



Growth of professional noticing of mathematics teachers: a comparative study of Chinese teachers noticing with different teaching experiences

Xinrong Yang¹ · Johannes König² · Gabriele Kaiser^{3,4}

Accepted: 27 December 2020 / Published online: 20 January 2021
© The Author(s) 2021

Abstract

The last decade has witnessed increasing interest in the study of teacher noticing in mathematics education research; however, little is known about the growth of teacher noticing and how it is influenced by teaching practice. Departing from the expert-novice-paradigm, in this paper we address this research gap by a cross-sectional study that investigates how Chinese mathematics teachers' noticing is affected by their developmental stage, measured by the length of their teaching experience. The study included 152 pre-service teachers at the end of their initial teacher education, 162 early career teachers with one to five years' teaching experience, and 123 experienced mathematics teachers with more than 15 years' teaching experience, who participated in a video-based assessment of their noticing competency conceptualized by the sub-facets of perception, interpretation, and decision-making. Our findings indicate a nearly linear growth in teacher noticing among Chinese mathematics teachers, with significant differences identified between pre-service and experienced teachers and only small differences between pre-service and early career teachers. Analyses using the method of Differential Item Functioning (DIF) further suggest that pre-service and early career teachers demonstrated strengths in aspects more related to reform-oriented or Westernized approaches to mathematics teaching, such as working with open-ended tasks, identifying characteristics of cooperative learning, and mathematical modeling tasks. By contrast, experienced teachers demonstrated strengths in perceiving students' thinking, evaluating teachers' behavior, and analyzing students' mathematical thinking. Our findings further highlight that the three sub-facets of teacher noticing develop differently within the three participating groups of teachers. These findings suggest that teaching experience acts as one influential factor in the development of teacher noticing in the Chinese context.

Keywords Teacher noticing · Video-based assessment · Pre-service teachers · Early career teachers · Experienced teachers

1 Introduction

The past decade has witnessed increasing interest in teachers' professional noticing, particularly in mathematics education research (e.g., Schack, Fisher, and Wilhelm 2017; Sherin, Jacobs, and Philipp 2011a, b). Teachers' noticing is generally accepted as a critical component of mathematics

teaching expertise and an important factor in the improvement of teaching effectiveness generally, and students' mathematical achievements in particular (Sherin et al. 2011a, b). Therefore, a clear understanding of the development process of teacher noticing seems essential not only for understanding the construct of teacher noticing but also for the effective understanding and promotion of the growth of teacher noticing.

Hitherto, little empirical evidence has been available to answer the question of “What trajectories of development related to noticing expertise exist for prospective and practicing teachers?” posed by Schoenfeld (2011, p. 234) nearly ten years ago. The study by Jacobs, Lamb, and Philipp (2010) is one of a few that used a cross-sectional design to compare the similarities and differences of teacher noticing among four groups of mathematics teachers with different

✉ Johannes König
johannes.koenig@uni-koeln.de

¹ Southwest University, Chongqing, China

² University of Cologne, Cologne, Germany

³ University of Hamburg, Hamburg, Germany

⁴ Australian Catholic University, Brisbane, Australia

teaching experience and—especially—professional development experience. Although they observed variability within each teacher group, consistent patterns and a monotonic growth were evident across the four groups. Schoenfeld's (2011) question suggests that the patterns of strengths and weaknesses of teacher noticing should be investigated in relation to the level of teaching experience in classrooms.

Therefore, the growth of teachers' noticing skills seems to be "worthy of attention" (Jacobs et al. 2010, p. 193) and studies investigating patterns are overdue. However, most studies on the development of teacher noticing are intervention studies aimed at identifying effective ways to enhance teacher noticing. Mainly video-based, the pre- and post-test results generally suggested clear improvement of teacher noticing (Santagata et al., 2021, this issue). The intervention compounds various effects that must be considered, such as increased teaching experience, differences in the observed videos, and a learning effect from seeing the same videos several times (Simpson, Vondrová, and Žalská, 2018). Moreover, such studies offer only limited insight into the possible development trajectories of teacher noticing since many interventions took place during the pre-service training period and do not reflect the impact of teaching practice on the growth of teacher noticing without specific interventions.

However, teachers' professional growth in general, and the development of teacher expertise in particular, is a "complex and continuing process" (Wilkie 2019, p. 96). Therefore, although teaching experience alone is insufficient for the development of teacher noticing (Jacobs et al. 2010; Simpson et al. 2018), teaching experience is undoubtedly necessary for the growth of teachers' noticing. Teachers, especially in-service teachers, will continue learning to teach mainly within a specific school community through informal learning opportunities, such as mentoring, observing other teachers' teaching, and peer group discussion (Kyndt, Gijbels, Grosemans, and Donche, 2016; Patrick 2010). Therefore, studies within a more authentic context are necessary to obtain a more comprehensive picture of the development of teacher noticing.

Teacher noticing has been described as "socially and culturally constructed" (Louie 2018, p. 61), referring to approaches that frame learning to teach mathematics as a cultural activity. Therefore, teacher noticing development should be more thoroughly understood as a process that takes place within a particular sociocultural context. Hitherto, however, most studies of mathematics teacher noticing in general, and of teacher noticing growth in particular, have been conducted in Western contexts. Little empirical evidence is available from non-Western countries, for example, from China as an influential East Asian country with a specific mathematics teacher education culture.

In China, pre-service teachers are trained at specific teacher training institutions, called normal universities, with

a strong focus on subject matter knowledge. Normal universities generally provide four-year bachelor programs for pre-service teachers. Around 60% of pre-service teacher curriculum hours are devoted to mathematics subject courses, such as advanced algebra, analytical geometry, functional analysis, abstract algebra, and topology (Paine, Fang, and Wilson 2003; Li, Huang and Shin, 2008). Owing to the lack of pedagogical content knowledge and teaching skills, newly graduated teachers in China are not regarded as qualified but as "semi-finished products" (Paine et al. 2003, p. 216), who will need to learn and continue to develop their teaching skills when they enter teaching positions in schools. The most common way for in-service teachers to develop their teaching skills in China is through school-based professional activities, such as compulsory mentoring programs, open lessons, and exemplary lessons (see Li and Huang, 2018).

Departing from these manifold research gaps, in this study we seek to answer the following research questions:

- (1) How do teachers' noticing skills develop globally in relation to their teaching experience? We addressed this question by comparing three cohorts of Chinese teachers with different degrees of teaching experience (pre-service teachers, early career teachers, and experienced teachers). For this purpose, we examined whether a growth in teacher noticing in relation to the degree of teaching experience can be identified and how it can be characterized. At a more fine-grained level, in the study we aimed to answer the second research question:
- (2) Is it possible to identify strengths and weaknesses of these different cohorts of secondary school mathematics teachers concerning different sub-facets of noticing and different aspects of noticing, and if yes, which of these are present?

As the study included an East Asian context, we examined whether the same trend of growth can be identified as was reported in the few already existing studies carried out in Western contexts.

