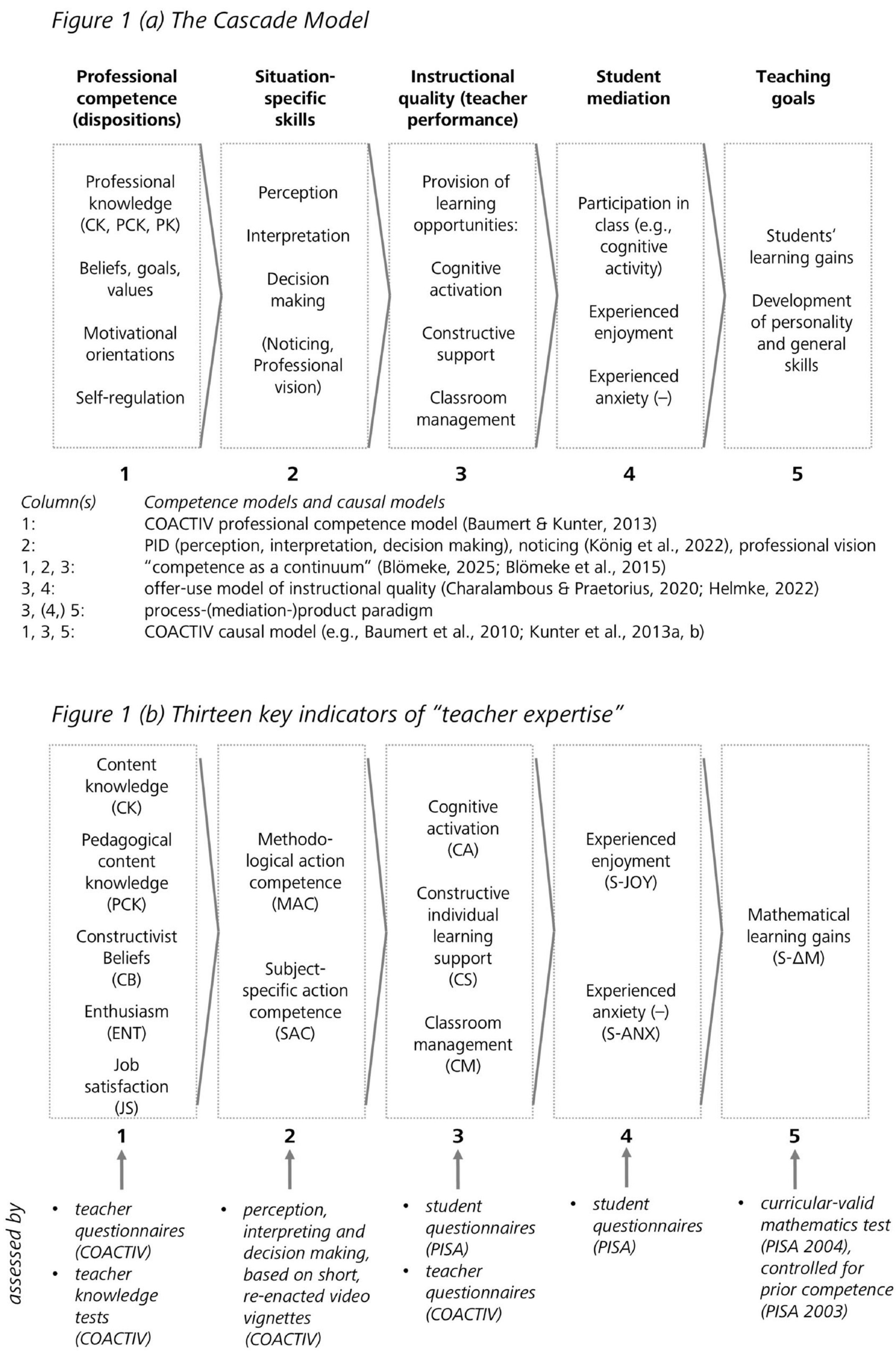

Fig. 1 **(a)** The cascade model (adapted from Krauss et al., 2020). **(b)** Thirteen corresponding key indicators of "teacher expertise" (including abbreviations) that were assessed in PISA/COACTIV 03/04

a single study, either to one another or to student-related outcome measures.

In this paper, we address the relation between expertise and experience from the perspective of the COACTIV research program ("Professional Competence of Teachers, Cognitively Activating Instruction, and the Development of Students' Mathematical Literacy"; Kunter et al., 2013a). In the COACTIV study, more than 1000 variables regarding mathematics teachers (including years of experience), their teaching in PISA (Program for International Student Assessment) classes, and their students were measured (Baumert et al., 2009). For our analyses, we selected 13 expertise indicators identified in previous publications as relevant for or connected to high-quality teaching, which have so far been analyzed only separately or in smaller groups (e.g., Kunter et al., 2013a). These indicators stem from various stages of typical causal models (e.g., Fig. 1a; see Fig. 1b for the implementation of the 13 constructs in the cascade model) and, historically, from different research paradigms (Table A1 in the online supplementary information).

To give all indicators the same weight (and for an easier interpretation and comparison), we transformed each indicator scale to a percentile rank scale (0 = lowest expertise; 100 = highest expertise). To reduce complexity, we furthermore created an overall expertise (OE) index for each teacher by considering the median of these 13 percentile scales and investigated the relationship to professional experience both for the individual expertise indicators and for the OE index. We considered the distribution of expertise particularly in the first years of teaching (Dreyfus, 2004; Stigler & Miller, 2018) but also in later career stages—an aspect that has been a blind spot in teacher expertise research.

2 Which indicators of expertise should be considered?

2.1 The perspective of expertise research

Expertise research from cognitive psychology has shown that in numerous fields, experts differ from laypersons primarily because of their well-organized, flexible, and quickly applicable *domain-specific knowledge* (Gruber et al., 2010). Furthermore, experts *perceive* domain-specific situations differently from novices: They immediately recognize meaningful deep structures, whereas novices notice only superficially visible features in the same situations (e.g., Brams et al., 2019).

In the 1980s, general expertise research was applied to the teaching profession (Krauss & Bruckmaier, 2014). The expertise paradigm provided the theoretical basis for research on the *professional knowledge* of teachers (Fig. 1a, column 1) and for the investigation of their *professional perception* ("noticing") of teaching situations (Fig. 1a, column 2). Both areas are now well established in teacher research.

The intraindividual development of expertise is a particular focus of the general expertise paradigm (Ericsson et al., 1993), including studies of cognitive internal processes such as knowledge restructuring when becoming an expert (e.g., Boshuizen et al., 2020). Because the expertise paradigm frequently concentrates on intraindividual change processes across the professional lifespan, methodologically it is based predominantly on qualitative research (e.g., Anderson & Taner, 2023).

