



Student Characteristics in the Eyes of Teachers: Differences Between Novice and Expert Teachers in Judgment Accuracy, Observed Behavioral Cues, and Gaze

Tina Seidel¹ · Katharina Schnitzler¹ · Christian Kosel¹ · Kathleen Stürmer² · Doris Holzberger¹

Published online: 11 May 2020
© The Author(s) 2020

Abstract

The present study investigates teacher diagnostic skills when observing student engagement and inferring to underlying student characteristic profiles. Five student profiles as empirically determined in previous studies are selected: three incoherent (overestimating, uninterested, and underestimating) and two coherent (strong and struggling) profiles. Teacher professional vision and underlying assumptions about processes of noticing and reasoning about the chosen diagnostic situation serve as a conceptual basis. In the empirical study ($N=41$ participants), it is investigated to what extent expert and novice teachers differ with regard to judgment accuracy of underlying student profiles, observed student cues used for judgment, and teacher gaze as perceptual indicator. The study task involved observing a video clip and diagnosing five marked students based on their underlying profiles. First, findings of the study suggest that expert teachers are more accurate in judging incoherent profiles compared to novices. Second, both novices as well as experts state valid behavioral cues when inferring from student engagement to underlying student profile. Third, experts spend more teacher gaze on student profiles which might need adaptive pedagogical action (struggling, underestimating, uninterested student). The study provides first evidence on teacher gaze during the professional task of diagnosing individual students in the process of teaching. Regarding the conceptual model of teacher professional vision teacher gaze can serve as an additional operationalization of the noticing component of teacher professional vision.

Keywords Teacher professional vision · Teacher gaze · Eye tracking · Teacher judgment accuracy · Diagnostic activities · Individual student characteristics · Student engagement

✉ Tina Seidel
tina.seidel@tum.de

In everyday classroom teaching, teachers have to make a myriad of decisions and adaptations, which strongly depend on the way their students behave (Borko et al. 2008). A central demand of being a teacher, therefore, is to continuously monitor students and observe how they react to planned activities (Berliner 1986; Shulman 1986).

In looking closer at student-related factors that influence how students respond to teachers' planned activities (e.g., by answering a teacher question or working on an instructional task), cognitive and motivational-affective characteristics come into play as prerequisites for learning (Snow et al. 1996). Pre-knowledge, self-concept of ability, and interest are major determinants for student engagement in class (Hattie 2008; Hoffmann 2002; Huber et al. 2015; Linnenbrink-Garcia et al. 2012; Lipowsky et al. 2008; Pauli and Lipowsky 2007).

As student characteristics are a central driver for student engagement, and student engagement, in turn, determines teachers' behavior, teachers need to monitor their students' engagement and make inferences with regard to underlying student characteristics. These kinds of formative, diagnostic activities are seen as a key component of teachers' professional competences (Heitzmann et al. 2019; Kunter et al. 2013). In this context, teachers' professional vision is another relevant aspect that must be considered (Lachner et al. 2016). Teachers' professional vision is characterized by teachers' ability to notice relevant events in a classroom and interpret these events based on professional knowledge (Seidel and Stürmer 2014). Teacher knowledge—such as content, pedagogical-content, and pedagogical-psychological knowledge—shapes professional vision because it affects the direction of the attention processes (Blömeke et al. 2015). In the context of observing student engagement and inferring underlying characteristics, teachers' pedagogical-psychological knowledge is particularly relevant (Kunter et al. 2013). Thereby, it has been shown that experts (representing professionals with high knowledge) are better able to monitor the responses and activities of all students constantly while, at the same time, being alert to those students and events that might require particular action or adaptations (Berliner 1991; Borko et al. 2008; Wolff et al. 2016).

Recently, an increasing number of studies has looked more deeply into aspects of teachers' visual perception of student engagement in classrooms (Seidel et al. 2016; van den Bogert et al. 2014; Wolff et al. 2016). Teachers' visual perception is seen as an important aspect of the noticing component of professional vision (Lachner et al., 2016). In this context, teacher gaze can be studied using an eye tracking methodology. So far, teacher eye movement studies have been mainly concerned with issues of classroom management (van den Bogert et al. 2014; Wolff et al. 2016), without considering the observation and diagnosis of individual students characteristics. However, monitoring individual students' engagement behaviors and linking these perceptions to interpretations of underlying student characteristics are also highly relevant diagnostic skills. Therefore, this study tries to expand the field and explore the potential of teacher eye movement methodology in the context of teachers observing student engagement and diagnosing underlying, individual student characteristics.

Diagnosing Student Characteristics as a Professional Diagnostic Task

Student characteristics as prerequisites for learning are the most decisive factor in predicting student engagement and also further learning achievement (Hattie 2008). Therefore, teachers' diagnostic skills at monitoring student engagement and accurately judging underlying student characteristics is regarded as an important element of professional teacher competences (Heitzmann et al. 2019; Herppich et al. 2018).

Student characteristics comprise both cognitive and motivational-affective components (Snow et al. 1996). Cognitive characteristics refer to students' general cognitive abilities as well as their acquired knowledge in subject domains. Motivational-affective components include aspects such as interest and self-concept of ability in subject domains. However, cognitive and motivational-affective components do not necessarily develop uniformly. Students, for example, may have high cognitive abilities but low self-concept (Seidel 2006).

Therefore, based on student data, combinations of cognitive and motivational-affective characteristics can be modeled by person-centered approaches (Seidel 2006). The resulting profiles represent the most prevalent patterns of student characteristic combinations. They vary in the consistency of the development of cognitive and motivational-affective characteristics. Thus, the names of the profiles represent the type of characteristics combination. Some groups of students acquire consistent high or low combinations of cognitive and motivational-affective characteristics (rated as either strong or struggling students in all characteristics), while other students show inconsistent combinations (Jurik et al. 2013; Linnenbrink-Garcia et al. 2012). An example of an inconsistent profile is so-called underestimating students. These students have developed a pattern of being highly knowledgeable but, at the same time, expressing low confidence in their abilities (self-concept). Another inconsistent pattern is the overestimating students. These students are comparably unknowledgeable but highly confident. Therefore, these students are in danger of overestimating their abilities. The third inconsistent pattern is uninterested students, who are intermediate in their knowledge and confidence but stand out because of their particular low interest in a subject domain.

Besides determining student characteristics based on student data, it is of interest to study how accurately teachers judge relevant student characteristics. Thereby, in a majority of studies a focus has been placed on judgments of single cognitive or motivational-affective characteristics (Kaiser et al. 2013). It was found that teachers diagnose student knowledge more accurately than they do general cognitive abilities, self-concept, or interest. In addition, teachers tend to perceive students holistically and intermingle distinct student characteristics when prompted to diagnose, for example, student achievement or motivation only (Südkamp et al. 2018). Recent teacher judgment studies, therefore, have expanded upon the knowledge stemming from student characteristic profiles and compare teacher and student perspectives of profiles (Huber and Seidel 2018; Südkamp et al. 2018). These studies have found that teachers tend to judge their students consistently as either over-average, average, or under-average across cognitive and motivational-affective characteristics. However, profiles from student reports again yield a much larger variety, including inconsistent profiles, such as underestimating, overestimating, or uninterested.