2 Literature review

2.1 Mathematics teachers' noticing and our own theoretical approach

Owing to the complexity of teaching and the different areas of focus in empirical studies, teacher professional noticing is defined in various ways, and different aspects of teaching practice are considered (Sherin et al. 2011a, b). For example, Sherin and others defined teacher professional noticing as comprised of the following three components: (a) identifying what is important in an instructional

setting; (b) making connections between specific events and broader principles of teaching and learning; and (c) using knowledge about the context to reason about a situation (Sherin 2007; Sherin and van Es 2005, 2009). Jacobs et al. (2010) subsequently refined the definition and proposed that teacher professional noticing includes (a) attending to students' strategies, (b) interpreting students' understanding, and (c) deciding how to respond based on students' understanding.

To date, these two definitions of teacher noticing are the most widely cited conceptualizations in the field of teacher noticing. Sherin et al. (2011a, b, p. 5) noted the consensus within research that this construct can be characterized as consisting of at least two components, namely, attending to important classroom incidents, and making sense of events in an instructional setting including interpreting and reasoning. Depending on the conceptualization of the last component, it may include instructional responses by the teachers. These two components were considered as the two phases of teachers' noticing, and were described as consequential, interrelated, and cyclical.

The current study TEDS-East–West is embedded in the TEDS-M (Teacher Education and Development Study in Mathematics) research program, specifically the TEDS-Follow-up study. Overall, TEDS-East–West aims to compare the influence of mathematics teachers' professional competence on instructional quality and students' mathematics achievement between China and Germany. The TEDS-FU study extended the cognitive approach to teacher competence (namely teacher knowledge and beliefs) to include situation-specific competence facets referring to the approach of 'noticing' (Kaiser et al. 2015; 2017).

In the extended theoretical framework developed within the TEDS-M research program, teacher noticing was defined as consisting of the following three facets: (a) perceiving particular events in an instructional setting; (b) interpreting the perceived activities in the instructional setting; and (c) decision-making, either anticipating responses to students' activities or proposing alternative instructional strategies (Kaiser et al. 2015, 2017). These three sub-facets of noticing—perception, interpretation, and decision-making—were called the 'PID model'. This approach refers to the expert-novice paradigm, in which the construct perception was widely used to describe the first phase of teachers' actions in an instructional setting, restricting the construct to observable, discernable incidents (Berliner 2001; Carter et al. 1988). The construct of attending was deliberately not used, although it is the more usual terminology in the noticing discourse, as this construct is already strongly connected to interpreting what is important in the classroom (Sherin et al. 2011a, b). In order to allow an empirical separation of the different facets of teachers noticing, our conceptualization refers to the construct of perception.

In contrast to other frameworks, this definition not only requires teachers to perceive and interpret particular events but also to make decisions and develop reasonable proposals for the continuation of classroom activities. This definition of teachers' noticing is not restricted to noticing of students' mathematical thinking, as in most of the earlier frameworks. This understanding of noticing comprises a broad understanding of the whole classroom situation and the aspects important for the quality of mathematics teaching, such as the design of mathematical teaching and learning processes, the potential for students' cognitive activation, individual learning support, and classroom management (Yang et al. in press). Overall, this model differentiates teacher professional noticing in two sub-domains: noticing based on general pedagogy (P_PID) and noticing based on mathematics pedagogy (M_PID). As the results reported in the present study were developed within this research program in an East Asian context, the theoretical framework used in studies concerning this conceptualization of teacher noticing was employed in this study.

2.2 Differences in teacher noticing in relation to teaching experience

In recent years, many researchers have investigated the development of teacher noticing. The most popular approach is to adopt the expert-novice paradigm by comparing teachers at different developmental stages or levels of expertise (e.g. Jacobs et al. 2010). Beginning in the 1980s, within general research on expertise (Chi et al. 1981), studies compared the differences in knowledge and classroom teaching behaviors between novice, early career, proficient, and expert teachers (e.g., Berliner 2001). Although the main focus of these studies was not teacher noticing, they "can be regarded as precursors" (Lachner, Jarodzka, and Nuckles 2016, p. 198) to current studies on teacher noticing, since many aspects of teacher behaviors are essentially related to teacher noticing. In earlier studies on teacher expertise, novice teachers were found to attend mainly to surface-level events, such as student behavior and disciplinary issues, and were sometimes able only to attend to one event and ignored the others (Berliner 2001; Leinhardt, Putnam, Stein, and Baxter 1991; Tsui 2003). By contrast, expert teachers were better able to read critical cues from students and attend to classroom teaching events swiftly, holistically, and accurately.

Expert teachers exhibited greater ability to interpret the attended events in greater detail and with more insight. For example, novice teachers were unable to provide in-depth explanations or struggled to develop accurate interpretations of what they noted (Carter et al. 1988). By contrast, expert teachers could relate teaching principles and concepts to the noticed events and could therefore make knowledge-based interpretations (Berliner 2001; Tsui 2003). Furthermore,

expert teachers were better able to make ongoing adjustments to their teaching or make fast decisions (e.g., Livingston and Borko 1989). Indeed, teaching with flexibility has been widely proposed as a characteristic of expert teachers (e.g. Berliner 2001, 2004), which implies that expert teachers will make necessary and proper decisions during teaching.

More recently, empirical studies also compared the differences in teacher noticing between expert and novice teachers from the theoretical frameworks of teacher noticing as mentioned above. For example, Huang and Li (2012) found in their study that both expert and novice teachers attended to the development of students' mathematics knowledge and mathematical thinking ability, but expert teachers paid greater attention than novice teachers to developing higher-order mathematical thinking and mathematics knowledge, with less attention to teachers' direct guidance. Wolff, Jarodzka, van den Bogert, and Boshuizen (2016) found by comparing novice and expert teachers from diverse subjects, based on eye-tracking technology, that expert teachers' perception was more knowledge-driven and focused; furthermore, expert teachers were more likely to attend to critical cues and interpret them in relation to relevant classroom management issues. By contrast, novice teachers' perception was found to be more image-driven and scattered; they described more superficially salient cues.

A cross-sectional study by Jacobs et al. (2010) that compared four cohorts of mathematics teachers with different degrees of teaching experience identified a significant monotonic trend for all three facets of noticing skills (attending, interpreting, and decision-making). They found that pre-service mathematics teachers struggled with all three noticing facets; early career teachers showed evidence of attending to students' strategies and interpreting their understanding; almost all advanced teachers and teacher leaders demonstrated evidence of attending to students' strategies and some evidence of interpreting students' understanding; most emerging teacher leaders demonstrated superior expertise in deciding how to respond.

3 Methodology

The study reported in this paper was conducted between 2016 and 2019 within the frame of the study TEDS-East–West embedded in the TEDS-M research program.

3.1 Participants

The present study's sample consisted of the following three cohorts:

- 1) 152 pre-service teachers close to completing their four years' pre-service teacher education at undergraduate level;
- 2) 162 early career teachers with one to five years' teaching experience;
- 3) 123 experienced teachers with over 15 years' teaching experience at junior secondary school level (grades 7–9).