The idea of the *expert–novice paradigm* is to explicitly compare a group of experts with a group of novices to detect further crucial cognitive or behavioral differences between the groups empirically. In this paradigm, the researcher has to decide which teachers are experts (and which are not) *before* starting the study. If the expert status is not clear by objective standards—as it is in well-structured domains, such as chess, where players have Elo ratings—this requirement almost inevitably leads to a "chicken-and-egg" problem. However, after several decades of using expert–novice paradigms in the ill-structured domain of teaching, a few criteria have emerged that, in part, also guided the present study. In such paradigms, typically one (or a combination) of the following criteria is used to preselect expert teachers (e.g., Bromme, 2008; Palmer et al., 2005):

(1) Level of education (e.g., student teachers vs. trainee teachers vs. practicing teachers),
(2) Professional achievements (e.g., professional promotions or certificates),
(3) Nominations by peers or supervisors (social recognition),
(4) Teaching quality (e.g., evaluation by students or observers),
(5) Improvement of students' performance (or other teaching target criteria),
(6) Teaching experience (by length of employment).

When teachers were selected according to these criteria (i.e., expert teachers vs. novices), substantial differences concerning many further constructs (e.g., those shown in Fig. 1) were identified between the two groups in multiple expert–novice studies (Anderson & Taner, 2023). In sum, we refer to all variables that may allow distinctions between experts and novices (i.e., criteria 1–5), as well as all constructs shown in Fig. 1, as (potential) *indicators of expertise*.

However, with regard to criterion 6, teaching experience, several authors have cautioned against simply equating experience with expertise (e.g., Berliner, 2001; Palmer et al., 2005; Stigler & Miller, 2018). As Stigler and Miller (2018) noted: "The problem with conflating experience and expertise has long been recognized (e.g., Berliner, 1986). Despite this, experience still has been the main variable used to indicate expertise in teaching, at least until recently" (p. 436). For instance, Caspari-Sadeghi and König (2018) found that one third of their analyzed expertise studies used experience as a criterion for preselecting "expert teachers".

Consequently, performance measures should be included when identifying experts (Berliner, 2001). Note that in the cascade model (Fig. 1), teaching quality (column 3) relates to *teacher performance* (i.e., the provision of learning opportunities), while also the improvement of *students' performance* (column 5) should be considered.

2.2 The perspective of the professional competence paradigm

German large-scale studies, such as COACTIV (Kunter et al., 2013a) and TEDS (Teacher Education and Development Study; Blömeke et al. 2020, 2022), adopted the initial focus on professional knowledge and perception from expertise research. Rather than qualitatively investigating intraindividual development or contrasting experts and novices, these studies constructed psychometric instruments for mathematics teachers and conducted quantitative statistical analyses on samples covering all levels of teacher competencies (rather than considering only "extreme groups" such as experts and novices). In both research programs, tests were developed based on Shulman's (1986, 1987) theoretical taxonomy of teacher knowledge, which contains the categories content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical knowledge (PK). In addition, both studies implemented video vignettes to capture teachers' professional perceptions of critical lesson scenes (e.g., Bruckmaier et al., 2016; Kaiser et al., 2017; Krauss et al., 2008a,b).

However, further aspects, many of which are more closely associated with personality, play a role in teaching success and should therefore be integrated into an overall model of important teacher characteristics. The left column in Fig. 1a, for instance, displays the four overarching constructs of the COACTIV model of *professional teacher* (*action*) *competence*. Because expertise research (see Sect. 2.1) traditionally focuses on the development of expertise, while empirical evidence on the malleability of teachers' motivational orientations or self-regulatory skills through education and training remains less clear, the classic notion of expertise in Germany has largely been replaced by the concept of teacher competence (Weinert, 2001). Thus, in general, both the expertise and competence concepts refer to—with slightly differing accentuations—teacher characteristics (Fig. 1a, first two columns) that are responsible for high-quality teaching and students' learning gains. In the present article, *indicators* of expertise then are the named competencies as well as their consequences, such as manifestations in teaching and learning, which can stem from all five columns in the model.

The notion of competence was extended in the international TEDS framework (e.g., Blömeke et al., 2022; Kaiser & König, 2020). Blömeke et al. (2015) proposed considering competence as a continuum (i.e., column 1–3 in Fig. 1a). In this generic model, perception-related situation-specific competencies (column 2) function as mediators between dispositions (column 1) and teacher performance (column 3). Situation-specific competencies, also discussed under concepts such as professional vision (Seidel & Stürmer, 2014) or noticing (König et al., 2022), describe the awareness of and appropriate responses to critical situations in the classroom (perception, interpretation, and decision making; Blömeke et al., 2015). Such competencies are often assessed using short video vignettes showing lesson scenes as stimuli (in contrast to the typical paper-and-pencil assessment of, e.g., teacher knowledge categories). Under the concept of competence as a continuum, it has been suggested that the actual *performance* of the teacher (i.e., the teaching action, column 3) can also be regarded as competence in a broader sense (Blömeke, 2025).

The competence paradigm—at least as it is typically interpreted in Germany—focuses neither on the intraindividual development of teachers nor on differences between experts and novices. It is instead concerned with *constructs and their mutual dependencies*, which are often analyzed quantitatively. These models are methodologically based on representative random samples of teacher populations. The statistical analyses—correlational by nature—then are typically based on the whole teacher sample (i.e., across all competence levels). Because no specific teachers are preselected in this paradigm (e.g., experts or novices), it implicitly avoids the challenge of defining the "expert teacher."

A key method of quantitative large-scale studies is the implementation of path models (PMs) or structural equation models (SEMs) to evaluate the impact of certain constructs within one area (i.e., columns in Fig. 1a) on constructs in subsequent areas. This could, for example, be the predictive validity of professional knowledge (column 1) or of professional perception (column 2) for aspects of teacher performance (column 3) or students' learning gains (column 5). Teacher performance is often referred to as instructional quality in German studies and modeled by the three basic dimensions: *cognitive activation*, *constructive individual learning support*, and *classroom management* (e.g., Kunter & Voss, 2013; Praetorius et al., 2018). Several studies have documented effects of professional knowledge, constructivist beliefs, and enthusiasm on aspects of instructional quality (e.g., Blömeke et al., 2022; Kelcey et al., 2019; Kunter et al., 2013b; Shechtman et al., 2010).

Because it is administratively very costly, only a few large-scale studies have so far investigated the impact of these constructs on student learning gains (see Table 1 in Krauss et al., 2020; for experimental evidence see, e.g., Hattie, 2024). In Appendix B9 (electronic supplementary information) we explain why this was particularly feasible in Germany in 2003/2004 in the framework of the COACTIV research program. In this study, for instance, PCK was the most influential factor for students' learning gains among more than 1000 teacher and lesson variables, after controlling for students' prior knowledge (e.g., Baumert et al., 2010; Kunter et al., 2013a,b). Furthermore, CK (Krauss et al., 2008a,b), constructivist beliefs (Voss et al., 2013), diagnostic skills (Binder et al., 2018), and enthusiasm (Kunter, 2013) also showed effects on student learning.

From the aforementioned large-scale studies, one can see how constructs can be identified that predict further desirable outcomes by implementing quantitative instruments (i.e., from left to right in the cascade model). The model of professional teacher competence (Fig. 1a, column 1), in turn, emphasizes the need to integrate factors such as motivational orientations, including enthusiasm for the subject or for teaching it. Furthermore, the German research tradition suggests considering the three basic dimensions of instructional quality (Fig. 1a, column 3; Praetorius et al., 2018). In the same way, it also points to the importance of including aspects of student mediation (e.g., experienced joy; Fig. 1a, column 4; Kunter et al., 2013b) that may ultimately contribute to student learning gains (Fig. 1a, column 5).