Given these inconsistencies between the perspectives of teachers and students further research is required to study which kind of misjudgments occur at which points in time. For example, to date, only a few studies have investigated student-related cues that teachers use when coming to conclusions (Spinath 2005). Demographic information about students, such as socio-economic status, migration status, and gender, has been investigated an influential factor (Glock et al. 2013; Kaiser et al. 2017), particularly in the context of high-stakes decisions, such as making recommendations for further educational careers. However, few studies discuss diagnostic activities in the teaching process that are necessary to learning about relevant behavioral cues that influence adaptations in the process of teaching (Huber and Seidel, 2018; Herppich et al. 2018).

Diagnostic Process of Interpreting Student Engagement in Relation to Characteristics

In order to judge and monitor student characteristic profiles in everyday classroom teaching accurately, teachers are required to observe students' behavioral cues (Back and Nestler 2016), which must be valid with regard to underlying student characteristics (Thiede et al. 2015). Therefore, it is necessary to study which kind of behavior is typically presented by groups of students who share a specific profile.

In video-based classroom research, student engagement is systematically observed and analyzed. Some studies differentiate between high- and low-achieving students and show that high-achieving students are typically more verbally engaged with their teachers compared to low-achieving students (Brophy and Good 1970; Lipowsky et al. 2007; Pauli and Lipowsky 2007; Sacher 1995). They also experience more frequently positive teacher feedback. Many previous studies, however, did not consider motivational-affective student characteristics or the combination of motivational-affective characteristics with achievement. Therefore, further studies (Huber et al. 2015; Jurik et al. 2013; Seidel and Stürmer 2014) analyzed student engagement based on the student profiles identified by Seidel (2006). They found that self-concept of ability was particularly predictive of student engagement. The strong and overestimating student profiles showed the same high-engagement pattern, despite the fact that the overestimating students were rather unknowledgeable. In comparison, struggling and underestimating students were rather unengaged and showed similar behavioral cues. The underestimating students' low engagement was in contrast to their underlying knowledge, which was comparable to that of the strong students. Finally, the fifth identified profile of uninterested students showed a rather negative engagement pattern, with reports of experiencing amotivation and no relevance in learning (Seidel 2006).

These classroom observation findings indicate that student engagement patterns, in some cases, can be quite valid with regard to underlying student characteristics (e.g., strong or struggling students). However, regarding incoherent profiles, students' behavioral cues are in danger of being misinterpreted (Huber et al. 2015; Jurik et al. 2013). Therefore, teachers need to observe and diagnose these students carefully in order to differentiate their instruction between them. An underestimating student, for example, might require pedagogically relevant actions such as positive feedback and reinforcement and a struggling student might need further instructional guidance (Huber and Seidel 2018). Regarding overestimating students, teachers might need to recognize that these students are not as knowledgeable as they may appear; similarly, strong students might need to be encouraged to continue their high-level learning.

When inferring underlying characteristics based on behavioral student engagement, teachers must be sensitive to the available student information (Berliner 1991). Some student cues related to the underlying student profiles may be more or less directly observable by teachers' visual perception. However, little research has been conducted into the relationship between perceptual processes of observing students' behavioral cues and judgments about underlying student characteristic profiles (Spinath 2005; Südkamp et al. 2018). Therefore, linking observable student engagement to underlying student characteristic profiles and relating these to possible challenges to accurate teacher judgments appear to be a logical step for future research.

Teacher Gaze in the Context of Studying Perceptual Processes

Teacher decision-making in the teaching process is highly dependent on perceptual processes when observing student engagement and, in turn, drawing inferences to underlying student characteristics (Borko et al. 2008). The way teachers' professional knowledge is organized and structured is one of the most important factors in explaining the qualitative differences involved in managing this professional task (Heitzmann et al. 2019). Expert–novice comparisons (Berliner 1986, 1991; Stürmer et al. 2017; Wolff et al. 2016; Wolff et al. 2015), for example, reveal that expert teachers (with high professional knowledge) are better able to distribute their attention equally across all students while teaching (as an indicator for successful classroom management) and to move more smoothly from one event to the next than novices. Novices, by comparison, are more likely to follow salient events or student cues erratically, moving from the perceiving written notes about their lesson plan, for example, to observing students in the classrooms and then the black board or other kind of media (McIntyre et al. 2017; Stürmer et al. 2017; Wolff et al. 2016).

Besides the ability to distribute attention equally, expert teachers are also able to focus on those cues and events that are relevant to further learning (Berliner 1986, 1991). Since their professional knowledge is organized based on typical events, schemata, and routines, single cues are sufficient for them to be alerted and activate their specific schemata and scripts (Borko et al. 2008). Novices, by comparison, have not yet structured their knowledge in this way and, therefore, tend to miss these kinds of cues and events (McIntyre et al. 2017; Wolff et al. 2016).

Studying teachers' visual perception processes and relating them to underlying professional knowledge structures has been a major advancement in teacher expertise research (Berliner 1986, 1991; Borko et al. 2008). Such research has recently expanded to include the concept of teachers' professional competences (Herppich et al. 2018). Teachers' professional vision (Goodwin 1994) is a concept currently applied in teacher research (Gaudin and Chaliès 2015) and is regarded as a situative application of skills in the interplay between professional knowledge and practical performance (Blömeke et al. 2015; Lachner et al. 2016). Teachers' professional knowledge drives the professional vision processes of noticing and interpreting classroom events. However, when observing classroom events, teachers' professional vision also activates their knowledge in the form of scripts, schemata, and routines reciprocally (Borko et al. 2008).

In a cognitive model of teachers' professional vision, teachers' perceptual processes play an important role, particularly with regard to the noticing component (Lachner et al. 2016). Teacher gaze as measured, for example, by an eye movement methodology (Rayner 2009), adds to the model since it provides a suitable operationalization for noticing processes in the forefront of further processing and interpreting observed events (Seidel and Stürmer 2014). Teacher gaze is driven by bottom-up and top-down processes (Gegenfurtner et al. 2011). Bottom-up drivers are autonomous, reactive, stimulus-driven, and related to salient features, such as the current flow of visual stimuli and motion (Schütz et al. 2011). Thereby, bottom-up drivers may play a critical role when teachers observe visually broader and ongoing motions of students in a classroom as for example during teacher–students interactions. In such situations, involving a teacher asking a question at the front of a classroom, students in the back of the room raising their hands, and one student being called upon and answering the question (Seidel and Prenzel 2006), a lot of motions occur simultaneously, and the complexity, immediacy, and unpredictability of teaching are fully apparent (Doyle 2006). In such motion-rich situations with a lot of active student engagement and salient behaviors, it can be assumed that teacher gaze follows the major motions of teachers and students and is highly driven by bottom-up processes.

Top-down processes have been shown to be driven by a specific task and personal goals (DeAngelus and Pelz 2009). They are also more likely to display differences in the expertise level (Gegenfurtner et al. 2011; Jarodzka et al. 2010). These top-down processes can occur when, for example, teachers observe a student more closely due to their personal goal of trying to judge a learning state or underlying student characteristic accurately. In this task-specific and goal-related context, top-down processes seem to be the more likely drivers of teacher gaze. In this context, differences between expert and novice teacher should also become apparent.