The 152 pre-service mathematics teachers were recruited from two normal universities in China. Fifty-one percent completed their teaching practicum in junior or lower secondary schools (Grades 7–9) and the others completed teaching practicum in senior or upper secondary school (Grades 10–12). Overall, all had four months of practical experience of school teaching lasting one semester, and all were trained to teach secondary school mathematics after graduation.

The 162 early career teachers were chosen from 18 provinces in China. Among them, 126 were part-time Master's degree students in mathematics education with at least one year of teaching experience in junior or senior secondary school. All were trained in the mathematics department at a normal university to teach junior (lower) and senior (upper) secondary school mathematics. The other 36 early career teachers with a first degree in mathematics education were chosen from the 18 junior secondary schools in which the highly experienced teachers were working. Among the 162 early career teachers, 59% were female and 17% taught in rural schools.

The 123 highly experienced junior secondary school mathematics teachers with teaching experience ranging from 15 to 36 years were chosen from different school types and school locations (rural and urban). In China, junior secondary school teachers teach only one subject, in this case mathematics. These teachers were chosen from 18 junior secondary schools in the same district in Chongqing, Western China's largest administrative area. Among them, 48% were female, and 35% taught in rural schools when the assessment was carried out.

3.2 Assessment instruments

The instruments used in the present study to test participants' noticing skills were adapted from the instruments designed in the TEDS-Follow-Up and TEDS-Instruct/Validate projects within the TEDS-M research program (for a detailed description of the adaptation and validation process, see Yang, Kaiser, König, and Blömeke 2018, 2019). The original three video-vignettes were developed within the TEDS-M research program to assess German mathematics teachers' professional noticing. As the instrument has already been described in other publications, we refrain from

describing item examples and refer to descriptions in various publications (e.g., Kaiser et al. 2015). The video assessment examined teachers' mathematics instruction-related noticing (M_PID) and general pedagogical noticing (P_PID) and distinguished the following three facets: perception, interpretation, and decision-making. Both the P_PID items and M_PID items required teachers to notice mathematics classroom teaching holistically—that is, items related to almost all aspects of classroom teaching.

Critical incidents in mathematics teaching and a range of typical teaching phases in a mathematics lesson were covered in the three video-vignettes, which explored the topics of functions, volumes and surfaces of geometrical solids. The three video-vignettes were based on scripted plots rather than episodes from real mathematics classroom teaching. Each video-vignette lasted around four minutes to provide participants with an overview of the whole lesson. Background information about the class and lessons prior to the lesson shown was also provided to help participants achieve a more comprehensive understanding of the teaching.

After watching each of the three videos, the mathematics teachers were asked to answer several items related to each of the videos within 15–20 min (around 60 min altogether). In total, there were 38 items (22 P_PID and 16 M_PID, see Fig. 1 for a sample item) based on Likert scales (four categories ranging, for example, from 'wholly correct' to 'incorrect') to assess teachers' perception. Thirty-six constructed-response items (18 P_PID and 18 M_PID) were

used to assess the teachers' interpretation and decision abilities. An expert rating was implemented in developing the test instrument to decide which answer could be regarded as correct with respect to the rating scales. A coding manual was developed and piloted before it was used in the German projects to improve its reliability and validity. Various different approaches, such as curricular analyses of the mathematical content and comprehensive expert workshops, were employed to ensure the instruments' content validity (Hoth et al. 2016).

To adapt the instruments for the Chinese context, the instruments were translated into Chinese and were checked by two mathematics education researchers and four junior secondary school mathematics teachers. Necessary modifications were made to several expressions used in the instrument. Several items closely related to the German mathematics curriculum and heterogeneity or multiple cultural backgrounds of students were deleted since they did not match the situation in China. Three Chinese junior secondary school mathematics teachers and their students retook the three video-vignettes and performed exactly as their German counterparts did. To further validate the instrument of teacher noticing in a Chinese context, both qualitative and quantitative methods, including content validity, 'elemental validity' (Hill, Dean, and Goffney 2007; Kane 2001), and construct validity, were employed to evaluate the psychometric properties of M_PID and P_PID, respectively (see Yang et al. 2018).

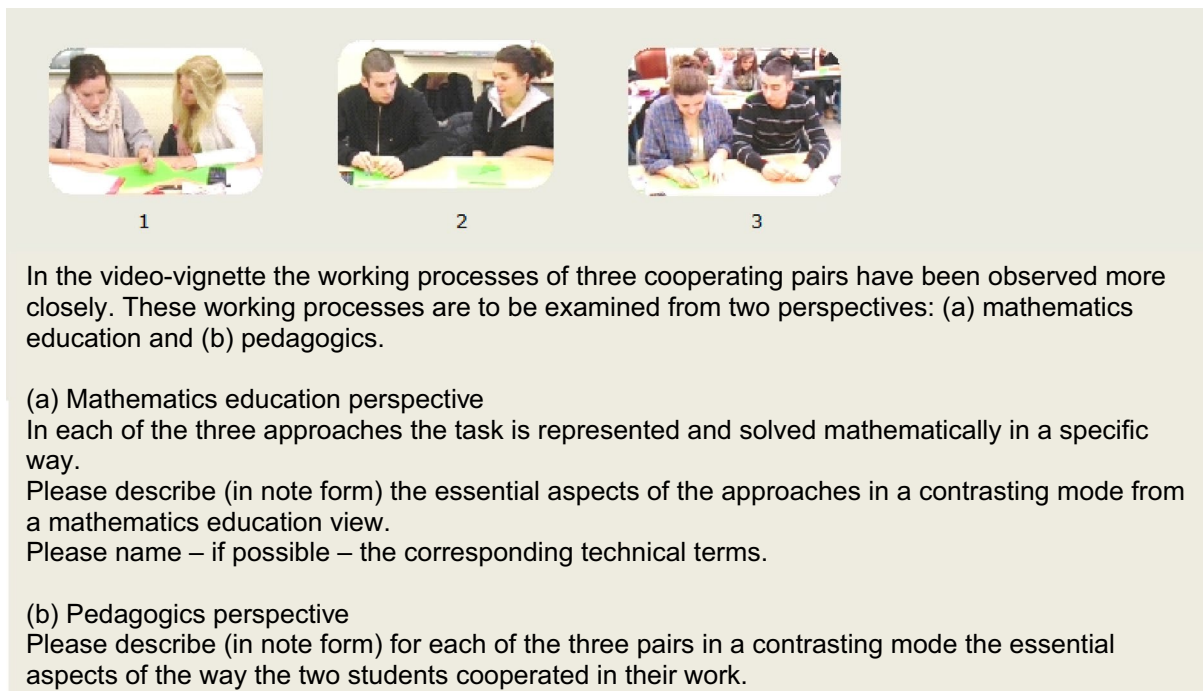


Fig. 1 Example of a high-inferential M_PID and P_PID item referring to interpreting a classroom situation (from Kaiser et al. 2015, pp. 381)

3.2.1 Scaling and data analysis

The data analysis comprised the following steps. First, the open response items were coded according to the coding manual's rubrics. Independent raters coded 56 of the questionnaires; good Cohen's Kappa values were reached ($k > 0.79$ and $K_{\text{average}} = 0.86$). For all open response items, items with no response or incorrect responses were scored 0, and each correct answer was scored 1 (for items with several sub-items, the sum of the correct answers was calculated). After completion of coding, the relative item difficulties for a one-parameter (Rasch model) item response theory (IRT) model were calculated separately on P_PID and M_PID. Items with extreme difficulty were removed from the final analysis because they showed weak discrimination and did not substantially contribute to the measurement of the construct (Bond and Fox 2007). The internal consistency of the remaining items in P_PID and M_PID were estimated using Cronbach's alpha reliability coefficient: these ranged from 0.712 to 0.789.