2.3 Expertise and competence paradigms: similar constructs but unclear role of experience

Despite the fundamental differences between the expertise and competence paradigms, the two approaches have identified very similar crucial teacher and lesson characteristics over the last decades. For instance, Anderson and Taner (2023) analyzed more than 100 empirical expertise studies and identified 73 specific features that were stressed in more than four of them. The authors informally grouped these features into six domains of an "expert teacher prototype": *knowledge base*, *cognitive processes*, *beliefs*, *personal attributes*, *professionalism*, and *pedagogic practice*. Features emphasized in more than ten studies included, among others, *PCK well developed*, *extensive CK*, *passion for profession*, and *creates positive, supportive learning environments*. Although the structure shown in Fig. 1 and the theoretical approaches described there are primarily used by quantitative researchers, most features identified by Anderson and Taner's meta-summary could easily fit into the cascade model (i.e., mainly into the first three columns).

With regard to (teaching) experience, practitioners in particular emphasize its central role in their profession. Indeed, within expert–novice paradigms, experience is frequently employed as a proxy for categorizing participants as experts (Stigler & Miller, 2018). However, although some growth in teacher expertise might occur during the first 5–7 years of practice (Berliner, 2004; Lopez, 1995), experience alone does not automatically lead to expert-level teaching. From the perspective of expertise research, *purposeful deliberate practice* with peer feedback and continuous reflection is needed as a potential trigger for individual development (Berliner, 2001; Ericsson et al., 1993). In fact, teachers might plateau after their initial years, and only those who actively reinvest their experience into innovative and responsive teaching potentially reach the expert level (Hattie, 2024; Palmer et al., 2005).

Experience might be a necessary but by no means sufficient condition for the development of teacher expertise. Thus, the question is, how valid is the common use of experience as a proxy for expertise? There is only limited evidence about the differential dependency of various indicators of expertise on years of professional experience. We are also not aware of any attempt to model an OE measure that comprises many relevant indicators and then to investigate the correlation of such a measure with professional experience.

3 Research questions and methods

3.1 Research questions and hypotheses

With respect to choosing the indicators, the approach of the present paper is exploratory in nature. We considered the following 13 expertise indicators (shown in Fig. 1b within the cascade model): content knowledge (CK), pedagogical content knowledge (PCK), constructivist beliefs (CB), enthusiasm (ENT), job satisfaction (JS), methodological action competence (MAC), subject-specific action competence (SAC), cognitive activation (CA), constructive individual learning support (CS), classroom management (CM), experienced enjoyment by students (S-JOY), experienced anxiety by students (S-ANX, negatively poled), mathematical learning gains of students (S-ΔM). In Sect. 3.3 we describe how these indicators were chosen from the variables assessed in COACTIV.

By relying on data from the COACTIV study, in which all 13 indicators—together with teaching experience—were measured in a single sample of German secondary mathematics teachers and their PISA classes, we aimed to address the following research questions (RQs):

- RQ1: How do the expertise indicators correlate with years of teaching experience?
- RQ2: How does an OE index formed by the indicators depend on experience when
  - a) the linear correlation is considered?
  - b) a nonlinear model is allowed?

Whereas the choosing of variables from the COACTIV data set has an explorative character, the analyses are hypothesis testing by nature. That the application of experience as a preselection criterion for "expert teachers" has been so prevalent—according to Stigler and Miller (2018), it is even the main criterion in expert–novice paradigms—carries the implicit hypothesis that years of teaching can be taken as a proxy of expertise. We tested this assumption with respect to 13 pertinent expertise indicators (RQ1) as well as regarding a newly formed overall expertise indicator (i.e., the OE index; RQ2).

By allowing the relationship between this overall expertise indicator and years of experience to be fitted using locally estimated scatterplot smoothing (LOESS, for details see Sect. 3.5), in addition (RQ2b), we aimed to identify specific phases of the teaching career and to eventually distinguish between beginning of career, midcareer, and end of career.

Note that the implicit hypothesis on experience being able to distinguish between experts and nonexperts should not be confused with that on intraindividual expertise development. Inferring intraindividual development from cross-sectional data would entail the risk of an ecological fallacy (Robinson, 1950). We instead sought to answer the question of whether it is a valid option to rely on experience when looking for expert teachers (i.e., at a specific point in time).

3.2 Context of the study: COACTIV 2003/2004

The COACTIV study,[1] conducted in Germany from April 2003 (end of 9th grade) to April 2004 (end of 10th grade), empirically examined the professional competence of secondary mathematics teachers in German PISA classes. It aimed to investigate which teacher competencies influence instructional quality and student learning gains (for more background see Appendix B9). Regarding the teachers, many valid and reliable measurement instruments were newly developed or adapted. These instruments tapped various facets of the four overarching competence aspects shown in column 1 of the cascade model (Fig. 1a). They also measured professional perception, interpretation, and decision making by implementing short video vignettes (column 2) as well as aspects of instructional quality (column 3). In PISA, student variables such as experienced joy or anxiety regarding mathematics (column 4) and achievement at the end of grades 9 and 10 were measured (column 5). In sum, more than 1000 variables were recorded for each teacher. For an overview of the COACTIV instruments, constructs, and scales, see Baumert et al. (2009).

Note that despite the age of the study, the present analyses are still relevant: While mean values of scales can and should change over time (e.g., 20 years later, means in beliefs or knowledge of teacher samples may be affected by changes in society or in teacher training), this does not necessarily apply to correlations between constructs (e.g., between expertise indicators and years of experience, or regarding predictive validity of constructs). This assumption is plausible, as such models represent structural empirical relations (i.e., in teaching environments), and there are no prevailing theories on the temporal change of these relations. Moreover, especially linear models are unaffected by changes in the means of the variables involved.

3.3 Constructs and instruments

Student variables do not appear in models of professional teacher competence (Kunter et al., 2013a) or in the list of Anderson and Taner (2023) on teacher expertise research. We took students' learning gains as the key target of teaching mathematics and modeled experienced joy as well as the absence of anxiety as supportive factors for students. The main criteria for choosing the indicators on the teacher side were (a) their availability in the COACTIV data set, (b) the demonstration of their predictive validity for the student variables in previous COACTIV analyses, and (c) their pertinence to both expertise and competence research.

To continue going from right to left in Fig. 1b, the three basic dimensions of instructional quality (i.e., CA, CS, and CM) have been successfully implemented in nearly all SEMs or PMs published to date that examined the predictive validity in the COACTIV study. Last but not least, we identified eight constructs regarding teacher competence that show predictive validity with respect to instructional quality or teaching goals. In sum, we obtained a set of 14 psychometric constructs.

Exploratory pre-analyses led us to exclude diagnostic skills from this set (Brunner et al., 2013). In COACTIV, diagnostic skills were measured by eight single items. However, it was not possible to build a reliable scale, so in COACTIV the term diagnostic *skills* was chosen deliberately (instead of diagnostic *competence*). The individual skills do not correlate systematically with each other, and it depends on the specific statistical method whether a certain diagnostic skill can predict instructional quality or student variables (see Binder et al., 2018). All remaining constructs are pertinent to expertise as well as to competence research (Anderson & Taner, 2023). Because of our focus on individual teachers and on high values of features (i.e., "expert teachers"; e.g., see Fig. D1 in Appendix D in the online supplementary information), we decided to deviate from the previous COACTIV convention and label all 13 constructs as *expertise* indicators.