The Present Study

Currently, teacher gaze is primarily studied in the context of classroom management (van den Bogert 2016; van den Bogert et al. 2014; Wolff et al. 2016; Wolff et al. 2015), without further consideration of diagnosing individual students and their characteristics. However, the ability to monitor not only the whole class but also individual students is a highly relevant professional skill for teachers (Heitzmann et al. 2019). Therefore, this study attempts to expand the field and explore the potential of teacher eye movement methodology in the context of teachers observing students' behavioral engagement and diagnosing underlying individual student characteristics. Based on previously identified student characteristic profiles (Seidel 2006), this study utilizes two consistent profiles, strong and struggling, and three inconsistent profiles: overestimating, underestimating, and uninterested. The following questions and hypotheses are addressed:

RQ1: Do novice and expert teachers vary in the accuracy of their judgment of the five selected student characteristic profiles when asked to judge profiles based on observed student engagement?

Based on previous research, it is assumed that expert teachers in general are more accurate than novice teachers at judging student profiles. However, even expert teachers tend to make more inaccurate judgments when it comes to the incoherent profiles (over- and underestimating) compared to consistent profiles.

RQ2: Which behavioral indicators do expert and novice teachers use when inferring underlying student characteristic profiles based on their observations?

Since teacher judgment research in relation to students' behavioral cues is lacking, it is tentatively conjectured that both expert and novice teachers' reporting of student behavior is linked to previous findings from classroom observation research. Indicators observed in those studies refer to body-related aspects (e.g., hand raising), content-related aspects (e.g., quality of answers), and emotional-motivational aspects (e.g., expressed confidence). In addition, given expertise findings regarding the saliency of cues, it might be tentatively conjectured that novices more often report salient student behaviors, such as hand raising compared to experts.

RQ3: Do expert and novice teachers differ in their gaze regarding the five observed student profiles? Are experts more attentive than novices to inconsistent student profiles which might require pedagogically adaptive action? In addition, is there a difference in teacher gaze when analyzing two different classroom scenes, which are assumed to differ regarding bottom-up or top-down drivers of teacher gaze?

Based on previous research, it is assumed that top-down or bottom-up drivers are relevant indicators of teacher gaze. Therefore, a classroom context with a lot of motion regarding teacher and student activities should represent a context that activates mainly bottom-up drivers and leads to teacher gaze patterns that mainly follow the course of motion. Such a context could be represented by a typical whole-class situation in which teachers stand in the

front of a classroom, pose questions, and chose various students sitting within the classroom to answer questions. Thus, no expertise differences are expected.

In contrast, a classroom context with less motion and less variation in activities should represent a context that activates top-down drivers. Such a context could be, for example, a typical setting of individual seatwork in which students solve a task independently and interact with the teacher from time to time. This context provides possibilities for top-down drivers in teacher gaze and should lead to expected differences of higher teacher gaze for experts compared to novice teachers.

Since expert teachers are also better able to focus on events that might require further action, it is conjectured that expert teachers exhibit longer fixations in teacher gaze regarding the inconsistent student profiles that potentially require additional teacher support. Therefore, it is assumed that expert–novice differences are particularly visible with these student profiles.

Methods

Participants

The sample for which high-quality eye tracking data was available (see section below) consisted of 41 participants (63.8% female) and comprised of 33 novice teachers (female = 60.4%) enrolled in a bachelor's university program and eight in-service teachers (female = 75%) with an average teaching experience of 12.25 years ($SD = 8.18$, range = 1.5–25.0 years). Regarding the second research question on observed student cues, open answers were provided by 27 novice and seven expert teachers.

Procedure

Data collection took place in the laboratory. Participants were first introduced to the student characteristics under investigation (cognitive abilities, performance, self-concept, and interest) and their appearance in strong, struggling, overestimating, underestimating, and uninterested student profiles. After the introduction, participants watched a short video trailer (2:30 min) of the class to become familiar with the lesson topic and the classroom environment.

Participants were then instructed to watch a video stimulus and observe five target students carefully in order to diagnose the student profiles afterward. The five students were marked with letters (B, E, K, P, T) so that the participants were always aware of which students to observe and diagnose (Fig. 1). Eye tracking was conducted while the participants watched the stimulus video. Finally, participants assigned each marked student to one profile. Each profile could only be assigned once. Moreover, for each assigned student, participants could indicate whether they thought they might have confused the assigned profile with another one by stating a second choice. For each assigned profile, participants were invited to write down which indicators they had used to judge the student profile in an open answer format.

Video Stimulus

The 11-min video showed an eighth-grade introductory geometry lesson from a German high school (Fig. 1). The video clip was identified in the context of a previous video study on teacher–student interactions in classrooms (Seidel et al. 2016). The video stimulus consisted of two major scenes.



Fig. 1 Video stimulus for eye movement analysis. Note: This is an exemplary screenshot of the classroom and used AOIs. AOIs are only marked for illustration in this paper; they were not visible to the participants. The blurring of student faces is only added for the presentation in the publication to ensure the protection of data privacy; it was not visible when drawing the AOIs. Students were marked with letters not referring to any underlying profile: B, E, K, P, and T

The first scene (3.5 min) was a whole-class instructional scene in which the teacher stands in the front of the classroom and introduces the topic in a question-answer teaching style: various students in the classroom raise their hands and the teacher calls upon one of them and, in turn, poses the next question. The features of this scene indicate that bottom-up drivers might be more involved because of the salient student behavior. At the end of the first scene, tasks for the following individual work are explained. Afterwards, students work individually for a couple of minutes on their tasks (this phase was not used for teacher gaze analysis).

The second scene follows (4 min) and presents a situation in which the teacher stands in the front of the classroom at the board and explains the solutions to the tasks (the teacher, in this case, is not visible, only audible in the video). Meanwhile, the students listen to the teacher, compare their solutions to those presented by the teacher, and make corrections if necessary. Students rarely pose questions during this second scene. Due to the absence of salient student behavior in this scene, top-down drivers are assumed to be more involved in teacher gaze.

Student Profiles

The video sequence used for this study stems from a larger video study in which the videotaped students participated over the course of a school year (Seidel et al. 2016). Therefore, the five marked students with their profiles of strong, struggling, overestimating, underestimating, and uninterested were identified empirically in the context of this larger video study. A latent profile analysis, a common method of identifying homogenous subgroups of students, was conducted with achievement and self-concept of abilities as two major indicators. This analysis identified five profiles (Seidel et al. 2016) of which four are similar to previous latent class analyses (Seidel 2006): strong (26%), struggling (21%), overestimating (12%), and underestimating (9%). In addition, a fifth group of intermediate students with moderate performance and self-concept was identified (31%) (Seidel et al. 2016), who also

reported about being low in their subject interest and, therefore, are quite comparable to the uninterested profile as determined in the Seidel (2006) study.

Findings regarding the profiles of all students in the classroom were available. The students were then ranked independently by three of the authors to represent a particular student profile with regard to their engagement behavior. If the authors did not agree on their first choice, they discussed possible choices. The strong student often raised the hand during the first segment. Throughout the rest of the video, the student looked attentively at the board, took notes, and wrote in the book. The uninterested student continually played with the hands and sometimes chewed fingernails. The underestimating student raised the hand twice to answer the teacher's questions and, in the second segment, wrote notes in the book. The overestimating student checked with a neighbor before raising the hand in response to the teacher's questions, but when classmates were answering the teacher, the student looked around the class and played with a necklace. The struggling student looked forward toward the teacher and her classmates but never raised her hand. The struggling student's head often rested on her hands.