Subsequently, a multi-group graded response model was applied for both the P_PID and M_PID aspects of their sub-facets (perception, interpretation and decision-making). Owing to the low number of items in the decision-making sub-facet, decision-making and interpretation were merged, supported by the strong relationship between these sub-facets. The calibration of the a-parameters (item discriminations) and b-parameters (item difficulties) was conducted based on three teacher groups (concurrent calibration). Thus, item parameters were constrained to be equal in all groups, ensuring the same metric in the groups. The person parameters (maximum likelihood estimates) were then computed. The overall P_PID and M_PID and the four sub-facets person parameters of the three groups were transformed to a scale with an average score of 500 and standard deviation of 100. One-way ANOVAs were performed separately to examine the differences among the three groups of teachers, and post hoc (Scheffé) significance tests were further conducted to examine the differences between each of the two groups of teachers.

Furthermore, three pairwise differential item functioning (DIF) analyses were conducted with the main aim of identifying which items were typically in favor of which group of teachers. DIF is typically used to investigate item bias (Holland and Wainer 1993): an item is labeled as exhibiting DIF when different groups of test-takers have different probabilities of answering the item correctly after controlling for their abilities on the construct being measured. DIF analyses have also been used to identify potential strengths and weaknesses of specific groups as a complement to other methods in examining cultural influences in cross-cultural comparative studies (e.g., Blömeke, Suhl, and Döhrmann 2013; Mesic 2012; Yang et al. 2019) or to examine the influences

affecting learning in specific subjects (Gess, Wessels and Blömeke 2017).

In the present study, DIF analyses were performed to identify possible strengths and weaknesses of a specific group of teachers at the item level due to differences in their teaching experience defined by group samples (Sect. 3.1). DIF was detected using manifest logistic regressions (Swaminathan and Rogers 1990), which can detect uniform and non-uniform DIF (Hambleton et al. 1991). An item showing uniform DIF indicated that a specific group of teachers outperformed another group systematically throughout all the ability levels. If an item showed non-uniform DIF, the teaching experience factor significantly impacts only teachers with either higher or lower competence. The magnitude of DIF in the present study was further classified into three levels according to the thresholds proposed by Jodoin and Gierl (2001): $\Delta R^2 \leq 0.035$ is negligible, $0.035 < \Delta R^2 \leq 0.07$ is moderate, and $0.07 < \Delta R^2$ is large. After the items with DIF had been detected and classified, content analysis of the items with DIF was further conducted to identify explanations.

4 Results

4.1 Overall performance differences on P_PID and M_PID

The overall teacher noticing performance results from the aspects of P_PID and M_PID and the two sub-facets under each for the three teacher groups are presented in Table 1. The overall mean of the three groups was transformed to 500 test points with a standard deviation of 100 test points. As Table 1 indicates, an initial pattern can be identified: a monotonic increase in the numerical values of mean scores of both the overall P_PID and M_PID and their sub-facets, which can be interpreted as almost linear growth of teacher noticing amongst the three groups of teachers. Aside from the perception sub-facet of P_PID, one-way ANOVA results showed significant differences among the three teacher groups for all other aspects and sub-facets.

However, as Table 1 demonstrates, the post hoc analysis results showed significant differences only between pre-service and experienced teachers and early career and experienced teachers. No significant differences could be identified between pre-service and early career teachers, indicating no significant development between pre-service and early career teachers. Indeed, as Table 1 indicates, pre-service teachers' achievement in the (sub)facets related to P_PID was only about 0.15 standard deviations lower than that of early career teachers. The differences in the mean scores related to M_PID (sub) facets were even smaller (less than 0.1 standard deviations).

Table 1 Mean differences between groups of teachers with different degrees of teaching experience in terms of P_PID and M_PID

	Pre-service teachers (PT)	Early career teachers (ECT)	Experienced teachers (ET)	F	Scheffé Post Hoc Test
	Mean ± SD	Mean ± SD	Mean ± SD		
P_PID	470.13 ± 90.88	486.16 ± 106.01	555.14 ± 79.07	30.712***	ET > PT, ET > ECT
Perception	487.25 ± 96.38	499.65 ± 115.17	516.22 ± 79.17	2.879	
Interpretation and decision-making	451.64 ± 75.80	468.93 ± 75.50	600.68 ± 81.79	146.702***	ET > PT, ET > ECT
M_PID	470.23 ± 86.84	475.45 ± 90.86	569.12 ± 93.76	50.256***	ET > PT, ET > ECT
Perception	484.10 ± 100.65	485.12 ± 104.55	539.25 ± 81.32	13.972***	ET > PT, ET > ECT
Interpretation and decision-making	468.94 ± 82.89	476.84 ± 78.83	568.89 ± 111.06	50.060***	ET > PT, ET > ECT

***p < 0.001

Another pattern can be identified, namely, that the difference in the interpretation and decision-making sub-facets among the three groups of teachers is much greater than the difference in perception. As shown in Table 1, experienced teachers outperformed pre-service and early career teachers by 0.92 to 1.5 standard deviations in the sub-facets of interpretation and decision-making in P_PID and M_PID. However, in the sub-facet of perception, the differences were much smaller. This result suggests that it is relatively difficult for pre-service and early career teachers to develop noticing skills related to interpretation and decision-making.

4.2 Specific strengths and weaknesses in P_PID as indicated by DIF

To identify possible strengths and weaknesses in the P_PID sub-facets of the three groups of teachers, three pairs of DIF analyses were carried out separately between each of the two teacher groups. Table 2 summarizes the distribution of the uniform DIF results considering the two facets of P_PID (perception and interpretation/decision-making) between two of the three teacher groups since no items showed non-uniform DIF.

As Table 2 demonstrates, the low number of DIF differences between pre-service and early career teachers can be identified as a clear pattern, which confirms the overall results. We found that more items showed DIF between pre-service and experienced teachers and early career and experienced teachers. By contrast, only three items showed moderate DIF between pre-service teachers and early career teachers. Between the early career and experienced teacher groups, more items were found to favor early career teachers on the sub-facet of perception, and by contrast, more items were found to favor experienced teachers on the sub-facet of interpretation and decision-making. However, between the pre-service and experienced teacher groups, a similar number of items were found to favor both groups on the sub-facet of perception, and relatively more items were again found to favor experienced teachers.

Content analysis was further conducted on the items with DIF for the sub-facet of perception. For items with DIF between pre-service and early career teachers, the item favoring pre-service teachers was related to classroom management (“it takes a long time for the students to calm down and the lesson to start”) and the item favoring early career teachers was related to students’ behavior (“most students take an active part in the lesson”).