[1]The COACTIV study was funded by the German research foundation (DFG) from 2002 to 2006 (Directors: Jürgen Baumert, Max Planck Institute for Human Development Berlin; Werner Blum, University of Kassel; Michael Neubrand, University of Oldenburg).

Table 1 provides an overview of the 13 expertise indicators (also see Fig. 1b). In the online supplementary information (Appendices B1–B8) the conceptualization and measurement of each indicator is described, including full scales (if brief) or example items. In Appendix B8 we explain how our current analysis differed from previous COACTIV analyses by including the calculation of the individual learning gain per class. Further background information on the study can be found in Appendix B9.

3.4 Sample

The COACTIV sample consists of $N = 402$ mathematics teachers, of whom some have missing values for one or more of the 13 facets targeted here. As a result, the available sample size per indicator varies ($194 \leq N_i \leq 283$). Regarding RQ1, the maximum sample size was used for each indicator. For shared analyses across all indicators (RQ2), we allowed a maximum of four missing values per teacher (see 3.5 for details). Data on at least 9 of the 13 facets were available for $N = 217$ mathematics teachers, who taught the PISA class in grade 10 in 2004. In the analysis for RQ2, these teachers and their 10th-grade classes, consisting of $N = 4{,}421$ students in total, were included (the mean class size was about $n = 21$).

Of the $N = 217$ teachers, 94 were female, 106 were male, and 17 declined to state their gender. Age and professional experience were reported by $n = 201$ teachers. The average age was 47.33 years ($SD = 8.53$) and the mean professional experience as a practicing teacher was 21.15 years ($SD = 9.92$).

Note that in Germany, there are different secondary school tracks. Empirical studies usually differentiate between the academic track ("Gy": *Gymnasium*; up to grade 13), qualifying for university access, and the other, vocational tracks ("NGy": non-*Gymnasium*; up to grades 9/10). Of the $N = 217$ teachers, 87 worked in the Gy track and 130 in the NGy tracks.

3.5 Statistical analyses

For answering RQ1 and RQ2a, Pearson's product-moment correlation coefficients were computed. Since combining all 13 facets (e.g., as a sum score) regarding RQ2 would require excluding many teachers, owing to missing values on individual indicators, we allowed up to four missing indicators per teacher to form the OE index. This scoring approach enabled the inclusion of a relatively large teacher sample of $N = 217$ secondary mathematics teachers. For the situation-specific skills (MAC and SAC), approximately 30% of the cases exhibited missing values owing to teacher turnover between 2003 and 2004, as both constructs were assessed exclusively in 2003 and could therefore not be transferred to teachers who entered in 2004. With regard to the remaining 11 indicators, the final sample of $N = 217$ teachers displayed a maximum of 23 missing cases (11%).

All expertise indicators for these $N = 217$ teachers were transformed into standardized percentile ranks (0 = lowest expertise, 100 = highest expertise), allowing the indicators to be included in the index with equal weights. Finally, we took the median of all percentiles that were available per teacher (permitting up to four missing values) as the OE index. The median (instead of the mean) was applied to reduce the influence of outliers and provide a comparable ranking system despite the varying sample sizes.

Regarding RQ2b, we used LOESS (Cleveland, 1979). This is a nonparametric regression method that fits a smooth curve to data by averaging subsets of nearby points (the scale defining the points included as neighboring being one parameter). LOESS performs a weighted linear regression for each point using only its neighbors, with weights declining as distance increases. The technique adapts to local patterns in a nonlinear manner and does not assume a fixed functional form (e.g., quadratic, asymptotic, etc.), which one would need to specify if applying a model of the generalized linear family. Thus, we did not explicitly assume or model a specific curve; rather, we exploratively allowed for nonlinearity in general.

4 Results

Table 2 presents the descriptive results for the 13 expertise indicators in the COACTIV study based on the maximum available N per indicator, with means, standard deviations, and school-type differences reported in their original metrics. In addition, the sample size, number of items, and reliability coefficient (Cronbach's alpha) are provided for each indicator.

4.1 RQ1: how do the expertise indicators correlate with years of teaching experience?

The rightmost column in Table 2 displays the relationships between professional experience (in years) and individual expertise indicators. Except for classroom management and the mathematical learning gains of the students, all indicators descriptively show a negative correlation with experience. Note that this is especially pronounced regarding the indicators in columns 1 and 2 of the cascade model (i.e., professional competence), where four of seven correlations are significantly negative.

The three highest negative correlations were found between expertise and content knowledge (CK), methodological action competence (MAC), and constructive individual learning support (CS). The linear correlation coefficient (Table 2) was $r = -.19$ ($p = .01$) for CK, $r = -.23$ ($p < .01$)

Table 1 The 13 expertise indicators (see Fig. 1b)

Expertise indicator (abbreviation) supplementary information	Column in the cascade model	Subscales / facets involved (# items / # codes per facet)	# items / codes (open[a]/ closed)	Year student / teacher test / questionnaire
Pedagogical content knowledge (PCK) B1	*1*: Professional teacher competence	Three facets: *knowledge of explaining and representing content* (11) *expertise on typical student errors and difficulties* (7) *familiarity with multiple solutions to tasks* (4)	22 items (open)	2004 teacher test
Content knowledge (CK) B1	*1*: Professional teacher competence	No facets	13 items (open)	2004 teacher test
Constructivist beliefs (CB) B2	*1*: Professional teacher competence	Three facets: *mathematics as process* (4) *independent and discursive learning* (12) *confidence in students' autonomous learning* (5)	21 items (closed)	2004 teacher questionnaire
Enthusiasm (Ent) B3	*1:* Professional teacher competence	Two facets: *enthusiasm for the subject* (3) *enthusiasm for teaching* (2)	5 items (closed)	2004 teacher questionnaire
Job satisfaction (JS) B4	*1*: Professional teacher competence	No facets	6 items (closed)	2004 teacher questionnaire
Methodological action competence (MAC) B5	2: Situation-specific skills	Two codes (for teachers' answers regarding three videos): *Dimension 1: student-centered, cognitively demanding activities* *Dimension 2: use of appropriate, non-subject-specific methods*	2 codes (open)	2003 teacher test (video)
Subject-specific action competence (SAC) B5	*2*: Situation-specific skills	Three codes (for teachers' answers regarding three videos): *Dimension 3: emphasis on conceptual understanding* *Dimension 4: correctness and precision of the mathematical content* *Dimension 5: recognizing a unique didactic opportunity*	3 codes (open)	2003 teacher test (video)

Table 1 (*Continued*)