Apparatus

The stimulus environment was created with Experiment Center 3.7. Eye movements were recorded with the SMI RED 500 binocular remote eye tracker with a 22-in. display monitor and a sampling frequency of 500 Hz. Light was kept stable across data collection times by closing the blinds and using a ceiling light. Participants were positioned about 65 cm in front of the eye tracker. A height-adjustable table ensured adjustment to the individual participants' requirements, and a chin rest prevented participants from performing strong head movements or leaning back to increase precision (Nyström et al. 2013).

Measures

Teacher Judgment Accuracy

A student profile was considered to be judged correctly if it was concordant with the underlying data-driven student characteristic profile. Participants received zero points for an incorrect diagnosis or one point for each correct assignment. In cases in which participants first assigned an incorrect profile but stated the correct profile in their second choice, half a point was awarded. The participants could gain a total score between zero (no correct judgments) and five points (only correct judgments).

Reported Behavioral Indicators Used for Judging Student Profiles

Responses to the open answer were coded inductively with a coding scheme of 108 codes by two coders for 80% of the sample with a sufficient interrater agreement ($\kappa = .95$). The 108 single codes were aggregated in 10 indicators: engagement, attention, emotions, confidence, work on worksheet, quality of answers, (pre)achievement, interaction with peers, hand raisings, and body posture and facial expression. These indicators were coded for each participant and student profile with either 0 (indicator not stated) or 1 (indicator stated). In the last step, indicators for the single profiles were added to express how often one participant used one indicator overall (0 = not used; 4 = used for all profiles).

Teacher Gaze

The final average tracking ratio for the sample was 97%, and accuracy resulted in a deviation on the horizontal x -axis = $.55^\circ$ and vertical y -axis = $.51^\circ$. Therefore, the quality of eye movement data can be regarded as high.

To analyze teacher gaze, the very end of the video stimulus was excluded because the teacher was visible. It was assumed that this event would attract participants' attention due to their saliency of teacher actions. Excluding this part left a total of 09:06 min of collected eye movements.

The five marked students were each defined as dynamic areas of interest (AOIs) with BeGaze 3.7 to analyze fixations on the students (Fig. 1). The exactness of the AOIs was manually controlled every 16 ms (equaling intervals of four video frames). A velocity-based algorithm was used to define fixations (Holmqvist et al. 2011), with the system default similar to that of Wolff et al. (2016). Fixation count (i.e., fixations inside specific AOIs) was used for further analysis. The total number of fixations was calculated for each participant and the five AOIs. Similar fixation duration (i.e., the length of the fixations within an AOI) and the number of glances at an AOI (i.e., how often an AOI was visited) were considered for analysis. To compare the number of fixations, fixation durations, and glances across student profiles (AOIs) and video segments, the relative proportion was calculated.

Data Analysis

A two-way analysis of variance (5×2 ANOVA) was used to examine the differences in the judgment accuracy of the five different student profiles (factor 1) and the level of professional experience (factor 2). There was homogeneity of variances, as assessed by Levene's test for the quality of variances, $p > .05$.

Unpaired t tests were conducted to investigate whether novice and expert teachers differ in their observation of student cues or used more indicators per profile and overall. Then, unpaired t tests were conducted to test whether the two expertise groups differed in the type of observed indicators. Additionally, the type of observed indicators for each profile was compared descriptively.

Unpaired t tests were also used to investigate differences between the number of fixations and the average fixation duration novice and expert teachers spent on each of the five student profiles.

Results

Teacher Judgment Accuracy Regarding Student Profiles

To answer the first research question, the accuracy of teacher judgments of each student profile (separated by the level of professional experience) was analyzed. The descriptive results of the diagnosis are shown in Table 1.

Most of the novices and experts were able to diagnose the underestimating and struggling student correctly. However, the overestimating student most often was diagnosed incorrectly; many of the novices (56.7%) and experts (50%) received zero points for this diagnosis, which indicates that their first and second diagnosis were both incorrect. One hundred percent of the experts identified the uninterested student compared to 63.6% of the novices. Diagnosis of the strong student received many of

Table 1 Expertise differences in teacher judgment accuracy (descriptive statistics)

Student profiles	Expertise level	Relative frequency of scores (%)		
		0	0.5	1
Overestimating	Novices	57.6	12.1	30.3
	Experts	50.0	25.0	25.0
Uninterested	Novices	21.2	15.2	63.6
	Experts	0.00	0.00	100.0
Underestimating	Novices	15.2	18.2	66.7
	Experts	12.5	12.5	75.0
Strong	Novices	24.2	48.5	27.3
	Experts	12.5	50.0	37.5
Struggling	Novices	15.2	18.2	66.7
	Experts	12.5	12.5	75.0

$N_{\text{novices}} = 33$, $N_{\text{experts}} = 8$. For accuracy scores 1 = first choice correct, 0.5 = second choice correct, and 0 = first and second choice incorrect

half points from both groups (48.5% novices, 50% experts), indicating that this student profile was often selected as a correct alternative solution.

A two-way ANOVA was conducted to examine the effects of student profiles and expertise level on teacher judgment accuracy. There was no statistically significant interaction between student profiles and expertise level for teacher judgment accuracy, $F(4, 195) = .552$, $p = .698$, partial $\eta^2 = .011$.

The main effect for student profiles was statistically significant, $F(4, 265) = 5.727$, $p < .001$, partial $\eta^2 = .105$, indicating that participants could diagnose the five student profiles with different levels of accuracy. Marginal means of the accumulated diagnosis scores and subgroup analysis based on Tukey's post hoc results are shown in Table 2. Thereby, the two profiles of uninterested and struggling were judged comparably and accurately ($p = .999$, means .70–.74). Also, the strong, underestimating, and overestimating profile were judged comparably ($p = .407$), however, with lower accuracy scores (means .30–.46).

Furthermore, there was a significant main effect for expertise level, $F(1, 195) = 5.046$, $p = .026$, partial $\eta^2 = .025$. However, expert teachers ($M = 3.39$, $SD = 1.65$) did not show a systematically higher judgment accuracy compared to novices regarding all student profiles

Table 2 Marginal means and standard deviations of the accumulated judgment accuracy score and homogenous subgroups based on Tukey's HSD

Student profile	Number	Homogenous subgroups			
		1		2	
		Marginal Mean	SD	Marginal mean	SD
Overestimating	41	.30	.43		
Underestimating	41	.45	.49		
Strong	41	.46	.36		
Struggling	41			.70	.40
Uninterested	41			.73	.40
Significance	–	.407		.999	

($M = 2.47$, $SD = 1.81$), $t(39) = 1.31$, $p = .198$, $d = -0.53$ (Table 3). When testing on the level of each student profile, experts did show higher judgment accuracy for the underestimating student, $t(39) = 2.03$, $p = .048$, $d = -0.73$, the uninterested student, $t(39) = -2.19$, $p = .035$, $d = -1.04$, and the overestimating student, $t(39) = 2.04$, $p = .046$, $d = -0.68$.