Table 2 Distribution of DIF Items on P_PID

		Number of class B items		Number of Class C Items	
		Perception	Interpretation and decision-making	Perception	Interpretation and decision-making
Pre-service teachers (PT) vs early career teachers (ECT)	Favor PT	1	0	0	0
	Favor ECT	1	1	0	0
Early career teachers (ECT) vs experienced teachers (ET)	Favor ECT	2	0	2	1
	Favor ET	0	0	1	3
Pre-service teachers (PT) vs experienced teachers (ET)	Favor PT	1	0	2	1
	Favor ET	2	0	2	3

Class B (moderate DIF): $0.035 < \Delta R^2 \leq 0.07$, Class C (large DIF): $0.07 < \Delta R^2$

For items with DIF between early career and experienced teachers, content analysis revealed that the main focus of the items favoring early career teachers was to investigate teachers’ perceptions related to classroom management (e.g., “it takes a long time for the students to calm down and the lesson to start”) and students’ behavior, focusing particularly on students’ participation in discussion (e.g., “the students take part in the lesson discussion”). The only item favoring experienced teachers was related to the accuracy of teachers’ instruction (“the teachers’ instruction is very precise”). For items with DIF between pre-service and experienced teachers, the items favoring pre-service teachers were mainly related to the teachers’ behavior (“the teacher presents the central question of the lesson orally and in writing”). By contrast, the four items favoring experienced teachers were mainly related to students’ thinking (e.g., “the teacher ensures that students have time for individual thinking”).

Concerning the sub-facet of interpretation/decision-making we found that for items with DIF between pre-service and early career teachers, only one item favored early career teachers, requiring teachers to suggest methods to make their teaching less teacher-centered. For items with DIF between early career and experienced mathematics teachers, the same item was again found to favor early career teachers. The three items favoring experienced teachers contained more teaching-related reflections, such as evaluating why the teachers’ comment is not helpful for developing students’ cognitive activities or suggesting teaching methods to use instruction time efficiently, and identifying the teaching phase and modifying it accordingly to better address the differences in individual students’ mathematical abilities. Interestingly, for items with DIF between pre-service and experienced teachers, the same item (requiring teachers to suggest teaching methods to make teaching less teacher-centered) was found to favor pre-service mathematics teachers, and the same three items were found to favor experienced teachers.

Overall, the DIF results on the P_PID items indicated no significant difference between pre-service and early career teachers. However, for the sub-facet of perception, pre-service mathematics teachers were found to be stronger than experienced teachers in aspects related to teachers’ behavior. Early career teachers exhibited strengths in paying attention to students’ roles, particularly students’ opportunities for participation in discussions. By contrast, experienced teachers exhibited strengths in items related to teaching accuracy and students’ thinking.

In addition, for the noticing sub-facet of interpretation/decision-making, pre-service and early career teachers exhibited strengths in suggesting methods to make teaching less teacher-centered, a more recent approach to teaching inspired by Western approaches. By contrast, experienced teachers performed better in evaluating teachers’ behavior and more traditional lesson organization, either by suggesting methods to effectively use lesson time or by analyzing individual students’ answers.

4.3 Specific strengths and weaknesses in M_PID as indicated by DIF

To identify which items on the sub-facets of M_PID typically favored which specific group of teachers, three separate pairs of DIF analyses were conducted. Again, no items showed non-uniform DIF. Table 3 summarizes the distribution of the DIF results considering the two facets of M_PID (perception and interpretation/decision-making) between each pair of the three groups of teachers.

As Table 3 indicates, for the sub-facet perception referring to M_PID, no items showed DIF between pre-service and early career mathematics teachers. Moreover, when compared with experienced teachers, only one item was found to favor early career or pre-service teachers, with relatively more items favoring experienced teachers. For the sub-facet of interpretation and decision-making referring to M_PID, generally speaking, between each pair of

Table 3 Distribution of DIF Items on M_PID

		Number of class B items		Number of class C items	
		Perception	Interpretation and decision-making	Perception	Interpretation and decision-making
Pre-service teachers (PT) vs Early career teachers (ECT)	Favor PT	0	0	0	1
	Favor ECT	0	2	0	0
Early career teachers (ECT) vs experienced teachers (ET)	Favor ECT	1	0	0	5
	Favor ET	3	2	2	3
Pre-service teachers (PT) vs experienced teachers (ET)	Favor PT	1	4	0	2
	Favor ET	1	3	2	2

Class B (moderate DIF): $0.035 < \Delta R^2 \leq 0.07$, Class C (large DIF): $0.07 < \Delta R^2$

the three groups of teachers, similar numbers of items were found to favor each of them. However, as Table 3 indicates, more items demonstrated DIF between pre-service and experienced teachers and early career and experienced teachers.

Content analysis was further conducted on all items demonstrating DIF. For the sub-facet of perception of M_PID, the one item that was found to favor early career or pre-service teachers was the same item that required the teachers to judge whether the task had the characteristics of an open-ended task. The other items favoring experienced teachers required teachers to evaluate the correctness of a student's statement, whether a specific topic such as function and space and form is important during teaching, or whether the task shown in the video is already accessible for lower grade students, or required to determine whether the statement of one individual student was helpful in solving the main task shown in the video-vignette.

For the sub-facet of interpretation/decision-making referring to M_PID, for items with DIF between pre-service and early career teachers, the only item favoring pre-service teachers required teachers to use a term with a meaning similar to the meaning of 'enactive' to describe the critical characteristics of students' group work. The two items favoring early career teachers both related to the aspect of 'reality'; for example, the teacher mentioned that the topic closely relates to reality and modified the lesson task to make it more realistic.

For items with DIF between early career and experienced teachers and pre-service and experienced teachers, the items were all found to relate to the same tasks. For the items favoring early career teachers or pre-service teachers, four are from a cooperation task requiring the teachers to highlight the critical characteristics of students' group work in relation to mathematics education, more specifically distinguishing the different kinds of representation in mathematics, such as enactive, iconic, or symbolic representation. The other items required the teachers to modify the lesson problem to make it more realistic or foster students' modeling competence (the latter only favoring pre-service teachers).

Two of the items favoring experienced teachers were related to the same problem, which required the teachers to provide three indicators from the answer given by one student shown in the video that she solved the problem using a purely algorithmic approach without deeper understanding. Another item required the teachers to identify the main difference between a student's statement and his classmates' statements from a mathematical perspective (i.e., to correctly answer the item, the teachers needed to notice that the student was deducing or making a linear assumption). The last two items also related to the cooperation task, which required the teachers to identify the different approaches ('enactive' and 'symbolic') to describe the critical characteristics of students' work.

Overall, the content analysis of the characteristics of the items with DIF on the aspect of M_PID emphasizes that for the sub-facet of perception, no major differences exist between pre-service and early career teachers. Compared with experienced teachers, both pre-service and early career teachers demonstrated weaker professionalism on the aspects related to evaluating the correctness and usefulness of a student's statement and making a judgment concerning a specific aspect of mathematical content. By contrast, the early career and pre-service teachers were found to demonstrate stronger professional noticing on judging whether a task was open-ended.