Expertise indicator (abbreviation) supplementary information	Column in the cascade model	Subscales / facets involved (# items / # codes per facet)	# items / codes (open[a]/ closed)	Year student / teacher test / questionnaire
Cognitive activation (CA) B6	*3*: Instructional quality	Five facets: *cognitively activating tasks* (8) *supporting cognitive autonomy* (4) *cognitively challenging approach to students' ideas and errors* (5) *independence and obligation to give reasons when working on tasks* (8) *discursive treatment of different student solutions* (5)	30 items (closed)	2004 teacher/student[b] questionnaire
Constructive individual learning support (CS) B6	*3*: Instructional quality	Five facets: *constructive responses to student errors* (3) *respectful treatment of students* (3) *adaptive support and explanation for difficult tasks* (4) *patience in dealing with comprehension difficulties* (3) *caring ethos* (3)	16 items (closed)	2004 student questionnaire
Classroom management (CM) B6	*3*: Instructional quality	Three facets: *disruptions in math lessons* (3) *time wasting in math lessons* (3) *disruptions and time wasting* (8)	14 items (closed)	2004 teacher questionnaire
Experienced enjoyment (S-JOY) B7	*4*: Student mediation	Two facets: *joy* (6) *interest* (5)	11 items (closed)	2004 student questionnaire
Experienced anxiety (S-ANX) B7	*4*: Student mediation	Two facets: *anxiety* (9) *anxiousness* (5)	14 items (closed)	2004 student questionnaire
Mathematical learning gains (S-ΔM) B8	*5*: Teaching goals	difference of z-values of two PISA-tests: *PISA 2003 mathematics* *PISA 2004 mathematics (curricular valid)*	PISA-scales	2003/2004 student test

[a]All open answers were coded by independent raters according to a coding scheme.

[b]Only discursive treatment assessed by students

Table 2 Descriptive results on the 13 expertise indicators and overall expertise (OE) index

Indicator	Number of items (internal consistency per indicator: Cronbach's α)	Theoretical range	School track M (SD)			School track differences	OE index[a] ($\alpha_{OE} = .70$)	
			Total	Gy	NGy	Effect size (significance) d (p) (Gy–NGy)	Part–whole corrected item–total correlation r_{it} ($N = 194–283$)	Bivariate correlation r (p) with experience (in years) ($N = 173–260$)
1. CK: Content knowledge[t4] ($N = 212$)	13 (.83)	0–13	5.92 (3.36)	8.49 (2.24)	4.07 (2.76)	1.74 (<.01)	.27	−.19 (.01)
2. PCK: Pedagogical content knowledge[t4] ($N = 202$)	22 (.78)	0–37	19.77 (6.24)	22.55 (5.86)	17.75 (5.73)	0.83 (<.01)	.34	−.13 (.07)
3. CB: Constructivist beliefs[t4] ($N = 223$)	21 (.77)	1–4	3.15 (0.40)	3.20 (0.45)	3.12 (0.36)	0.20 (.16)	.39	−.11 (.11)
4. ENT: Enthusiasm[t4] ($N = 221$)	5 (.80)	1–4	3.18 (0.46)	3.20 (0.49)	3.16 (0.45)	0.09 (.54)	.43	−.15 (.04)
5. JS: Job satisfaction[t4] ($N = 224$)	6 (.88)	1–4	2.85 (0.73)	2.83 (0.71)	2.86 (0.74)	−0.04 (.78)	.29	−.07 (.29)
6. SAC: Subject-specific action competence[t3] ($N = 283$)	3 (.84)	0–6	2.06 (1.39)	2.86 (1.41)	1.66 (1.20)	0.94 (<.01)	.30	−.17 (.01)
7. MAC: Methodological action competence[t3] ($N = 283$)	2 (.80)	0–6	3.21 (1.61)	3.66 (1.48)	2.98 (1.62)	0.43 (<.01)	.33	−.23 (<.01)
8. CA: Cognitive activation[s4/t4] ($N = 229$)	30 (.65)	1–4	2.86 (0.28)	2.93 (0.30)	2.82 (0.25)	0.39 (<.01)	.47	−.02 (.79)
9. CS: Constructive individual learning support[s4] ($N = 213$)	16 (.97)	1–4	2.87 (0.44)	2.86 (0.46)	2.87 (0.43)	−0.03 (.86)	.45	−.27 (<.01)
10. CM: Classroom management[t4] ($N = 229$)	14 (.83)	1–4	2.74 (0.55)	2.75 (0.54)	2.74 (0.55)	0.03 (.83)	.32	.08 (.25)
11. S-JOY: Experienced enjoyment[s4] ($N = 214$)	11 (.95)	1–4	2.19 (0.25)	2.15 (0.20)	2.22 (0.27)	−0.27 (.04)	.34	−.05 (.47)
12. S-ANX: Experienced anxiety (inverted)[s4] ($N = 214$)	9 (.93)	1–4	2.03 (0.26)	1.96 (0.21)	2.07 (0.27)	−0.44 (<.01)	.24	−.04 (.54)
13. S-ΔM: Mathematical learning gains[s4] ($N = 194$)	[b]	–	0 (0.37)	−0.07 (0.35)	0.04 (0.37)	−0.30 (.04)	.17	.04 (.59)

Note: Cohen's effect sizes, d: $d = 0.2$: small; $d = 0.5$: medium; $d = 0.8$: large effect.

Gy = *Gymnasium* (academic track); NGy = non-*Gymnasium* (vocational tracks).

For s3, s4, t3, t4, s = students assessed, t = teachers assessed, 3 = 2003, 4 = 2004.

[a]To form the OE index, each indicator was transformed into standardized percentile ranks (0: lowest expertise; 100: highest expertise).

[b]Difference of z-values.

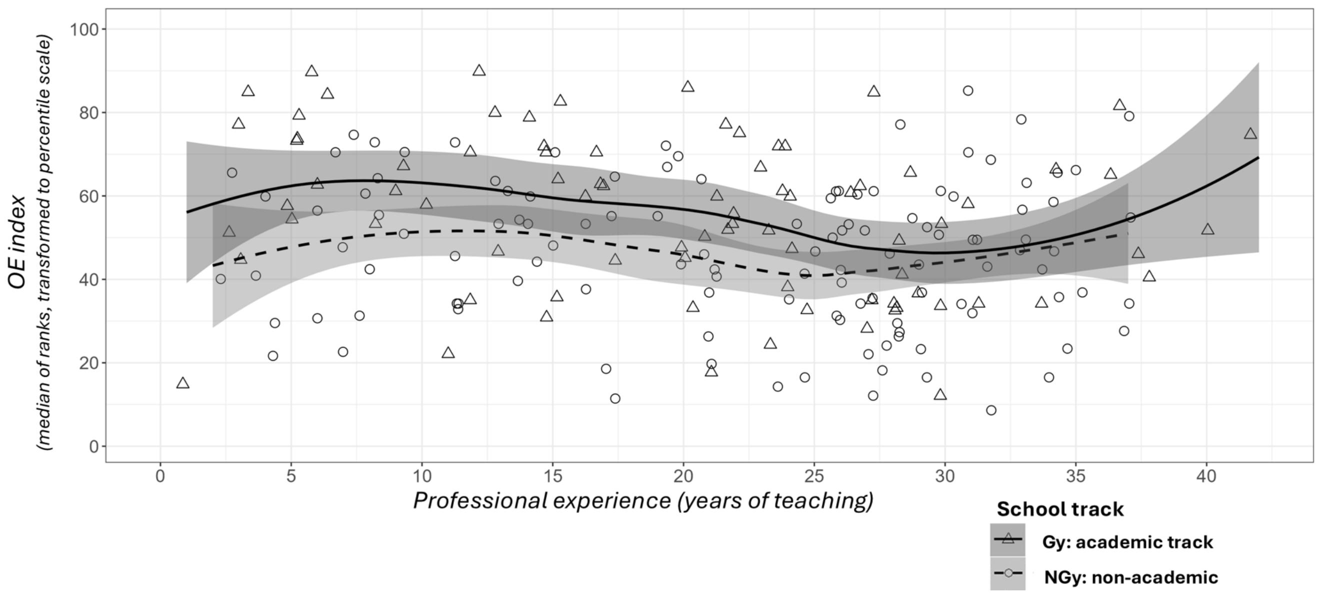

Fig. 2 Relationship of the overall expertise (OE) index and professional experience (cross-sectional data)

for MAC, and $r = -.27$ ($p < .01$) for CS. Therefore, at least regarding these indicators, the data seem to suggest that when looking for expert teachers, it is even preferable to look among younger teachers.