Hypotheses regarding the first research question were partially confirmed. In line with our assumption is the result that the two inconsistent profiles of underestimating and overestimating were judged less accurate than the consistent struggling profile. However, both experts and novices judged the consistent profile of a strong student comparably inaccurate, as well as the inconsistent profile of an uninterested student quite accurate. Contrary to our expectation, experts in general were not more accurate than novices in judging the student profiles. However, in line with our hypotheses, they judged all inconsistent student profiles of underestimating, overestimating, and uninterested more accurate than novices.

Observed Behavioral Cues Used for Inferring Student Profiles

To address the second research question, open-ended answers regarding student cues that were used when inferring underlying profiles based on observations are reported in this section. Overall, both novices and experts stated about two to three indicators per student profile ($M_{\text{novice}} = 2.66$, $SD_{\text{novice}} = 0.72$; $M_{\text{expert}} = 2.94$, $SD_{\text{expert}} = 0.97$), resulting in no significant difference in expertise level, $t(25) = -0.82$; $p = .422$; $d = 0.36$). Regarding analyses of each student profile, no significant differences between experts and novices were found for strong, $t(29) = -0.50$; $p = .622$; $d = 0.21$, struggling, $t(29) = -0.39$; $p = .697$; $d = 0.17$, or underestimating students, $t(29) = -0.44$; $p = .663$; $d = -0.19$. However, experts stated more indicators than novices did for diagnosing the overestimating, $t(30) = -2.18$; $p = .037$; $d = 0.93$, and the uninterested student, $t(30) = -2.29$; $p = .029$; $d = 0.98$.

As to whether novice and expert teachers report different indicators, it was found that novices used student hand raising and the quality of answers significantly more often than experts. Experts used all other indicators more often than novices did, although significance was not reached (Table 4).

Altogether, expert and novice teachers stated valid behavioral student cues. Expert teachers seemed to use a range of indicators to diagnose student profiles while novices focused on salient indicators such as hand raising and the quality of answers.

Table 3 Expertise differences regarding the accumulated judgment accuracy score (descriptives and t tests)

Student profiles	Expertise level				t	df	95% CI	p	d
	Novices		Experts						
	n	M (SD)	n	M (SD)					
Overestimating	33	0.24 (0.31)	8	0.52 (0.49)	2.04	39	[-0.55; -0.18]	<i>.046</i>	-0.68
Uninterested	33	0.69 (0.42)	8	1.00 (0.00)	-2.19	39	[-0.64; -0.25]	<i>.035</i>	-1.04
Underestimating	33	0.35 (0.32)	8	0.62 (0.41)	2.03	39	[-0.54; -0.01]	<i>.048</i>	-0.73
Strong	33	0.47 (0.36)	8	0.50 (0.38)	-0.32	39	[-0.34; 0.25]	<i>.753</i>	-0.08
Struggling	33	0.72 (0.40)	8	0.75 (0.37)	-0.33	39	[-0.37; 0.27]	<i>.743</i>	-0.07
Overall	33	2.47 (1.81)	8	3.39 (1.65)	1.31	39	[-2.34; 0.51]	<i>.198</i>	-0.53

Note. Significant p values are in italics

Table 4 Expertise differences regarding observed student cues used for inferring to student profile (descriptives and *t* tests)

Indicator	Expertise level		95% CI	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
	Novices (<i>n</i> = 20) <i>M</i> (<i>SD</i>)	Experts (<i>n</i> = 7) <i>M</i> (<i>SD</i>)					
Body-related							
Body and face	0.65 (0.81)	1.14 (1.07)	[−1.29; 0.30]	−1.27	25	.214	0.56
Hand raising	2.90 (1.25)	1.00 (1.73)	[0.65; 3.15]	3.13	25	<i>.004</i>	−1.37
Interaction with peers	0.85 (0.75)	1.00 (0.82)	[−0.84; 0.54]	−0.45	25	.658	0.20
Content-related							
(Pre)achievement	0.40 (0.50)	0.57 (0.98)	[−1.08; 0.74]	−0.45	7.15 ^a	.670	0.26
Quality of answers	2.30 (1.03)	0.86 (1.21)	[0.47; 2.42]	3.05	25	<i>.005</i>	−1.34
Worksheet	1.25 (1.37)	1.88 (1.57)	[−1.89; 0.68]	−0.97	25	.341	0.43
Motivation-related							
Confidence	1.15 (0.93)	1.86 (1.07)	[−1.58; 0.17]	−1.66	25	.109	0.73
Attention	1.10 (0.93)	2.14 (1.57)	[−2.51; 0.43]	−1.65	7.65 ^a	.140	0.91
Emotions	0.65 (1.04)	1.43 (1.40)	[−1.81; 0.25]	−1.56	25	.131	0.69
Engagement	1.15 (1.23)	2.00 (1.53)	[−2.03; 0.33]	−1.48	25	.150	0.65

^aNo homogeneity of variances therefore adjusted statistics are reported. The single indicators include both positive (e.g., frequent hand raising) and negative (e.g., seldom hand raising) occurrences

Note. Significant *p* values are in italics

Teacher Gaze in Relation to Student Profiles

To answer the third research question regarding expertise differences in teacher gaze, the number of fixations and the average fixation duration for each student profile were analyzed. Then results are shown in Table 5 (whole-classroom instruction scene) and Table 6 (individual seatwork scene).

As can be seen in Table 5, there were no significant differences in the number and duration of fixations between novices and experts in the whole-classroom instruction scene. In the individual seatwork scene (Table 6), experts showed a significantly higher number of fixations on the uninterested, underestimating, and struggling students. Moreover, the average fixation duration of the novices was significantly longer for the overestimating and strong student.

Therefore, hypotheses regarding the third research question were partially confirmed. In line with our conjectures, no expertise differences were found for the whole-classroom instruction scene. Regarding the individual seatwork scene, expected expertise differences were found. In line with our hypotheses, expert teachers spend more fixations on the inconsistent profiles of the uninterested and underestimating student, as well as on the consistent struggling profile. No expertise difference, in contrast to our expectation, was found for the inconsistent overestimating student.

Discussion

Observing individual students as they engage in classroom activities and inferring underlying student characteristics based on these observations are highly relevant aspects of teachers' diagnostic competences (Heitzmann et al. 2019; Herppich et al. 2018; Kunter et al. 2013). The present study applied an eye movement methodology to describe teachers' perceptual processes in the context of ongoing diagnostic activities, a context many researchers have previously expressed the need to explore (Herppich et al. 2018; Spinath 2005). To date, eye

Table 5 Expertise differences in teacher gaze in whole-class instruction: descriptives and *t* tests for the number and duration of fixations