For the sub-facet of interpretation and decision-making, stronger differences could be identified between pre-service and experienced teachers and between early career and experienced teachers. In general, the pre-service and early career teachers were found to demonstrate stronger professional noticing on modern and more Westernized themes, such as mathematical modeling or cooperative learning. By contrast, the experienced teachers showed professional noticing strengths on aspects related to more abstract topics and the inner nature of mathematics, such as using a specific term to describe students' group work and analyzing students' mathematical thinking.

5 Discussion

Our main aim in the present study was to compare teacher noticing among three teacher cohorts with different degrees of mathematics teaching experience, namely pre-service, early career, and experienced teachers, to describe the development of teachers' noticing influenced by teaching experience and to identify possible patterns of teacher noticing among teachers at specific developmental stages. The results suggest that for the two facets of teacher noticing investigated in the present study, almost a linear growth in teacher noticing can be traced as teaching experience increases. However, post hoc analysis demonstrated significant differences only between pre-service and experienced teachers, and between early career and experienced teachers. No significant differences were identified between pre-service and early career teachers for both P_PID and M_PID and their sub-facets. The findings are thus consistent with those of Jacobs et al. (2010). They also identified monotonic trends for all three sub-facets of teaching noticing among teachers with different professional development experiences of students' mathematical thinking. Such findings further suggest that teaching experience indeed acts as a main—though not sufficient—factor in the development of teacher noticing (Schoenfeld, 2011).

The linear growth of teacher noticing among the three teacher cohorts and the weak differences between

pre-service and early career teachers may first be explained by the tradition of Chinese mathematics teacher education. As mentioned above, pre-service mathematics teachers have few opportunities to improve their teaching practice skills within initial teacher education and are expected to develop their teaching skills further after they enter teaching positions (Paine et al. 2003). Such traditions hinder newly graduated teachers from the acquisition of necessary mathematics pedagogical content knowledge and teaching practice skills. It is thus understandable that there are no significant differences in teacher noticing between pre-service and early career teachers. However, the school-based professional development culture provides every in-service Chinese teacher with the opportunity to continuously develop his or her practical skills (Han and Paine 2010; Lu, Kaiser, and Leung 2020). Therefore, the existence of linear growth in teacher noticing among Chinese mathematics teachers is understandable as well as the significant differences between pre-service and experienced teachers and early career and experienced teachers.

The relationship between knowledge and teacher noticing may also further help to explain the present study's findings. It has hitherto been widely accepted that teacher knowledge has a fundamental impact on teacher noticing (Schoenfeld 2011; König et al. 2014; Yang et al. in press). In terms of teacher knowledge, such as mathematics content knowledge and mathematics pedagogical content knowledge, empirical studies have found that pre-service and early career mathematics teachers performed significantly more poorly than experienced teachers (Han, Ma, and Wu 2016; Kleickmann et al. 2013). However, early career teachers' knowledge remained largely unchanged during their first years of teaching (Blömeke et al. 2015a, b). Therefore, owing to the differences in teacher knowledge foundation among teachers with different teaching experience, it can be expected that they will also perform differently on noticing-related tasks.

DIF detection results reveal further interesting differences among the three teacher cohorts. First, several items showed DIF between pre-service and early career teachers. As reported above, only six items in total showed DIF, with four favoring early career teachers and two favoring pre-service teachers. The findings again confirm that relatively little development occurs in terms of teacher noticing between pre-service and early career teachers. However, it is worth highlighting several small differences between these two groups of teachers. Those items favoring early career teachers are mainly related to the perception of students' roles within teaching and learning processes, suggesting teaching methods to make teaching less teacher-centered and modifying instructional tasks. However, the tasks favoring pre-service teachers are mainly related to classroom management. These findings suggest that although the development is small, after a couple of years of teaching in the classroom,

Chinese early career teachers shift their focus from classroom management to students' roles and instructional methods. Similar findings have been found in non-Chinese contexts. After an intervention period, teachers were also found to attend to salient characteristics of mathematics instruction and students' situations, such as students' mathematical engagement (e.g., McDuffie et al. 2014; Mitchell and Marin 2015; Stockero, Rupnow and Pascoe 2017).

Secondly, more items were found to show DIF between pre-service and experienced teachers and between early career and experienced teachers, which again confirms the significant difference between less-experienced and experienced teachers. For the sub-facet perception within P_PID, the items favoring pre-service mathematics teachers were mainly related to teachers' behaviors, the items favoring early career teachers were related to students' discussion opportunities, and the items favoring experienced teachers were related to accuracy of teaching and students' thinking. To a certain degree, these results are consistent with the findings in previous studies. As reviewed above, according to expertise research, pre-service mathematics teachers typically attend to disciplinary issues and teacher moves during teaching rather than the salient characteristics of mathematics instruction (Berliner 2001; McDuffie et al. 2014). Moreover, earlier studies from expertise research found that, compared with teachers at other developmental stages, expert teachers organize learning processes quite differently: expert teachers tend to spend considerable time at the beginning of the school year establishing classroom norms and routines so that they can focus on teaching afterwards (Berliner 2004; Tsui 2003, 2009). As experienced teachers are used to well-established classroom routines, they may pay less attention to classroom management issues such as teaching disruptions but attend more to events closely related to teaching, such as students' errors and thinking.

Thirdly, relatively more items concerning the sub-facet interpretation/decision-making were found to favor the experienced teachers, requiring them to use either pedagogy-related or mathematics-related knowledge to interpret students' work and develop corresponding decisions, such as suggesting ways to use lesson time effectively, and analyzing students' mathematical thinking or answers. Similar findings have been reported in earlier research; for example, previous studies found that expert or more experienced teachers were more able to interpret students' understanding based on the mathematical elements used (e.g., Callejo and Zapatera 2017) or showed superior expertise in deciding how to adequately respond to their perception of students' understanding (Jacobs et al. 2010).

The differences between the experienced teachers and pre-service and early career teachers in the sub-facet interpretation/decision-making can be explained by the differences in the knowledge foundation among the three groups

of teachers. As mentioned above, experienced mathematics teachers were found to possess a more solid foundation in MCK and MPCK (Han et al. 2011). Recent studies on teacher noticing found that MPCK was more strongly needed during the process of interpreting and responding to students' mathematical understanding (Sánchez-Matamoros, Fernandez and Llinares 2019). Therefore, a profound understanding of mathematics itself and more developed and accessible MPCK allowed the experienced mathematics teachers in the study to identify students' mathematical misconceptions and to propose different ways to deal with these misconceptions, with approaches spanning from graphic to symbolic explanations.

Moreover, China's school-based professional development culture also helps to explain the differences. As mentioned above, Chinese teachers are expected to further develop their teaching skills after they have begun teaching through school-based activities (Han and Paine 2010; Lu et al., 2020). In these activities, experienced teachers will typically take the leading role to help newly graduated teachers devise different lesson plans for the same topic or to comment on or evaluate less-experienced teachers' teaching with the aim of repeatedly modifying lessons to improve teaching quality (Huang, Su and Xu, 2014). Therefore, it was expected that the experienced teachers in this study would exhibit superior skills in modifying lessons and suggesting ways to facilitate students' understanding.