Note that the negative (or zero, respectively) correlations of the indicators with experience cannot be explained by some low values in "older" teachers alone (e.g., caused by exhaustion or similar effects). Instead, it is striking that all 13 scatterplots displaying the relation between each expertise indicator and experience revealed "unexperienced" teachers scoring substantially above average (also see Fig. 2).

In sum, regarding the 13 individual expertise indicators, there is no evidence of higher expertise in more experienced teachers and thus no support for the frequent application of experience as criterion for assigning teachers to the expert teacher group. However, from the negative correlations one *cannot* automatically conclude an intraindividual decline of, for example, CK, MAC, and CS during the professional career (also see Sect. 5). The question answered is instead: When one is searching in a cross-sectional sample of mathematics teachers for expert teachers (i.e., teachers with high expertise indicators), should the years of teaching guide this search?

4.2 RQ2: how does an OE index formed by the indicators depend on experience?

Before we tackle RQ2a and b, it is an intriguing question whether the COACTIV data allow the construction of a meaningful OE index. Even though we consider the index to be neither a reflexive nor a unified construct, the correlations between the individual indicators are informative (see the desideratum formulated by Anderson & Taner, 2023, above). On the basis of the percentile scales, we calculated the internal consistency of this newly constructed OE index using Cronbach's α, which was $\alpha_{\text{OE}} = .70$ based on $N = 217$ teachers (this is not an indicator of homogeneity; Stadler et al., 2021). Note that for interpreting Cronbach's α as reliability, unidimensionality must be assumed. Instead, in the way we use α here, it is an approximation of the mean split-half correlation (since tau equivalence cannot be assumed for the OE index).

In the second-to-last column of Table 2, the corresponding part–whole corrected item–total correlations of the individual indicators with the OE index are reported. Two of the three instructional quality dimensions—CA and CS—together with the enthusiasm of the teachers (ENT) proved to be the "most representative" expertise indicators. The item–total correlations were $r_{\text{it}} = .47$ for CA, $r_{\text{it}} = .45$ for CS, and, $r_{\text{it}} = .43$ for ENT. Note that two of these three indicators that are particularly representative for expert teachers—ENT and CS—simultaneously relate significantly negatively with experience (while CA is uncorrelated in this respect).

Although relations between some indicators can be expected, such an overall relation ($\alpha_{\text{OE}} = .70$) cannot be readily assumed for the following reasons: The 13 included expertise indicators (a) stem from structurally disjunct columns of the cascade model, (b) differ strongly in content, (c) are based on different measurement methods and tools (e.g., questionnaires, tests, coding of open answers), and (d) were evaluated from different perspectives (i.e., students vs. teachers vs. external coders). The individual expertise profiles of four exemplary teachers taken from the "top 10 expert teachers" referring to the OE index (Appendix D in the online supplementary information) also illustrate that high values on the indicators are not randomly distributed but tend to cluster (this finding is in line with the high Cronbach's α of the OE index). In any case, it is meaningful to reduce complexity and consider the OE index in the following.

4.2.1 Linear correlation between the OE index and experience (RQ2a)

The Pearson's product-moment correlation coefficient between professional experience (in years) and the OE index (comprising 13 relevant expertise indicators) was $-.18$ ($p = .01$), meaning that the two variables show a significantly negative correlation (Appendix C in the online supplementary information). Like the individual expertise indicators, the OE index also does not hint at higher expertise in more experienced teachers and thus suggests that professional experience should not be used as a criterion for identifying expert teachers.

However, the small but significant negative correlation coefficient requires a linear dependency between years of experience and expertise. It is questionable whether the assumption of linearity indeed best captures the underlying nature of the relationship, especially given the long and varied careers of teachers, typically spanning 40 eventful years.

4.2.2 Nature of the relationship between the OE index and experience (RQ2b)

Since Pearson's product-moment correlation captures only linear relations, it is interesting to allow for a nonlinear modeling of the relationship between the OE index and professional experience (in years). Figure 2 displays the corresponding scatterplot, fitted using LOESS (see Sect. 3.5) with two curves representing teachers from the two German secondary school types (Gy vs. NGy). First, the curves in Fig. 2 show a nonlinear relationship between the OE index and professional experience. Second, the two curves exhibit a very similar shape. Since the two groups (Gy vs. NGy) have no teachers in common (i.e., they are completely disjunct), this comparison may indicate a certain stability of the underlying relationship.

A closer look at Fig. 2 descriptively reveals a positive relationship between OE and experience in the first 0–8 years of a teaching career. Despite the impossibility of inferring longitudinal effects from cross-sectional data as a matter of principle (Robinson, 1950), it has to be noted that this relation is at least consistent with the oft-reported intraindividual increase of expertise in the beginning of a career (e.g., Lopez, 1995; Stigler & Miller, 2018). However, in the interval of about 9–27 years of experience, the diagram suggests a negative relationship, followed by an again positive relation in the phase of about 28–42 years. Since these three distinct phases can be roughly approximated as linear, Appendix C in the online supplementary information reports, among other results, the linear correlation coefficients restricted to these three intervals of interest. Together with the stable variability of expertise across the professional lifespan (see Fig. 2), we found no empirical evidence for the validity of taking experience as a prerequisite for expertise, neither as a necessary nor as a sufficient condition.

5 Discussion

5.1 Summary

We investigated the dependency of indicators of teacher expertise on professional experience (i.e., years as a practicing teacher). In the theoretical part, we consulted two prevalent paradigms in research on teaching: expertise and professional competence. Despite differing research methods and foci, similar competence aspects and features of an expert teacher prototype were identified within the two paradigms so far. Since these competence aspects on the one hand and consequences such as instructional quality and students' learning gains on the other belong to distinct columns of the cascade model (Fig. 1), we referred to them as expertise indicators.