Student profiles	Eye-tracking metrics	Expertise level		95% CI for mean difference	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
		Novices <i>M</i> (<i>SD</i>)	Experts <i>M</i> (<i>SD</i>)					
Overestimating	No. of fixations	62.42 (18.85)	65.25 (15.07)	[- 16.92; 11.27]	-0.41	39	.687	-0.17
	Av. fixation duration (ms)	392.82 (98.76)	338.14 (149.35)	[- 32.66; 142.03]	1.26	39	.213	0.43
Uninterested	No. of fixations	71.33 (20.17)	84.25 (26.39)	[- 29.99; -4.16]	-1.53	39	.134	-0.55
	Av. fixation duration (ms)	418.47 (123.05)	377.31 (164.14)	[- 63.57; 145.88]	0.79	39	.431	0.28
Underestimating	No. of fixations	93.69 (35.18)	101.75 (36.58)	[- 36.42; 20.07]	-0.57	39	.562	-0.23
	Av. fixation duration (ms)	369.20 (117.16)	338.36 (144.39)	[- 66.81; 128.48]	0.63	39	.527	0.24
Strong	No. of fixations	80.67 (21.21)	76.88 (20.78)	[- 13.05; 20.63]	0.45	39	.651	0.18
	Av. fixation duration (ms)	324.08 (116.36)	365.96 (104.46)	[- 126.93; 43.16]	-0.99	39	.325	-0.38
Struggling	No. of fixations	66.18 (28.93)	83.38 (41.33)	[- 42.31; 7.93]	-1.38	39	.174	-0.48
	Av. fixation duration (ms)	435.61 (140.75)	395.01 (206.48)	[- 82.66; 163.84]	0.66	39	.509	0.23

Table 6 Expertise differences in teacher gaze in seatwork scene: descriptives and *t* tests for the number and duration of fixation

Student profiles	Eye tracking metrics	Expertise level		95% CI for mean difference	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
		Novices <i>M</i> (<i>SD</i>)	Experts <i>M</i> (<i>SD</i>)					
Overestimating	No. of fixations	74.88 (30.89)	88.00 (32.35)	[- 37.96, 11.72]	-1.08	39	.292	-0.42
	Av. fixation duration (ms)	400.98 (119.36)	309.50 (101.67)	[1.29, 181.25]	2.20	39	<i>.048</i>	0.83
Uninterested	No. of fixations	95.88 (32.81)	144.75 (57.49)	[- 79.51, -18.24]	-3.23	39	<i>.003</i>	-1.04
	Av. fixation duration (ms)	382.85 (111.20)	322.95 (105.13)	[- 27.89, 147.69]	1.38	39	.175	0.55
Underestimating	No. of fixations	114.94 (40.91)	149.51 (27.85)	[- 65.56, -3.56]	-2.26	39	<i>.030</i>	-0.99
	Av. fixation duration (ms)	360.73 (108.79)	297.63 (115.49)	[- 24.59, 150.81]	1.46	39	.154	0.56
Strong	No. of fixations	83.15 (28.78)	105.13 (48.54)	[- 48.44, 4.49]	-1.68	39	.101	-0.56
	Av. fixation duration (ms)	336.79 (84.51)	264.33 (66.66)	[7.43, 137.51]	2.25	39	<i>.022</i>	0.95
Struggling	No. of fixations	76.88 (34.02)	102.13 (23.82)	[- 51.08, -3.28]	-1.98	39	<i>.027</i>	-0.86
	Av. fixation duration (ms)	425.36 (154.39)	318.64 (122.77)	[- 12.21, 225.65]	1.82	39	<i>.077</i>	0.77

Note. Significant *p* values are in italics

movement methodology has been mainly applied to the study of teacher gaze in the context of classroom management, with a focus on attention processes regarding the entire class (Stürmer et al. 2017; van den Bogert 2016; van den Bogert et al. 2014; Wolff et al. 2016; Wolff et al. 2015).

In the present study, teacher gaze is investigated in the context of attending to individual students and inferring underlying student characteristics based on observations (Borko et al., 2008). Based on empirically identified student profiles (Seidel 2006), teachers participating in the present study were asked to observe five marked students in a natural, videotaped, and mathematics classroom setting and make judgments regarding the assumed underlying profiles of the students based on their observations. Each student represented one specific profile: strong, struggling, overestimating, underestimating, or uninterested. Teachers' eye movements were recorded while they observed the students in the two major scenes of the video sequence: first, a whole-classroom instruction, and second, an individual seatwork phase.

For these two scenes, different affordances regarding bottom-up versus top-down drivers for teacher gaze were assumed (DeAngelus and Pelz 2009; Gegenfurtner et al. 2011; Jarodzka et al. 2013; Schütz et al. 2011). During the whole-class scene more motion of interaction between teachers and students was present due to frequent and simultaneous hand raisings and verbal contributions in comparison to the seatwork scene, in which students worked individually. Therefrom, it was conjectured that bottom-up drivers might be more involved during the whole-class instruction scene because the described active student behavior is very salient and thus a likely driver of bottom-up cognitive processing (Schütz et al. 2011). For the seatwork scene, top-down drivers were assumed to play a more dominant role in teacher gaze due to the absence of active and salient student behaviors such as hand raisings. In order to test this assumption, the design of an expert–novice comparison was chosen to identify assumed top-down, knowledge-related differences, particularly for the seatwork scene. We outline this new study design for investigating teacher gaze in the context of teacher diagnostic competences to provide a better basis for interpreting the findings as summarized in the follow paragraphs.

To address the first research question, previous research regarding accuracy of teacher judgment of student profiles was acknowledged and applied to the context of this study (Huber and Seidel 2018; Kaiser et al. 2017; Linnenbrink-Garcia et al. 2012; Pielmeier et al. 2018; Praetorius et al. 2017; Südkamp et al. 2018). Previous studies mainly included in-service and experienced teachers, without providing systematic comparisons to novices and their typical misjudgment patterns. The findings of the present study regarding the judgment accuracy of experienced teachers are in line with previous findings: expert teachers in the present study were also comparatively more inaccurate in judging incoherent profiles than coherent ones (except for the incoherent profile of the uninterested student, who was also judged highly accurately by the experienced teachers of this sample). In addition, the expert–novice comparison revealed that expert teachers were significantly more accurate than novices at judging the incoherent profiles of the overestimating, underestimating, and uninterested students. This finding is interesting from a pedagogical perspective since these students require adaptive teacher support regarding further instructional aids, encouragement, and motivation (Hattie and Timperley 2007).

The second research question explored which observational cues expert and novice teachers used when inferring to underlying student profiles. Participants' written responses were first analyzed inductively and then categorized along three major aspects of body-related, content-related, and motivation-related indicators. The categorizations were based on observational classroom research (cf. Seidel et al. 2005). Since research on the connection between teacher

judgments and observed behavioral indicators is lacking (Spinath 2005; Südkamp et al. 2018), answering the second research question was more of explorative nature. However, in line with our tentative conjectures, the findings of the present study show that both novices and experts stated about two to three behavioral indicators for each student profile. While experts tended to state more and a broader range of observed indicators than novices did, these differences yielded in no significant effect. However, in line with our tentative conjecture, novices more frequently cited hand raising and quality of answers compared to experts, two indicators which might be regarded as more salient compared to other student cues, such as expressed confidence. Experts' statements, by contrast, were manifold and typically covered a broad range of body-, content-, and motivation-related indicators.

Overall, the qualitative analysis of behavioral cues revealed a valid pattern regarding the kind of indicators used for making inferences to underlying student profiles. Previous findings of classroom studies show comparable results regarding observed student engagement in real classrooms (Huber et al. 2015; Jurik et al. 2013). Despite these valid responses, however, further research is required to learn more about possible misinterpretations when determining underlying student profiles based on observations. Therefore, think-aloud protocols might be of additional value (Cortina et al. 2015).