The identified differences in teacher noticing related to the sub-facet interpretation/decision-making among the three groups of teachers can be further explained by results from the expertise research, including expert teachers' ability to teach more flexibly and swiftly and to provide more meaningful decisions for further teaching (Berliner 2001, 2004; Livingston and Borko 1989). Since the experienced teachers in the study all had over 15 years' teaching experience, it was expected that many would have developed the expertise to interpret teaching events meaningfully and adjust their teaching flexibly.

Aside from the patterns elaborated above, several other findings merit discussion. First, concerning the sub-facet of interpretation/decision-making from P_PID, compared with experienced teachers, pre-service and early career teachers were better able to identify the essential weaknesses or shortcomings in students' group work, as shown in the video-vignettes. Furthermore, for the same sub-facet from the perspective of M_PID, pre-service and early career teachers demonstrated stronger noticing skills concerning aspects related to open-ended tasks or modifying modeling tasks. Moreover, for the sub-facet of perception of P_PID, early career teachers demonstrated specific strength on the aspects related to students' opportunities for discussion. These more unexpected results can be interpreted and understood considering the mathematics education tradition

and culture in China. Group work, students' discussion, and mathematical modeling are relatively new topics in China, introduced only in the most recent mathematics curriculum from 2000, which encouraged cooperative learning and mathematical modeling. Previously, mathematics teaching in China could be described as relatively traditional, with dominant teacher talk, routine tasks and many mathematical exercises (Leung 2001). Consequently, experienced teachers in the study may have lacked the knowledge or experience to organize effective cooperative mathematics learning and carry out mathematical modeling and could therefore neither identify students' weaknesses in group work nor modify modeling tasks. Furthermore, experienced teachers may even doubt the usefulness of group work and modeling tasks in teaching. Recent studies found that teachers' transmissive beliefs hinder them from professionally observing classroom situations (Meschede, Fiebranz, Möller and Stefensky 2017).

Finally, the present study's findings further confirm that the three sub-facets of teacher noticing—perception, interpretation, and decision-making—may develop differently and at different paces. More specifically, both the overall and DIF results suggest that perception is better developed at the pre-service and beginning stages of teaching. The pre-service and early career teachers were better able to perceive events related to classroom management, teacher behavior, and process-oriented mathematical skills such as open-ended tasks and mathematical modeling, which they may have learned during their university study. By contrast, the findings reported above suggest that the sub-facets of interpretation and decision-making are more difficult to develop. This is consistent with findings identified in other contexts (e.g., Callejo and Zapatera 2017; Jacobs et al. 2010). As these authors observe, the sub-facets of interpretation and decision-making, especially those related to mathematics instruction, may need more deliberate practice to achieve a certain level of proficiency. However, the sub-facet of perception may be the easiest to develop, particularly in relation to general pedagogical issues. However, more studies are needed in this respect before secure conclusions can be drawn.

6 Conclusions and limitations

Although many previous studies investigated the development of teacher noticing in mathematics education, little empirical evidence has hitherto been available in relation to possible patterns of strengths and weaknesses in noticing for teachers at different developmental stages. The present study's central goal was to close this gap by evaluating the growth of teacher noticing among three cohorts of mathematics teachers with different lengths of teaching

experience, within authentic contexts rather than intervention contexts, as in many of the previous studies. As few studies have been conducted in non-Western cultures, this study seemed overdue, as noticing is considered a culturally shaped construct (Louie 2018; Yang et al. 2019).

As reported above, the findings in the present study indicate a nearly linear growth of teacher noticing among Chinese mathematics teachers, as was pointed out in the study by Jacobs et al. (2010). However, significant differences were identified only between pre-service and experienced teachers, no significant differences appeared between pre-service and early career teachers. DIF results further suggest that for the noticing sub-facet of perception, pre-service mathematics teachers tended to demonstrate strength in relation to teacher behavior and judging whether a task was open-ended; early career teachers demonstrated particular strength in relation to students' opportunities for discussion; and experienced teachers tended to show strength in aspects related to teaching accuracy, students' thinking, and correctness in students' statements. For the noticing sub-facet of interpretation and decision-making, pre-service and early career teachers exhibited strength in aspects related to more reformed or Western approaches to mathematics teaching, such as identifying critical characteristics of cooperative learning and mathematical modeling-related tasks. By contrast, experienced teachers showed strength in relation to evaluating teachers' behavior and analyzing students' mathematical thinking. The present study's findings also indicate differences in the development rates of the three sub-facets of teacher noticing.

Although the present study is one of the few empirical studies hitherto to have used cross-sectional design in a non-laboratory setting in China to investigate the development process of teacher noticing, its limitations should be mentioned. First, the experienced teachers were mainly chosen from one administrative area, and the pre-service teachers were mainly chosen from two normal universities in China. Therefore, the samples may not be sufficiently typical and representative to reflect the general situation or diversity of mathematics teacher noticing in China. Future studies should consider teachers from a wider geographical range and include teachers from other school levels, such as primary or senior upper secondary school level.

In addition, only three cohorts of teachers with different teaching experiences were included. Further studies are needed, including teachers with a greater variety of teaching experience. Such studies would provide further insightful and meaningful information about teachers' noticing trajectories in mathematics education.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463–482.
- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science Technology and Society*, *24*(3), 200–212.
- Blömeke, S., Gustafsson, J., & Shavelson, R. J. (2015a). Beyond dichotomies—Competence viewed as a continuum. *Zeitschrift für Psychologie*, *223*(1), 3–13.
- Blömeke, S., Hoth, J., Döhrmann, M., Busse, A., Kaiser, G., & König, J. (2015b). Teacher change during induction: development of beginning primary teachers' knowledge, beliefs and performance. *International Journal of Science and Mathematics Education*, *13*(2), 287–308.
- Blömeke, S., Suhl, U., & Döhrmann, M. (2013). Assessing strengths and weaknesses of teacher knowledge in Asia, Eastern Europe and Western countries: differential item functioning in TEDS-M. *International Journal of Science and Mathematics Education*, *11*, 795–817.
- Bond, T., & Fox, C. (2007). *Applying the Rasch model: fundamental measurement in the human sciences (2nd)*. Mahwah: LEA.
- Callejo, M. L., & Zapatera, A. (2017). Prospective primary teachers' noticing of students' understanding of pattern generalization. *Journal of Mathematics Teacher Education*, *20*(4), 309–333.
- Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education*, *39*(3), 25–31.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.
- Gess, C., Wessels, I., & Blömeke, S. (2017). Domain-specificity of research competencies in the social sciences: evidence from differential item functioning. *Journal for Educational Research Online*, *9*(2), 11–36.
- Han, J., Ma, Y., & Wu, Q. (2016). A study of pre-service mathematics teachers' professional knowledge (in Chinese). *Teacher Education Research*, *28*(3), 67–72.
- Han, J., Wong, N., Ma, Y., & Lu, N. (2011). Research on teacher knowledge of middle school teachers—based on maths teachers in municipalities in Northeast China (in Chinese). *Educational Research*, *375*(4), 91–95.
- Han, X., & Paine, L. (2010). Teaching mathematics as deliberate practice through public lessons. *The Elementary School Journal*, *110*(4), 519–541.

- Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement*, 5(2–3), 81–92.
- Holland, P., & Wainer, H. (1993). Concluding remarks and suggestions. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 419–422). Hillsdale: Lawrence.
- Hoth, J., Schwarz, B., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Uncovering predictors of disagreement: ensuring the quality of expert ratings. *ZDM—Mathematics Education*, 48(1–2), 83–95.
- Huang, R., & Li, Y. (2012). What matters most: a comparison of expert and novice teachers' noticing of mathematics classroom events. *School Science and Mathematics*, 112(7), 420–432.
- Huang, R., Su, H., & Xu, S. (2014). Developing teachers' and teaching researchers' professional competence in mathematics through Chinese lesson study. *ZDM—Mathematics Education*, 46(2), 239–251.
- Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, 41, 169–202.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329–349.
- 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.
- Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the complexities of video-based assessments: theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *International Journal of Science and Mathematics Education*, 13(2), 369–387.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319–342.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, 38, 76–88.
- Kyndt, E., Gijbels, D., Grosemans, I., & Donche, V. (2016). Teachers' everyday professional development: mapping informal learning activities, antecedents and learning outcomes. *Review of Educational Research*, 86(4), 1111–1150.
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Sciences*, 44, 197–203.
- Leinhardt, G., Putnam, R. T., Stein, M. K., & Baxter, J. (1991). Where subject knowledge matters. In J. E. Brophy (Ed.), *Advances in research on teaching: teachers' subject matter knowledge and classroom instruction* (Vol. 2, pp. 87–113). Greenwich: JAI Press.
- Leung, F.K.S. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics*, (1), 35–51.
- Li, S., Huang, R., & Shin, Y. (2008). Discipline knowledge preparation for prospective secondary mathematics teachers: an East Asian perspective. In P. Sullivan & T. Wood (Eds.), *Knowledge and beliefs in mathematics teaching and teaching development* (pp. 63–86). Rotterdam: Sense Publishers.
- Li, Y., & Huang, R. (2018). *How Chinese acquire and improve mathematics knowledge for teaching*. Boston: Sense Publishers.
- Livingston, C., & Borko, H. (1989). Expert-novice differences in teaching: a cognitive analysis and implications for teacher education. *Journal of Teacher Education*, 40(4), 36–42.
- Louie, N. L. (2018). Culture and ideology in mathematics teacher noticing. *Educational Studies in Mathematics*, 97(1), 55–69.
- Lu, X., Kaiser, G., & Leung, F. (2020). Mentoring early career mathematics teachers from the mentees' perspective—a case study from China. *International Journal for Science and Mathematics Education*, 18, 1355–1374.
- McDuffie, A. R., Foote, M. Q., Bolson, C., Turner, E. E., Aguirre, J. M., Bartell, T. G., et al. (2014). Using video analysis to support pre-service K-8 teachers' noticing of students' multiple mathematical knowledge bases. *Journal of Mathematics Teacher Education*, 17(3), 245–270.
- Meschede, N., Fiebranz, A., Möller, K., & Steffensky, M. (2017). Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teaching and Teacher Education*, 66, 158–170.
- Mesic, V. (2012). Identifying country-specific cultures of physics education: a differential item functioning approach. *International Journal of Science Education*, 34(16), 2483–2500.
- Mitchell, R., & Marin, K. (2015). Examining the use of a structured analysis framework to support prospective teacher noticing. *Journal of Mathematics Teacher Education*, 18(6), 551–575.
- Paine, L. W., Fang, Y., & Wilson, S. (2003). Entering a culture of teaching. In E. Britton, L. Paine, D. Pimm, & S. Raizen (Eds.), *Comprehensive teacher induction: systems for early career learning* (pp. 20–82). Dordrecht: Kluwer.
- Patrick, W. (2010). *Recognising non-formal and informal learning outcomes, policies and practices: outcomes, policies and practices*. Paris: OECD Publishing.
- Sánchez-Matamoros, G., Fernández, C., & Llinares, S. (2019). Relationships among prospective secondary mathematics teachers' skills of attending, interpreting and responding to students' understanding. *Educational Studies in Mathematics*, 100(1), 83–99.
- Santagata, R., König, J., Scheiner, T., Nguyen, H., Adleff, A.-K., Yang, X., & Kaiser, G. (2021). Mathematics teacher learning to notice: A systematic review of studies of video-based programs. *ZDM—Mathematics Education*, **this issue**.
- Schack, E. O., Molly, H. F., & Wilhelm, J. A. (Eds.). (2017). *Teacher noticing: bridging and broadening perspectives, contexts, and frameworks*. New York: Springer.
- Schoenfeld, A.H. (2011). Noticing matters. A lot. Now what? In M.G. Sherin, V.R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 223–238). New York: Routledge.
- Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011a). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.
- Sherin, M.G., Jacobs, V.R., & Philipp, R.A. (2011). Situating the study of teacher noticing. In M.G. Sherin, V.R. Jacobs, & R.A. Philipp (2011). *Mathematics teacher noticing. Seeing through teachers' eyes* (pp. 3–13). New York: Routledge.
- Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, 13, 475–491.
- Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60, 20–37.
- Simpson, A., Vondrová, N., & Žalská, J. (2018). Sources of shifts in pre-service teachers' patterns of attention: the roles of teaching experience and of observational experience. *Journal of Mathematics Teacher Education*, 21(6), 607–630.
- Stockero, S. L., Rupnow, R. L., & Pascoe, A. E. (2017). Learning to notice important student mathematical thinking in complex classroom interactions. *Teaching and Teacher Education*, 63, 384–395.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361–370.

- Tsui, A. B. M. (2003). *Understanding expertise in teaching: case studies of second language teachers*. Cambridge: Cambridge University Press.
- Tsui, A. B. M. (2009). Distinctive qualities of expert teachers. *Teachers and Teaching*, 15(4), 421–439.
- Wilkie, K. J. (2019). The challenge of changing teaching: Investigating the interplay of external and internal influences during professional learning with secondary mathematics teachers. *Journal of Mathematics Teacher Education*, 22(1), 95–124.
- Wolff, C. E., Jarodzka, H., Van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, 44(3), 243–265.
- Yang, X., Kaiser, G., König, J., & Blömeke, S. (in press). Relationship between Chinese mathematics teachers' knowledge and their professional noticing. *International Journal of Science and Mathematics Education*, <https://doi.org/10.1007/s10763-020-10089-3>.
- Yang, X., Kaiser, G., König, J., & Blömeke, S. (2018). Measuring Chinese teacher professional competence: adapting and validating a German framework in China. *Journal of Curriculum Studies*, 50(5), 638–653.
- Yang, X., Kaiser, G., König, J., & Blömeke, S. (2019). Professional noticing of mathematics teachers: a comparative study between Germany and China. *International Journal of Science and Mathematics Education*, 17, 943–963.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.