In the empirical part, we tested the implicit assumption of many expert–novice paradigms that manifests as choosing experience as a criterion for assigning expert status to teachers by using the data of the COACTIV study 2003/2004. With this data, so far several SEMs have been implemented to investigate the predictive validity of teacher competencies or aspects of instructional quality for student variables such as joy or anxiety about mathematics and learning gains (e.g., Baumert et al., 2010; Kunter et al., 2013a,b). In previous publications on the COACTIV study, we identified eight competence aspects that had an impact on instructional quality or teaching goals. Thus, in addition to the three relevant student variables and the three basic dimensions of instructional quality, there were 14 potential constructs indicating teacher expertise. We excluded the variables on diagnostic skills (as they are very heterogeneous), leaving 13 individual expertise indicators (Fig. 1b, Table 1).

To investigate RQ1, we considered the bivariate correlations of these 13 indicators with the teachers' professional experience (i.e., years of teaching), revealing negative or zero correlations (Table 2). The sample for RQ2 consisted of all teachers with values on at least nine expertise indicators ($N = 217$). Using this sample, after ranking all teachers on each indicator and transforming these ranks into a percentile scale, we formed an OE index, determined as the median across all available valid measures per teacher.

By constructing and validating a comprehensive OE index using 13 theoretically grounded indicators from multiple stages of the cascade model, we offer a new integrative approach for operationalizing teacher expertise that allows for more general analyses within the professional lifespan.

Regarding RQ2a, relating this index again to professional experience reproduced the negative linear effects of the individual indicators. Concerning RQ2b (allowing the scatterplot to be fitted using LOESS) a nonlinear relationship between the OE index and experience emerged. We descriptively found evidence for a positive relation in the first 8 years of teaching (even in line with the longitudinal state of the art; e.g., Lopez, 1995; Stigler & Miller, 2018). In the midphase of a teaching career, however, a negative linear relationship was found in the data. In the last 10–15 years of the career (a phase that seems to be a blind spot in teacher expertise research), the relationship becomes positive again. Remarkably, similar curves also reappear in the two subsamples (GY vs. NGy track teachers).

Note that one cannot derive a conclusion about intraindividual development from the lack of a cross-sectional correlation between experience and expertise. Nor can one use this to make a judgment about the quality of an educational system (maybe—or maybe not—the teachers' education was successful and the prevailing expertise is "sufficient" in some way; see Sect. 5.2).

5.2 Limitations

A possible limitation of the present analyses is the age of the data. However, we argue that there are no substantial theories on the temporal change of correlational structures in this domain. Thus, even 20-year-old data remain relevant in this respect and may contribute to our understanding of the relationships between the analyzed variables today (also see Appendix B9).

Answering our research questions did not require us to infer an existing or not existing intraindividual increase in expertise from the cross-sectional COACTIV data. However, this is an issue worth discussing, especially because such an interpretation is tempting. Moreover, if we integrated cross-sectional data, for example, on the PCK of a sample of teacher students (Krauss et al., 2017), the combined scatterplot would show a strong "increase" during the university phase, a slightly weaker "increase" in the subsequent practical phase (in Germany the second-phase of teacher education, which is a seminar-based practical education, so-called Vorbereitungsdienst), and afterward a course like the curve of Fig. 2 (slight increase in the beginning years of a teacher, then approximating some kind of maximum). Thus, the "whole picture" then indeed would look exactly like typical logarithmical individual learning curves, be it regarding expertise research or concerning general learning theories.

However, the ecological fallacy (Robinson, 1950) teaches us that longitudinal and cross-sectional correlations can have the same or inverse signs in principle. For instance, potential cross-sectional correlations of expertise (indicators) with teachers' professional experience (in years) may arise from (multiple) individual developments in expertise over time but also for several other reasons:

First, it could be possible to identify generation effects. For instance, if the highest educational degree of the teachers' mothers is correlated with teachers' experience, this relation should be strongly negative (as it is in the COACTIV data). This, however, does not document a development within a teacher's career (the mother always has the same educational degree) but rather reflects that, due to substantial changes in societal roles over the last five decades, more and more women have pursued higher education and careers of their own.

Second, age effects are perceivable. For example, in the COACTIV reaction time test (teachers had to respond as quickly as possible to evaluate a student's answer; Binder et al., 2018; Krauss & Brunner, 2011), older teachers may have reacted more slowly simply because of an age-related reduction in general reaction time. Although such effects could theoretically be considered a decrease in expertise, they are not a direct effect of increasing experience.

Third, sometimes, all three factors may contribute to a correlation with experience. For instance, if older teachers show less patience with students, this may be due to a generation effect ("in the past, teachers were stricter"), an age effect ("increasing distance from the world of the students"), or a direct effect of (even negative) development in expertise ("learned impatience," e.g., by exhaustion).

Thus, a cross-sectional correlation would be only a possible indicator of intraindividual change processes. Note that in the same way, a lack of correlation cannot rule out longitudinal development. For example, when—just as a thought experiment—explicitly modeling an intraindividual increase at a second measurement point for every single teacher, a zero-correlation scatterplot from the first measurement point could simply move upward to the right overall in the second measurement point, preserving the cross-sectional zero-correlation despite multiple individual expertise increases.

Still, a longitudinal interpretation of our results in any case would stimulate discussion as to why the instruments typically used in quantitative studies are not sensitive to expertise development. Certainly, there are multiple intraindividual changes within a teacher's career, for instance, the development of higher automaticity or the formation of routines (yet both automaticity and routines have been said to prevent expertise development in general expertise theory; Gruber et al., 2010). If one believes in intraindividual expertise development, one must then determine how to design quantitative instruments to reveal this development. If one does not interpret the results of this article in a longitudinal way, however, the interesting question remains how to design expertise measurement instruments so that more experienced teachers score higher than novices.

The approach we used in the present paper with respect to choosing the expertise indicators as well as forming the OE index (including the restriction to a maximum of four missing values), is exploratory. Still, given the frequent attribution of expertise based on professional experience (c.f., Stigler & Miller, 2018) and the availability of a large data set that can verify this assumption from multiple perspectives (Appendix B9), we consider this approach to be justified for obtaining evidence. We hope, however, that the present study may inform future longitudinal designs.

Moreover, in the analyses regarding RQ2b, no particular relationship between professional experience and the OE index was modeled. Future studies should test specific hypotheses regarding the expected curve of the relationship. Without assumptions about the relationship, the fitted curve (Fig. 2) also remains exploratory.

5.3 Conclusion

In a cross-sectional COACTIV sample of $N = 217$ secondary mathematics teacher we found no positive relation between experience (years of teaching) and expertise when the whole teaching career is considered. This is true for the 13 identified pertinent expertise indicators from various stages of the cascade model as well as for a newly formed OE index.

Taking additionally into account the stable variability of expertise across the professional lifespan (see Fig. 2), we found no empirical evidence for the validity of the common practice of using professional experience as a selection criterion for expert teachers in expert–novice paradigms. Note that there are teachers with very high expertise even in the early years of their teaching career (see the values in the upper left corner of the scatterplot, Fig. 2), which is also reproduced for every singly indicator. Thus, the present analyses raise not only the question of whether experience is a sufficient criterion for the selection of expert teachers but strikingly also the question of whether professional expertise is a necessary criterion.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11858-025-01748-2.

Funding Information Open Access funding enabled and organized by Projekt DEAL. The COACTIV study was supported by the German research foundation (DFG, grant number Ba-03/2001). The authors have no financial or proprietary interests in any material discussed in this article.

Declarations

Competing Interests The authors declare no competing interests.