To address the third research question, teacher gaze was investigated in order to learn more about the perceptual processes of noticing and knowledge-based reasoning (Lachner et al. 2016). Thereby, the present study, and the seatwork scene, in particular, could provide first evidence that expert teachers more frequently fixate on student profiles that might be in need for further teacher support: struggling, underestimating, and uninterested students. Considering the diagnostic task, it might be that experts focused on these profiles simply because they were difficult to judge. However, these three profiles were also those which were most often correctly assigned. Therefore, the first explanation seems quite likely and experts might have observed the struggling, underestimating, and uninterested student more closely because of activated schemas that also involve possible teacher supportive actions (Borko et al. 2008). Nevertheless, to clarify which one of the two interpretations fit best, future research needs to combine eye tracking data with additional data sources that give possibility for studying underlying cognitive processes (Cortina et al. 2015). Novices, by comparison, exhibited longer fixations overall compared to experts, with significant effects for the strong and overestimating students. Both findings indicate that top-down drivers might have been involved in the seatwork scene (Gegenfurtner et al. 2011), resulting in higher frequencies for expert teachers due to faster information processing, and longer durations for novices possibly due to slower processing.

Nevertheless, for the whole-class instruction scene, no significant differences between novice and expert teachers were found although experts showed a tendency, similar to the seatwork scene, toward more and shorter fixations on all profiles besides the strong one. Therefore, this finding needs to be interpreted with caution. It supports on the one hand the hypothesis, that both novices and experts' gaze are driven by bottom-up processes during whole-class instruction, when many students show simultaneously active and salient behaviors such as hand raisings or verbal contributions. However, on the other side the described tendency contradicts the hypothesis and suggests that in line with the information-reduction theory (Haider and Frensch 1996) experts are even in such complex classroom situations able to focus their attention on relevant students. To clarify which cognitive processes are underlying novices and experts gaze when observing different classroom scenes, future research should apply combined measures of gaze data and for example think-aloud data (Cortina et al., 2015).

The findings regarding the third research question further expand the method of eye movement to the field of teacher research. Previous studies on teacher gaze have been predominantly based on classroom management and focus on answering the question of how teachers split their attention between monitoring the class as a whole and focusing on incidences that need specific attention (e.g., specific disturbances) (van den Bogert 2016; van den Bogert et al. 2014; Wolff et al. 2016; Wolff et al. 2015). In the present study, the professional tasks of observing individual students as they respond to teachers' instructional activities and relating those observations to underlying student characteristics was investigated (Borko et al. 2008).

Studying teacher gaze in this professional context, the extent to which top-down drivers might be present should be considered in more depth when video scenes are selected. While no systematic differences were found in the present study for the first scene of whole-class teaching, the second scene of students participating in individual seatwork revealed hypothesized expert–novice differences as connected with top-down drivers of teacher gaze. Thereby, it is interesting to consider that expert teachers paid more attention to students who might require additional teacher support in this scene. From a pedagogical viewpoint, teachers are often required to implement more individualized working phases since these phases give teachers more opportunities to monitor the students' learning progress (Hattie and Timperley, 2007). In the present study, it seems that expert teachers paid closer attention to specific students and their underlying profiles during the seatwork scene.

Since this is a first study, the findings are limited with regard to their generalizability and replicability which implies that further studies with different teaching scenes are needed. However, the present study provides the first insights to applicable research designs and first findings regarding the interplay between perceptual processes, observed cues, and inferences to underlying student characteristics. Moreover, the marking of the students with visible letters throughout the whole video might have forced bottom-up processing due the letters' saliency in comparison to top-down processing. In the present study, research questions were based on a cognitive model of teachers' professional vision, including both perceptual noticing as well as knowledge-based reasoning (Lachner et al. 2016). To date, the noticing component has been mainly determined by post hoc identification of marked events and then linked to knowledge-based reasoning regarding those noticed events (van Es et al. 2017). In our theoretical conceptualization, we argued that professional vision research might be moved forward through the use of teacher gaze as a way to operationalize perceptual noticing in the professional vision model. The findings of the present study tentatively suggest that adding teacher gaze to the model can enrich our understanding of the processes involved in knowledge-based reasoning.

Conclusion

The present study is the first to expand teacher gaze research to the context of teachers' diagnostic activities in the process of teaching. Expert teachers tended to pay more attention than novices did to those students who might be in particular need of further support. In addition, experts tended also to be more accurate in judging incoherent student profiles based on observed student engagement. In addition, the present study explored possible relationships between perceptual processes and teacher judgment by analyzing behavioral cues used to infer to underlying student profiles. The findings indicate that the observed cues are quite valid with regard to student engagement, as found in observational classroom research. It can be tentatively concluded that

assumed expert–novice differences in teacher gaze can also be found for diagnostic activities of individual students. Furthermore, using teacher gaze as a way to operationalize the noticing component of teachers’ professional vision seems to be promising.

Funding Information Open Access funding provided 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

- Back, M. D., & Nestler, S. (2016). Accuracy of judging personality. In J. A. Hall, M. S. Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 98–124). Cambridge: Cambridge University Press.
- 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. (1991). Perceptions of student behavior as a function of expertise. *Journal of Classroom Interaction*, 26(1), 1–8.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Approaches to competence measurement in higher education. *Zeitschrift für Psychologie*, 233(1), 1–2. <https://doi.org/10.1027/2151-2604/a000249>.
- Borko, H., Roberts, S. A., & Shavelson, R. (2008). Teachers’ decision making: from Alan J. Bishop to today. In P. Clarkson & N. Presmeg (Eds.), *Critical issues in mathematics education* (1. Aufl. ed., pp. 37–67). New York: Springer.
- Brophy, J., & Good, T. (1970). Teachers’ communication of differential expectations for children’s classroom performance. *Journal of Educational Psychology*, 61(5), 365–374.
- Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge. *International Journal of Science and Mathematics Education*, 13(2), 389–403. <https://doi.org/10.1007/s10763-014-9610-5>.
- DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements. *Visual Cognition*, 17(6–7), 790–811. <https://doi.org/10.1080/13506280902793843>.
- Doyle, W. (2006). Ecological approaches to classroom management. In C. M. Everson & C. S. Weinstein (Eds.), *Handbook of classroom management: research, practice, and contemporary issues* (pp. 97–125). Mahwah, NJ: Erlbaum.
- Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development. A literature review. *Educational Research Review*, 16, 41–67. <https://doi.org/10.1016/j.edurev.2015.06.001>.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations. A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>.
- Glock, S., Krolak-Schwerdt, S., Klapproth, F., & Böhmer, M. (2013). Beyond judgment bias. How students’ ethnicity and academic profile consistency influence teachers’ tracking judgments. *Social Psychology of Education*, 16(4), 555–573. <https://doi.org/10.1007/s11218-013-9227-5>.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633.
- Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, 30(3), 304–337. <https://doi.org/10.1006/cogp.1996.0009>.
- Hattie, J. (2008). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.