References

Anderson, J., & Taner, G. (2023). Building the expert teacher prototype: A metasummary of teacher expertise studies in primary and secondary education. *Educational Research and Reviews*, *38*, 100485. https://doi.org/10.1016/j.edurev.2022.100485.

Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 25–48). Springer.

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M., & Yi-Miau, T. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente*. Max Planck Institute for Human Development.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., 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.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher*, *15*(7), 5–13. https://doi.org/10.3102/0013189X015007007.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6.

Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, *24*(3), 200–212. https://doi.org/10.1177/0270467604265535.

Binder, K., Krauss, S., Hilbert, S., Brunner, M., Anders, Y., & Kunter, M. (2018). Diagnostic skills of mathematics teachers in the COACTIV study. In T. Leuders, K. Philipp, & J. Leuders (Eds.), *Diagnostic competence of mathematics teachers – unpacking a complex construct in teacher education and teacher practice* (pp. 33–53). Springer. https://doi.org/10.1007/978-3-319-66327-2_2.

Blömeke, S. (2025). Intelligence, knowledge, skills, behavior: Refining the Blömeke, Gustafsson, and Shavelson model of competence-as-a-continuum. In A. Gegenfurtner & R. Stahnke (Eds.), *Teacher professional vision: Theoretical and methodological advances*. Routledge. https://doi.org/10.4324/9781003370901-7.

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194.

Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, *79*, 101600. https://doi.org/10.1016/j.learninstruc.2022.101600.

Blömeke, S., Kaiser, G., König, J., & Jentsch, A. (2020). Profiles of mathematics teachers' competence and their relation to instructional quality. *ZDM – Mathematics Education*, *52*, 329–342. https://doi.org/10.1007/s11858-020-01128-y.

Boshuizen, H. P., Gruber, H., & Strasser, J. (2020). Knowledge restructuring through case processing: The key to generalise expertise development theory across domains? *Educational Research and Reviews*, *29*, 100310. https://doi.org/10.1016/j.edurev.2020.100310.

Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A., & Helsen, W. (2019). The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological Bulletin*, *145*(10), 980–1027. https://doi.org/10.1037/bul0000207.

Bromme, R. (2008). Lehrerexpertise. In W. Schneider & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* (pp. 159–167). Hogrefe.

Bruckmaier, G., Krauss, S., Blum, W., & Leiss, D. (2016). Measuring mathematical teachers' professional competence by using video clips (COACTIV video). *ZDM – Mathematics Education*, *48*(1–2), 111–124. https://doi.org/10.1007/s11858-016-0772-1.

Brunner, B., Anders, Y., Hachfeld, A., & Krauss, S. (2013). The diagnostic skills of mathematics teachers. 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. https://doi.org/10.1007/978-1-4614-5149-5_11.

Caspari-Sadeghi, S., & König, J. (2018). On the adequacy of expert teachers: From practical convenience to psychological reality. *International Journal of Higher Education*, *7*(5), 1–19. https://doi.org/10.5430/ijhe.v7n5p1.

Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM – Mathematics Education*, *50*(3), 355–366. https://doi.org/10.1007/s11858-018-0914-8.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836.

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society*, *24*(3), 177–181. https://doi.org/10.1177/0270467604264992.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363.

Gruber, H., Jansen, P., Marienhagen, J., & Altenmueller, E. (2010). Adaptions during the acquisition of expertise. *Talent Development & Excellence*, *1*, 3–15.

Hattie, J. (2024). *Visible learning 2.0*. Schneider Verlag Hohengeren.

Helmke, A. (2022). In *Unterrichtsqualität und Professionalisierung. Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*, Klett.

Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers – cognitive versus situated approaches. *Educational Studies in Mathematics*, *94*(2), 161–182. https://doi.org/10.1007/s10649-016-9724-5.

Kaiser, G., & König, J. (2020). Analyses and validation of central assessment instruments of the research program TEDS-M. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student learning in German higher education* (pp. 29–51). Springer. https://doi.org/10.1007/978-3-658-27886-1_3.

Kelcey, B., Hill, H. C., & Chin, M. J. (2019). Teacher mathematical knowledge, instructional quality, and student outcomes: A multilevel quantile mediation analysis. *School Effectiveness and School Improvement*, *30*(4), 398–431. https://doi.org/10.1080/09243453.2019.1570944.

König, J., Santagata, R., Scheiner, T., Adleff, A. K., Yang, X., & Kaiser, G. (2022). Teacher noticing: A systematic literature review of conceptualizations, research designs, and findings on learning to notice. *Educational Research and Reviews*, *36*, 100453. https://doi.org/10.1016/j.edurev.2022.100453.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM – Mathematics Education*, *40*(5), 873–892. https://doi.org/10.1007/s11858-008-0141-9.

Krauss, S., & Bruckmaier, G. (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (2nd ed., pp. 241–261). Waxmann.

Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N., & Blum, W. (2020). Competence as a continuum in the COACTIV study: The "cascade model". *ZDM – Mathematics Education*, *52*, 311–327. https://doi.org/10.1007/s11858-020-01151-z.

Krauss, S., & Brunner, M. (2011). Quick assessment of the accuracy of students' answers: A reaction time test for mathematics teachers. *Journal für Mathematik-Didaktik*, *32*, 233–251. https://doi.org/10.1007/s13138-011-0029-z.

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

Krauss, S., Lindl, A., Schilcher, S., Fricke, M., Göhring, A., Hofmann, B., Kirchhoff, P., & Mulder, R. H. (Eds.) (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. Waxmann.

Kunter, M. (2013). Motivation as an aspect of professional competence: Research findings on teacher enthusiasm. 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. 273–289). Springer. https://doi.org/10.1007/978-1-4614-5149-5_13.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.) (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers*. Springer. https://doi.org/10.1007/978-1-4614-5149-5.

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

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. https://doi.org/10.1007/978-1-4614-5149-5_6.

Lopez, O. S. (1995). *The effect of the relationship between classroom student diversity and teacher capacity on student performance. Executive summary. Conclusions and recommendations for educational policy and practice* [Report]. The Strategic Management of the Classroom Learning Enterprise Research Series.

Palmer, D. J., Stough, L. M., Burdenski, T. K., & Gonzales, M. (2005). Identifying teacher expertise: An examination of researchers' decision making. *Educational Psychologist*, *40*(1), 13–25. https://doi.org/10.1207/s15326985ep4001_2.

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM – Mathematics Education*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357.

Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links from teacher knowledge, to classroom practice, to student learning in the instructional system of the middle-school mathematics classroom. *Cognition and Instruction*, *28*(3), 317–359. https://doi.org/10.1080/07370008.2010.487961.

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

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

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411.

Stadler, A., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology*, *60*, 100832. https://doi.org/10.1016/j.newideapsych.2020.100832.

Stigler, J. W., & Miller, K. F. (2018). Expertise and expert performance in teaching. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & M. A. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 431–452). Cambridge University Press. https://doi.org/10.1017/9781316480748.024.

Voss, T., Kleickmann, T., Kunter, M., & Hachfeld, A. (2013). Mathematics teacher' beliefs. 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. 235–258). Springer. https://doi.org/10.1007/978-1-4614-5149-5_12.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Saganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe & Huber.