- Heitzmann, N., Seidel, T., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., ... Fischer, F. (2019). Facilitating diagnostic competences in simulations: a conceptual framework and a research agenda for medical and teacher education. *Frontline Learning Research*, 7(4), 1–24. <https://doi.org/10.14786/flr.v7i4.384>.
- Herppich, S., Praetorius, A. K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence. Integrating knowledge-, process, and product-oriented approaches into a competence-oriented conceptual mode. *Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>.
- Hoffmann, L. (2002). Promoting girls' learning and achievement in physics classes for beginners. *Learning and Instruction*, 12(4), 447–465. [https://doi.org/10.1016/S0959-4752\(01\)00010-X](https://doi.org/10.1016/S0959-4752(01)00010-X).
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking*. Oxford: Oxford University Press.
- Huber, S. A., & Seidel, T. (2018). Comparing teacher and student perspectives on the interplay of cognitive and motivational-affective student characteristics. *PLoS One*, 13(8), e0200609. <https://doi.org/10.1371/journal.pone.0200609>.
- Huber, S. A., Häusler, J., Jurik, V., & Seidel, T. (2015). Self-underestimating students in physics instruction: development over a school year and its connection to internal learning processes. *Learning and Individual Differences*, 43, 83–91. <https://doi.org/10.1016/j.lindif.2015.08.021>.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder. How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154. <https://doi.org/10.1016/j.learninstruc.2009.02.019>.
- Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see. *Learning and Instruction*, 25, 62–70. <https://doi.org/10.1016/j.learninstruc.2012.11.004>.
- Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction*, 23, 33–42. <https://doi.org/10.1016/j.learninstruc.2012.09.002>.
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement. How student characteristics influence teacher judgments. *Learning and Instruction*, 28, 73–84. <https://doi.org/10.1016/j.learninstruc.2013.06.001>.
- Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy. *Journal of Educational Psychology*, 109(6), 871–888. <https://doi.org/10.1037/edu0000156>.
- 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>.
- Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science*, 44(3), 197–203. <https://doi.org/10.1007/s11251-016-9376-y>.
- Linnenbrink-Garcia, L., Pugh, K. J., Koskey, K. L. K., & Stewart, V. C. (2012). Developing conceptual understanding of natural selection. *The Journal of Experimental Education*, 80(1), 45–68. <https://doi.org/10.1080/00220973.2011.559491>.
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K., & Klieme, E. (2007). Gleicher Unterricht - gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft*, 35(2), 125–147.
- Lipowsky, F., Pauli, C., & Rakoczy, K. (2008). Schülerbeteiligung und Unterrichtsqualität. In M. Gläser-Zikuda & J. Seifried (Eds.), *Lehrerexpertise* (pp. 67–90). Münster: Waxmann.
- McIntyre, N. A., Jarodzka, H., & Klassen, R. M. (2017). Capturing teacher priorities. *Learning and Instruction*, 60, 215–224. <https://doi.org/10.1016/j.learninstruc.2017.12.003>.
- Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45(1), 272–288. <https://doi.org/10.3758/s13428-012-0247-4>.
- Pauli, C., & Lipowsky, F. (2007). Student participation in whole-class discussions and teacher-student interactions in mathematics classrooms. *Unterrichtswissenschaft*, 35(2), 101–124.
- Pielmeier, M., Huber, S., & Seidel, T. (2018). Is teacher judgment accuracy of students' characteristics beneficial for verbal teacher-student interactions in classroom? *Teaching and Teacher Education* 76, 255–266. <https://doi.org/10.1016/j.tate.2018.01.002>.
- Praetorius, A.-K., Koch, T., Scheunpflug, A., Zeinz, H., & Dresel, M. (2017). Identifying determinants of teachers' judgment (in) accuracy regarding students' school-related motivations using a Bayesian cross-classified multi-level model. *Learning and Instruction*, 52, 148–160. <https://doi.org/10.1016/j.learninstruc.2017.06.003>.

- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology* (2006), 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>.
- Sacher, W. (1995). *Meldungen und Aufrufe im Unterrichtsgespräch*. Augsburg: Wissner.
- Seidel, T. (2006). The role of student characteristics in studying micro teaching–learning environments. *Learning Environments Research*, 9(3), 253–271. <https://doi.org/10.1007/s10984-006-9012-x>.
- Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: findings from a video study. *Learning and Instruction*, 16(3), 228–240. <https://doi.org/10.1016/j.learninstruc.2006.03.002>.
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51(4), 739–771. <https://doi.org/10.3102/0002831214531321>.
- Seidel, T., Prenzel, M., & Kobarg, M. (Eds.). (2005). *How to run a video study: technical report of the IPN Video Study*. Münster: Waxmann.
- Seidel, T., Jurik, V., Häusler, J., & Stubben, S. (2016). Micro-environments within classrooms: how cognitive and motivational-affective student characteristics affect students' classroom perception and behavior. In N. McElvany, W. Bos, H. G. Holtapples, M. M. Gebauer, & F. Schwabe (Eds.), *Conditions and effects of good teaching [Bedingungen und Effekte guten Unterrichts]* (pp. 65–88). Münster: Waxmann.
- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception. A selective review. *Journal of Vision*, 11(5). <https://doi.org/10.1167/11.5.9>.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(4), 3–14.
- Snow, R. E., Corno, L., & Jackson III, D. (1996). Individual differences in affective and conative functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 243–310). New York: Macmillan.
- Spinath, B. (2005). Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95. <https://doi.org/10.1024/1010-0652.19.12.85>.
- Südkamp, A., Praetorius, A.-K., & Spinath, B. (2018). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education*, 76, 201–213. <https://doi.org/10.1016/j.tate.2017.09.016>.
- Stürmer, K., Seidel, T., Müller, K., Häusler, J., & Cortina, K. S. (2017). What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft*, 20(1), 75–92. <https://doi.org/10.1007/s11618-017-0731-9>.
- Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., Oswalt, S., Snow, J. L., Sutton, J., & Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education*, 49, 36–44. <https://doi.org/10.1016/j.tate.2015.01.012>.
- van den Bogert, N. (2016). *On teachers' visual perception and interpretation of classroom events using eye tracking and collaborative tagging methodologies*. Eindhoven: Technische Universiteit Eindhoven.
- van den Bogert, N., van Bruggen, J., Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teaching and Teacher Education*, 37, 208–216. <https://doi.org/10.1016/j.tate.2013.09.001>.
- van Es, E. A., Cashen, M., Barnhart, T., & Auger, A. (2017). Learning to notice mathematics instruction: using video to develop preservice teachers' vision of ambitious pedagogy. *Cognition and Instruction*, 35(3), 165–187. <https://doi.org/10.1080/07370008.2017.1317125>.
- 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. <https://doi.org/10.1007/s11251-016-9367>.
- Wolff, C. E., van den Bogert, N., Jarodzka, H., & Boshuizen, H. P. A. (2015). Keeping an eye on learning: Differences between expert and novice teachers' representations of classroom management events. *Journal of Teacher Education*, 66(1), 68–85. <https://doi.org/10.1177/0022487114549810>.

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

Affiliations

Tina Seidel¹ · Katharina Schnitzler¹ · Christian Kosel¹ · Kathleen Stürmer² · Doris Holzberger¹

¹ TUM School of Education, Technical University of Munich (TUM), Arcisstraße 21, 80333 Munich, Germany

² Hector Research Institute of Education Sciences and Psychology, University of Tuebingen, Europastrasse 6, 72072 Tübingen, Germany